

META-ANALYSES OF GENE EXPRESSION IN AGE-DEPENDENT DISEASES

By

Lavida Rashida Kenera Rogers

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Microbiology and Molecular Genetics – Doctor of Philosophy

2019

ABSTRACT

META-ANALYSES OF GENE EXPRESSION IN AGE-DEPENDENT DISEASES

By

Lavida Rashida Kenera Rogers

Physiological changes with age such as immune system decline and brain aging cause an increased risk for diseases. Age-related diseases are of major concern especially in the elderly population due to there being an increase in the average lifespan. This dissertation explores neurodegenerative and respiratory diseases and how gene expression varies due to disease status, age tissue and sex.

Alzheimer's disease (AD) has been categorized by the Centers for Disease Control and Prevention (CDC) as the 6th leading cause of death in the United States. AD is a significant health-care burden because of its increased occurrence (specifically in the elderly population), and the lack of effective treatments and preventive methods. AD targets neuronal function and can cause neuronal loss due to the buildup of amyloid-beta plaques and intracellular neurofibrillary tangles.

The respiratory disease, chronic obstructive pulmonary disease (COPD), was classified by the Centers for Disease Control and Prevention in 2014 as the 3rd leading cause of death in the United States. The main cause of COPD is exposure to tobacco smoke and air pollutants. In addition to exploring genetic variation due to disease state, sex and age we also explored the role of smoking status on expression profiles.

Additionally, the respiratory infections, influenza and pneumonia affect thousands of people worldwide. Young children, elderly and immunocompromised individuals are at higher risk for being infected by the influenza virus and *Streptococcus pneumoniae*. Host responses to these pathogens and vaccinations vary by the state of one's immune system.

This dissertation includes multiple meta-analyses to assess genetic variation in Alzheimer's disease, COPD and Influenza, and an assessment of pneumococcal disease and aging. To identify significant differentially expressed genes we ran an analysis of variance with a linear model with disease state, age, sex, tissue, smoking status and study as effects that also included binary

interactions.

Our meta-analysis approach effectively combined multiple publicly available microarray datasets to identify gene expression differences across diseases including full age, sex, smoking status and tissue type considerations. Our findings provide potential gene and pathway associations that can be targeted to improve treatment and prevention of diseases.

Copyright by
LAVIDA RASHIDA KENERA ROGERS
2019

To my siblings

ACKNOWLEDGEMENTS

The last five years at Michigan State University has comprised of five years of commitment, diligence and hard work, which was often accompanied by teary eyes, long days and long nights. I would like to acknowledge a number of people whose guidance and support made this journey a lot easier.

Firstly, I would like to thank my PhD advisor, Dr. George Mias. Thank you for always believing in me and motivating me to go above and beyond at every task I tackled. Thank you for promoting a positive work environment and for always being available. Thank you for challenging me to think critically, giving me both computational and experimental experience, and pushing me out of my comfort zone. Thank you for the many lessons in statistics and calculus - you took a keen interest in my scientific career and I appreciate you for that. Thank you for being patient and for also allowing me to spazz, but somehow always finding a way to show me that everything will work out in the end. Thank you for being supportive of all my activities outside of lab which all helped me to develop scientifically and professionally. Thank you for answering all my career-related questions, reading over my emails and many letters of recommendation! I would not have made it to this point without your support.

To the past and present members of the Mias lab: Vikas Singh, Raeuf Roushangar and Eren Veziroglu - thank you for being great lab mates! To the current and past undergraduates: Alisha Ungkuldee, Connor Schury, Curtis Bunker, Ashley Garvin, Jayna Lenders, Jennifer Abel, Michael Bennet and Maddie Verlinde - it was an honor mentoring and doing science with you.

I would also like to thank the members of my guidance committee - Dr. Patrick Venta, Dr. Shin-han Shui, Dr. Chris Waters and Dr. Kefei Yu. Thank you all for always pushing me to think critically about my research. Thank you for always challenging my research topics and methods as this allowed me to delve deeper into the literature to truly understand why my approach was fitting to achieve my research goals. Thank you for ensuring every meeting ended on a great and positive note. Your expertise and support has made producing this dissertation possible.

I would like to thank the Michigan State University (MSU) chapter of Graduate Women in Science (GWIS), the MSU Alliance for Graduate Education and the Professoriate (AGEP), and the Charles Drew Science Scholars. Thank you for offering experiences outside of lab and for providing opportunities for professional development, mentoring and teaching experience over the years.

I'd like to thank the Microbiology and Molecular Genetics Department, Biochemistry and Molecular Biology Department and The Institute For Quantitative Health Science & Engineering and their staff for assistance with all the logistics and paper work of being in graduate school. Special thanks to Dr. Poorna Viswanathan for her guidance throughout my teaching assistantship and all my teaching endeavors. Thank you to Dr. Donna Koslowsky and Dr. Victor DiRita for their guidance through the MMG program.

To the friends that I made along the way shout out to you for also enduring the pressures of graduate school with me. To Dr. Taylor Dunivin, thank you for agreeing to study with me in BMB 801. I am so grateful for our friendship over the past 5 years, and I thank you for the many subway dates. I am so glad I got to share this journey with you. To Dr. Isola Brown, thank you for always being available to give advice and for being a phone call away. I appreciate you! To my best friends - thank you for always allowing me to vent and for your continuous support.

To my family, I could not have made it here without your never-ending support and encouragement. To my siblings, Rajeem, Laesha, Leondre and Rejanae, as your big sister, you push me to be my best self. Laesha you have been my rock while here in Michigan, and I appreciate you for just being there to listen to me, and for making sure I was meeting deadlines. To my mother, aunts and uncles, grandparents, and in-laws thank you for always believing in me! #LongLiveTheHobsonRogersSupportSystem

Lastly, to my husband, Jerome, thank you for your patience and for providing unwavering love and support throughout my time in graduate school. Thank you for remaining confident in my ability to reach the end and for comforting and uplifting me through the trying times - I love you.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xiii
KEY TO ABBREVIATIONS	xxi
CHAPTER 1 OVERVIEW AND AIMS OF DISSERTATION	1
1.1 Overview	2
1.2 Aims of Dissertation	3
CHAPTER 2 DATA-DRIVEN ANALYSIS OF AGE, SEX, AND TISSUE EFFECTS ON GENE EXPRESSION VARIABILITY IN ALZHEIMER'S DISEASE	4
2.1 Abstract	5
2.2 Introduction	6
2.3 Methods	8
2.3.1 Microarray Data Curation	9
2.3.2 Pre-processing and Data Normalization	10
2.3.3 Visualizing Variation due to Batch Effects	11
2.3.4 Analysis of Variance	12
2.3.5 Identifying Up and Down Regulated Genes by Factor	13
2.3.6 Gene Ontology and Reactome Pathway Analysis	14
2.4 Results	15
2.4.1 ComBat Batch Effect Visualization	15
2.4.2 Analysis of Variance on Gene Expression By Disease State	16
2.4.3 Up and Down- Regulated Gene Expression in AD and Sex Specific Differences	18
2.4.4 Aging and Tissue Differences in AD Gene Expression	23
2.5 Discussion	28
2.5.1 Significant Gene Expression Differences Due to Disease Status and Bio- logical Significance	29
2.5.2 Sex, Age and Tissue Effect on Disease Status Biologically Significant Genes	34
2.5.3 Limitations of the Study	37
2.5.4 Future Directions and Recommendations	38
CHAPTER 3 <i>STREPTOCOCCUS PNEUMONIAE</i> 'S VIRULENCE AND HOST IM- MUNITY: AGING, DIAGNOSTICS AND PREVENTION	40
3.1 Abstract	41
3.2 Introduction	41
3.3 Pneumococcal Disease, Epidemiology and Transmission	44
3.4 Transmission	52
3.4.1 Transmission Via Coinfections:	53
3.5 <i>S. pneumoniae</i> 's Virulence Factors	54
3.6 Host Immune System Responses to <i>S. pneumoniae</i>	66

3.6.1	Innate Immune Responses	67
3.6.2	Adaptive Immune Responses (B and T Cells)	74
3.6.3	Additional Immune Response Considerations	76
3.7	Diagnosis, Age-dependent Response, Prevention and Disease Prognosis	79
3.7.1	Diagnosis	79
3.7.2	Prevention, Antibiotic Resistance and Age-Dependent Immune Responses	81
3.7.3	Post-infection Prognosis	84
3.8	Discussion	86
CHAPTER 4 META-ANALYSIS OF GENE EXPRESSION MICROARRAY DATASETS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE.		89
4.1	Abstract	90
4.2	Introduction	90
4.3	Materials and Methods	92
4.3.1	Microarray Data Curation from Gene Expression Omnibus and Array Express	94
4.3.2	Microarray Pre-processing and BoxCox Normalization	96
4.3.3	Identifying and Visualizing Batch Effects	97
4.3.4	Analysis of Variance to Identify Differentially Expressed Genes by Factor	97
4.3.5	Machine Learning with COPD	99
4.4	Results	99
4.4.1	Visualizing Batch Effects and Batch Effect Correction	99
4.4.2	Variance in Gene Expression Due to Disease Status	100
4.4.3	Up and Down- Regulated Gene Expression in COPD	104
4.4.4	Sex and Age on COPD Expression	109
4.5	Discussion	113
CHAPTER 5 MICROARRAY GENE EXPRESSION DATASET RE-ANALYSIS RE- VEALS VARIABILITY IN INFLUENZA INFECTION AND VACCI- NATION.		120
5.1	Abstract	121
5.2	Introduction	122
5.3	Methods	123
5.3.1	Data Curation: Gene Expression Omnibus	125
5.3.2	Data Pre-Processing in R and Mathematica	128
5.3.3	Linear Mixed Effects Modeling	129
5.3.4	Determining Gene Expression Variability between Influenza Infection and Vaccination	130
5.4	Results	130
5.4.1	Differentially Expressed Genes in Influenza Disease and Vaccination	131
5.4.2	Age and Sex Effect on Gene Expression in Influenza	139
5.5	Discussion	144
CHAPTER 6 CONCLUSION AND FUTURE DIRECTIONS		151
6.1	Conclusion	152
6.2	Limitations and Future Directions	153

APPENDICES	155
APPENDIX A DATA-DRIVEN ANALYSIS OF AGE, SEX, AND TISSUE EF- FECTS ON GENE EXPRESSION VARIABILITY IN ALZHEIMER'S DISEASE SUPPLEMENTARY DATA	156
APPENDIX B META-ANALYSIS OF GENE EXPRESSION MICROARRAY DATASETS IN CHRONIC OBSTRUCTIVE PULMONARY DIS- EASE SUPPLEMENTARY DATA	171
APPENDIX C MICROARRAY GENE EXPRESSION DATASET RE-ANALYSIS REVEALS VARIABILITY IN INFLUENZA INFECTION AND VACCINATION SUPPLEMENTARY DATA	180
BIBLIOGRAPHY	188

LIST OF TABLES

Table 2.1: Curated microarray datasets and the study description.	8
Table 2.2: Patient characteristics for curated datasets.	11
Table 2.3: Top 25 KEGG Pathways using differentially expressed genes.	18
Table 2.4: Top 25 up- and down- regulated genes in Alzheimer’s disease compared to healthy controls.	21
Table 3.1: Occurrence of pneumococcal diseases (Cases and Death Rates) from 1995 to 2015 as reported by the Centers for Disease Control. Rates are per 100,000 population for Active Bacterial Core surveillance (ABCs) areas	46
Table 3.2: Selected virulence factors of <i>S. pneumoniae</i> , their location, and function.	47
Table 4.1: Description of datasets used in the meta-analysis	94
Table 4.2: Sample Characteristics By Dataset	96
Table 4.3: Enriched KEGG Pathways using the ANOVA Differentially Expressed Genes from Disease Factor	102
Table 4.4: Top 25 up and down regulated differentially expressed genes in COPD based on effect size	105
Table 5.1: Demographics of curated influenza microarray datasets.	125
Table 5.2: Enriched KEGG Pathways from Statistically Significant Genes with an Interaction Between Disease Status and Age.	137
Table 5.3: Top 10 Up- and Down- Regulated Differentially Expressed Genes from the Influenza Infected and Influenza Vaccination Biologically Significant Gene Lists (based on estimates).	138
Table A.1: Additional information reported from datasets on samples used for the meta-analysis.	167
Table A.2: Quantiles on differences of means between group comparisons from TukeyHSD analysis for each factor with the 10% and 90% highlighted.	168

Table A.3: TukeyHSD results (male-female) table of statistically significant differentially expressed disease genes with sex effect.	169
---	-----

LIST OF FIGURES

Figure 2.1: Alzheimer’s disease meta-analysis framework. (A) Simplified workflow used for the meta-analysis, (B) Pipeline for curating microarray data, (C) Pipeline for pre-processing the microarray data, (D) Methods used for meta-analysis of raw expression microarray data.	9
Figure 2.2: Principal component analysis of the study factor before and after batch correction with ComBat.	16
Figure 2.3: Principal component analysis of the tissue factor before and after batch correction with ComBat.	16
Figure 2.4: Enriched genes from the ANOVA statistically significant disease status gene list (p-value <0.05) found in the KEGG Alzheimer’s disease pathway (hsa05010) [[1–3]]. The yellow shading represents up-regulated and the blue shading represents down-regulated in AD samples. These genes were not yet filtered for biological significance.	19
Figure 2.5: Pathway-gene network of top 10 enriched Reactome pathways from down-regulated genes in Alzheimer’s disease patients.	22
Figure 2.6: Pathway-gene network of top 10 enriched Reactome pathways from up-regulated genes in Alzheimer’s disease patients.	23
Figure 2.7: Heatmap with gene clustering to visualize age group effect (difference in means) on the differentially expressed disease (control-AD) gene list that have agegroup:disease status interaction.	25
Figure 2.8: Heatmap with gene clustering to visualize tissue effect (difference in means) on the differentially expressed disease (control-AD) gene list that have tissue:disease status interaction.	27
Figure 3.1: Global distribution of lower respiratory infections by sex. Highlighted in this figure is the distribution of the disability adjusted life year (DALY) per 100,000 (2016) for four major lower respiratory infections worldwide by sex. Data obtained from Institute for Health Metrics and Evaluation [4]	42
Figure 3.2: Global distribution of lower respiratory infections with age. This figure shows the age-dependent disease burden to lower respiratory infections especially pneumococcal pneumonia based on the disability adjusted life year (DALY) data from 2016. Data obtained from Institute for Health Metrics and Evaluation [4]	44

Figure 3.3: Schematic cross section of <i>Streptococcus pneumoniae</i> cell wall. The bacterial cell wall composes of teichoic acids, a thick peptidoglycan layer, and a phospholipid bilayer	45
Figure 3.4: Virulence factors of <i>Streptococcus pneumoniae</i> . There are a variety of proteins and toxins that are expressed by <i>S. pneumoniae</i> that drive its pathogenesis. The major virulence factors are highlighted in the figure. Abbreviations: PsaA, pneumococcal surface adhesin A; PspA, pneumococcal surface protein A; PspC, pneumococcal surface protein C; PiaA, pneumococcal iron acquisition A; PiuA, pneumococcal iron uptake A; PitA, pneumococcal iron transporter.	48
Figure 3.5: Worldwide disability adjusted life year (DALY) of pneumococcal pneumonia. Global distribution of pneumococcal pneumonia on a log10 scale of the 2016 DALY per 100,000 pneumococcal pneumonia data obtained from Institute for Health Metrics and Evaluation [4]	50
Figure 3.6: Global distribution of lower respiratory infections over time. This figure depicts how the burden for four major lower respiratory infections changes over time in response to the introduction of antibiotic treatments and vaccine implementation. Disability adjusted life year (DALY) data obtained from Institute for Health Metrics and Evaluation [4]	51
Figure 3.7: Host surface and intracellular receptors necessary for immune response to <i>Streptococcus pneumoniae</i> . Highlighted in this figure are the major pathogen recognition receptors necessary for binding to pneumococcal ligands and eliciting an immune response. Upon binding to the ligands, receptors and signaling pathways are activated, which leads to the overall production of inflammatory cytokines and recruitment of immune cells. There are 10 toll-like receptors (TLRs) that have been discovered in humans—TLRs involved in pneumococcal disease are depicted in the figure	67
Figure 3.8: <i>Streptococcus pneumoniae</i> 's interaction with host epithelial cells. Two types of epithelial cells are depicted: goblet cells and ciliated epithelial cells. The cilia on the epithelial cells together with the mucus produced by goblet cells clear the pathogen via mucociliary clearance. Epithelial cells can also secrete antimicrobial peptides that directly kill <i>S. pneumoniae</i> or produce cytokines, which leads to a state of inflammation and the recruitment of immune cells. . .	69

Figure 3.9: Toll-like receptors (TLRs) assist in the activation of adaptive immune cells. In this figure, TLR2 recognizes the <i>Streptococcus pneumoniae</i> 's lipoproteins. Upon activation, TLR2 secretes cytokines and co-stimulatory molecules. These co-stimulatory molecules are essential for co-stimulation and activation of T cells. The T cell is presented an antigen with major histocompatibility complex (MHC)II and antigen-presenting cell. The recognition of the antigen–MHCII complex and the co-stimulatory molecules activates the T cell and leads downstream to differentiation into Th1 and Th2 cells, that can release various cytokines such as interferon- gamma (IFN)- γ and interleukin (IL)-4	73
Figure 4.1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. Data were curated from Gene Expression Omnibus (GEO) and Array Express (AE). The PRISMA flow diagram shows the identification, screening, eligibility and inclusion of samples in our analysis.	93
Figure 4.2: Meta-analysis pipeline for Chronic Obstructive Pulmonary Disease. (A)Summary of workflow used for the meta-analysis, (B) Pre-processing steps used on the microarray data,(C) Data analysis post ANOVA, (D) post-hoc analysis steps using ANOVA results.	94
Figure 4.3: Visualizing batch effects introduced by using multiple studies in our meta-analysis. (A) PCA before and (B) PCA after batch effect correction with ComBat.	100
Figure 4.4: Highlighted Primary Immunodeficiency KEGG Pathway (hsa05340) with enriched genes from the ANOVA (BH-adjusted p-value < 0.05)[1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in COPD samples.	102
Figure 4.5: Highlighted Cytokine-cytokine receptor interaction KEGG Pathway (hsa04060) with enriched genes from the ANOVA (BH-adjusted p-value < 0.05) [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in COPD samples.	103
Figure 4.6: Enriched Reactome pathway-gene network from up-regulated disease genes in COPD subjects. The enrichment analysis was based on the 304 statistically significant differentially expressed genes filtered for effect size.	106
Figure 4.7: Enriched Reactome pathway-gene network from down-regulated disease genes in COPD subjects. The enrichment analysis was based on the 304 statistically significant differentially expressed genes filtered for effect size.	107

Figure 4.8: Heatmap of statistically significant interacting genes across disease states and smoking statuses. Difference in means calculated using control non-smokers as the baseline.	108
Figure 4.9: Heatmap of age effect on the statistically significant disease gene list. The enrichment analysis was based on the 304 statistically significant differentially expressed genes filtered for effect size. The clustered groups are color-coded, with the corresponding genes in each group listed in the table.	110
Figure 4.10: Trained logistic regression model can classify COPD and healthy profiles. (A)The logistic regression model trained on all the data achieves $87.0 \pm 3.0\%$ accuracy), with the (B) confusion matrix and (C) ROC curves indicating good performance overall, with AUC 0.979. Training with 10-fold cross validation gives an average accuracy of 84.2%, with the worst testing model shown in (D) and its ROC for (E) Controls and (F) COPD shown respectively, with an AUC of 0.882.	112
Figure 5.1: Meta-analysis Workflow to Assess Gene Expression Variation in Influenza Disease and Vaccination	124
Figure 5.2: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist.	127
Figure 5.3: Flowchart of Gene Filtering Steps for Influenza Meta-analysis.	131
Figure 5.4: Highlighted NF-Kappa B Signaling KEGG Pathway (hsa04040) with Enriched Genes from the LRT Analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Infected Subjects [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Infected Subjects. . . .	132
Figure 5.5: Highlighted NF-Kappa B Signaling KEGG Pathway (hsa04040) with Enriched Genes from the LRT Analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Vaccinated Subjects [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Vaccinated Subjects. . .	133
Figure 5.6: Highlighted Influenza A KEGG Pathway (hsa05164) with Enriched Genes from the LRT analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Infected Subjects [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Infected Subjects.	134
Figure 5.7: Highlighted Influenza A KEGG Pathway (hsa05164) with Enriched Genes from the LRT Analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Vaccinated Subjects [1–3] . Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Vaccinated Subjects.	135

Figure 5.8: Heatmap of Statistically Significant (Bonferroni-adjusted p-value <0.05) Genes with an Interaction Between Disease State and Age for Healthy Controls. Difference in means calculated by comparing control subjects in age groups 2-4 to control subjects in age group 1 (baseline).	141
Figure 5.9: Heatmap of Statistically Significant (Bonferroni-adjusted p-value <0.05) Genes with an Interaction Between Disease State and Age for Influenza Infected Subjects. (A) Difference in means calculated by comparing influenza infected subjects in age groups 2-4 to influenza infected subjects in age group 1 (baseline). (B) Comparison of influenza infected subjects to control subjects in the different age groups by calculating the difference between the baseline-adjusted means for influenza infected subjects (A) and control subjects (Figure 5.8).	142
Figure 5.10: Heatmap of Statistically Significant (Bonferroni-adjusted p-value <0.05) Genes with an Interaction Between Disease State and Age for Influenza Vaccinated Subjects. (A) Difference in means calculated by comparing influenza vaccinated subjects in age groups 2-4 to influenza vaccinated subjects in age group 1 (baseline). (B) Comparison of influenza vaccinated subjects to control subjects in the different age groups by calculating the difference between the baseline-adjusted means for influenza vaccinated subjects (A) and control subjects (Figure 5.8).	143
Figure 5.11: Heatmap of Statistically Significant (Bonferroni-adjusted p-value <0.05) Genes with an Interaction Between Disease State and Age for Influenza Vaccinated Subjects Compared to Influenza Infected Subjects. Comparison of baseline-adjusted means for influenza vaccinated subjects (Figure 5.10A) and influenza infected subjects (Figure 5.9A)	144
Figure A.1: Principal component analysis of the disease factor before (A) and after (B) batch correction with ComBat.	156
Figure A.2: Principal component analysis of the sex factor before (A) and after (B) batch effect correction with ComBat.	157
Figure A.3: Principal component analysis of the age group factor before (A) and after (B) batch effect correction with ComBat.	157
Figure A.4: Heatmap with gene clustering of the top 25 differentially expressed disease (control-AD) gene list.	158
Figure A.5: Reactome pathway analysis bar plot of enriched pathways and number of gene hits. Gene list: Genes that were down-regulated in Alzheimer's disease but up-regulated in healthy controls.	159

Figure A.6: Reactome pathway analysis bar plot of enriched pathways and number of gene hits. Gene list: Genes that were up-regulated in Alzheimer's disease but down-regulated in healthy controls.	160
Figure A.7: Gene Ontology (biological processes) network of differentially expressed genes by disease factor from BINGO in Cytoscape. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.	161
Figure A.8: Pathway-gene network of enriched Reactome pathways using the differentially expressed disease genes with a sex effect (prior to selecting for interacting genes) that were up-regulated in males	162
Figure A.9: Heatmap with gene clustering to visualize gene expression of differentially expressed disease (control-AD) gene list with a sex effect (prior to selecting for interacting genes).	163
Figure A.10: Heatmap with gene clustering to visualize age group effect (prior to selecting for interacting genes) using difference in means on the differentially expressed disease (control-AD) gene list.	164
Figure A.11: Heatmap with gene clustering to visualize tissue (hippocampus as baseline) effect using the difference in means (prior to selecting for interacting genes) on the differentially expressed disease (control-AD) gene list.	165
Figure A.12: Heatmap with gene clustering to visualize tissue (blood as baseline) effect using the differences in means between binary comparisons (prior to selecting for interacting genes) on the differentially expressed disease (control-AD) gene list	166
Figure B.1: Principal Component Analysis to visualize changes in variation in datasets before and after combat.	172
Figure B.2: Highlighted Pathways in Cancer KEGG Pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor) [1–3]	173
Figure B.3: Highlighted Lysosome KEGG Pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor). [1–3]	174
Figure B.4: Highlighted Adherens KEGG Pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor)[1–3]	175
Figure B.5: Highlighted Hematopoietic Cell Lineage KEGG pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor) [1–3] . . .	176

Figure B.6: Highlighted Measles KEGG pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor) [1–3]	177
Figure B.7: Enriched Reactome pathway-gene network using the differentially expressed disease genes with a sex effect (no significant interaction between sex and disease) that were up-regulated in males).	178
Figure B.8: Gene ontology results from BINGO using our 207 unique statistically significant disease genes filtered for biological effect. Our 304 biologically significant genes were compared to Reinhold et al., [5]	179
Figure C.1: Gene Ontology of Biologically Significant Genes for Influenza Infected Subjects using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.	180
Figure C.2: Gene Ontology of Biologically Significant Genes for Influenza Vaccinated Subjects using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.	181
Figure C.3: Gene Ontology of Biologically Significant Genes Only in the Influenza Infected Subjects Gene List using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.	182
Figure C.4: Gene Ontology of Biologically Significant Genes Only in the Influenza Vaccinated Subjects Gene List using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.	183
Figure C.5: Gene Ontology of Biologically Significant Common Genes for the Influenza Infected and Vaccinated Subjects using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.	184
Figure C.6: Heatmap of Biologically Significant Common Genes for the Influenza Infected and Vaccinated Subjects with an Interaction Between Disease State and Age. Comparison of baseline-adjusted means for influenza vaccinated subjects and influenza infected subjects	185
Figure C.7: Heatmap of Biologically Significant Genes Only in the Influenza Infected Gene List with an Interaction Between Disease State and Age. Comparison of baseline-adjusted means for influenza infected subjects and controls	186

Figure C.8: Heatmap of Biologically Significant Genes Only in the Influenza Vaccinated
Gene List with an Interaction Between Disease State and Age. Comparison
of baseline-adjusted means for influenza vaccinated subjects and controls 187

KEY TO ABBREVIATIONS

AATD alpha-1 antitrypsin deficiency

A β amyloid-beta

ABCA1 ATP Binding Cassette Subfamily A Member 1

ABCs Active Bacterial Core surveillance

AD Alzheimer's Disease

AE Array Express

AJC Apical junctional complex

ANOVA Analysis of variance

APC Antigen presenting cells

APH-1 Gamma-Secretase Subunit APH-1A

APOD Apolipoprotein D

APOE Apolipoprotein D

APP Amyloid Beta Precursor Protein

AUC Area under the curve

BACE Beta-Secretase

BCL11B BAF Chromatin Remodeling Complex Subunit BCL11B

BH-adjusted Benjamini-Hochberg adjusted

BINGO Biological Networks Gene Ontology tool

bp Base pairs

BTK Bruton tyrosine kinase

C11orf74 Chromosome 11 Open Reading Frame 74

C21orf91 Chromosome 21 Open Reading Frame 91

CACNG3 Calcium Voltage-Gated Channel Auxiliary Subunit Gamma 3

CAP Community acquired pneumonia

CAPN1 Calpain 1

CAPN2 Calpain 2
CAPN3 Calpain 3
CASP7 Caspase 7
CASP9 Caspase 9
CBP Choline binding protein
CBPC Choline binding protein C
CBPG Choline binding protein G
CCL2 C-C Motif Chemokine Ligand 2
CCL5 C-C Motif Chemokine Ligand 5
CCR6 C-C Motif Chemokine Receptor 5
CCR8 C-C Motif Chemokine Receptor 8
CD14 CD14 molecule
CD177 CD177 molecule
CD2 CD2 molecule
CD27 CD27 molecule
CD3E CD3E molecule
CD40 CD40 molecule
CD7 CD7 molecule
CD81 CD81 molecule
CD96 CD96 molecule
CDC Centers for Disease Control and Prevention
CEACAM6 Carcinoembryonic Antigen Related Cell Adhesion Molecule 6
CHRM1 Cholinergic Receptor Muscarinic 1
CLEC4D C-Type Lectin Domain Family 4 Member D
CLU Clusterin
COL21A1 Collagen Type XXI Alpha 1 Chain
COPD Chronic obstructive pulmonary disease

COX Cytochrome c oxidase

COX5B Cytochrome C Oxidase Subunit 5B

COX6A1 Cytochrome C Oxidase Subunit 6A1

COX6C Cytochrome C Oxidase Subunit 6C

CRP C-reactive protein

CT Computed tomography

CXCL12 C-X-C Motif Chemokine Ligand 12

CXCR3 C-X-C Motif Chemokine Receptor 3

CXCR4 C-X-C Motif Chemokine Receptor 4

CXCR6 C-X-C Motif Chemokine Receptor 6

CYBRD1 Cytochrome B Reductase 1

CYP1B1 Cytochrome P450 Family 1 Subfamily B Member 1

DALY Disability adjusted life year

DCLRE1C DNA Cross-Link Repair 1C

DDR2 Discoidin Domain Receptor Tyrosine Kinase 2

DEFA4 Defensin Alpha 4

DEG Differentially expressed genes

DIRAS2 DIRAS Family GTPase 2

DUSP7 Dual Specificity Phosphatase 7

ECM Extracellular matrix

EEF1A Eukaryotic Translation Elongation Factor 1 Alpha 1

EEF1A2 Eukaryotic Translation Elongation Factor 1 Alpha 2

EEF1E1 Eukaryotic Translation Elongation Factor 1 Epsilon 1

EFEMP2 EGF Containing Fibulin Extracellular Matrix Protein 2

EPS8 Epidermal Growth Factor Receptor Pathway Substrate 8

F5 Coagulation Factor V

FAM107B Family With Sequence Similarity 107 Member B

FAM13A Family With Sequence Similarity 13 Member A

Fas Fas Cell Surface Death Receptor

FLT3LG Fms Related Tyrosine Kinase 3 Ligand

GABA Gamma-Aminobutyric Acid

GABRA1 Gamma-Aminobutyric Acid Type A Receptor Alpha1 Subunit

GABRG2 Gamma-Aminobutyric Acid Type A Receptor Gamma2 Subunit

GAD1 Glutamate Decarboxylase 1

GAPDH Glyceraldehyde 3-phosphate dehydrogenase

GEO Gene Expression Omnibus

GJA1 Gap Junction Protein Alpha 1

GlcNac N-acetylglucosamine

GLRB Glycine Receptor Beta

GMPR Guanosine Monophosphate Reductase

GNAQ G Protein Subunit Alpha Q

GNG12 G Protein Subunit Gamma 12

GO Gene Ontology

GOLD Global Initiative for Chronic Obstructive Lung Disease

GPR15 G Protein-Coupled Receptor 15

GRIN1 Glutamate Ionotropic Receptor NMDA Type Subunit 1

GWAS Genome-wide association study

H2O2 Hydrogen peroxide

HA Hemagglutinin

HIP1 Huntingtin Interacting Protein 1

HK3 Hexokinase 3

HP Haptoglobin

HVCN1 Hydrogen Voltage Gated Channel 1

ICAM Intercellular Adhesion Molecule 1

ICAM2 Intercellular Adhesion Molecule 2

IFN Interferon

IFNG Interferon-gamma

IgA Immunoglobulin A

IKBKG Inhibitor Of Nuclear Factor Kappa B Kinase Regulatory Subunit Gamma

IL-18 Interleukin 18

IL-1 β Interleukin 1 beta

IL24 Interleukin 24

IL2RB Interleukin 2 Receptor Subunit Beta

IL2RG Interleukin 2 Receptor Subunit Gamma

IL6 Interleukin6

IL7R Interleukin 7 Receptor

IP6K3 Inositol Hexakisphosphate Kinase 3

IS Immune Space

ITGB8 Integrin Subunit Beta 8

ITIH5 Inter-Alpha-Trypsin Inhibitor Heavy Chain 5

ITPKB Inositol-Trisphosphate 3-Kinase B

ITPR3 Inositol 1,4,5-Trisphosphate Receptor Type 3

JAK3 Janus Kinase 3

KCNJ10 Potassium Voltage-Gated Channel Subfamily J Member 10

KCNJ16 Potassium Voltage-Gated Channel Subfamily J Member 16

KCNJ2 Potassium Voltage-Gated Channel Subfamily J Member 2

KCNQ2 Potassium Voltage-Gated Channel Subfamily Q Member 2

KEGG Kyoto Encyclopedia of Genes and Genomes

KLRB1 Killer Cell Lectin Like Receptor B1

KLRG1 Killer Cell Lectin Like Receptor G1

LAIV Live attenuated vaccine

LRI Lower respiratory tract infection

LRP LDL Receptor Related Protein

LRP1 LDL Receptor Related Protein 1

LRT Likelihood ratio tests

LTA Lipoteichoic acids

LTF Lactotransferrin

LytA Lytic Amidase

MAP3K1 Mitogen-Activated Protein Kinase Kinase Kinase 1

MAPK Mitogen-Activated Protein Kinase

MARCO Macrophage receptor with collagenous structure

MBL Mannose binding lectin

MET Mesenchymal Epithelial Transi- tion

MHC Major histocompatibility complex

MMP8 Matrix Metallopeptidase 8

MPO Myeloperoxidase

mRNA Messenger rna

MRPL1 Mitochondrial Ribosomal Protein L1

MRPL13 Mitochondrial Ribosomal Protein L13

MRPL15 Mitochondrial Ribosomal Protein L15

MRPL3 Mitochondrial Ribosomal Protein L3

MRPS18C Mitochondrial Ribosomal Protein S18C

MRPS28 Mitochondrial Ribosomal Protein S28

MS4A1 Membrane Spanning 4-Domains A1

MS4A3 Membrane Spanning 4-Domains A3

MurNac N- acetylmuramic

MYD88 Myeloid differentiation primary response 88

MYOM2 Myomesin 2

NA Neuraminidase

NAE1 NEDD8 Activating Enzyme E1 Subunit 1

NCALD Neurocalcin Delta

NCBI National Center for Biotechnology Information

NCSPS Non-classical Surface Proteins

NDUFC2 NADH:Ubiquinone Oxidoreductase Subunit C2

nec NormExp Background Correction

NEFL Neurofilament Light

NET Neutrophil extracellular traps

Nfkb Nuclear Factor Kappa B

NLR Nod-like receptors

NLRP3 NLR Family Pyrin Domain Containing 3

NOD2 Nucleotide Binding Oligomerization Domain Containing 2

NRXN3 Neurexin 3

OTOF Otoferlin

OXPHOS Oxidative phosphorylation

PAI Pathogenicity islands

PAMPs Pathogen associated molecular patterns

PARP1 Poly(ADP-Ribose) Polymerase 1

PavA Pneumococcal adherence and virulence factor A

PBMC Peripheral blood mononuclear cell

PCA Principal component analysis

Pcho Phosphorylcholine

PCR Polymerase chain reaction

PCV13 Pneumococcal conjugate vaccine 13

PiaA Pneumococcal iron acquisition A

PitA Pneumococcal iron transporter

PiuA Pneumococcal iron uptake A
PLCB1 Phospholipase C Beta 1
pm Perfect match
POU2AF1 POU Class 2 Homeobox Associating Factor 1
PPSV23 Pneumococcal polysaccharide vaccine 23
PRR Pattern Recognition Receptors
Psaa Pneumococcal surface antigen A
PSEN Presenilin
PSEN2 Presenilin 2
PSMA3 Proteasome Subunit Alpha 3
PSMC6 Proteasome 26S Subunit, ATPase 6
PspA Pneumococcal surface protein A
PspC Pneumococcal surface protein C
PSRP Pneumococcal serine-rich repeat protein
PTGDS Prostaglandin D2 Synthase
RFXAP Regulatory Factor X Associated Protein
RIG-I- Retinoic acid-inducible gene I
RMA Robust Multi-array Average
RNA-seq RNA sequencing
RNF135 Ring Finger Protein 135
ROC Receiver operating characteristic
ROS Reactive oxygen species
RPA3 Replication Protein A3
RRAS2 RAS Related 2
RSV Respiratory syncytial virus
S. pneumoniae *Streptococcus pneumoniae*
SAP Serum amyloid p

SASH1 SAM And SH3 Domain Containing 1

SDHA Succinate Dehydrogenase Complex Flavoprotein Subunit A

SELENOP Selenoprotein P

SEM1 SEM1 26S Proteasome Complex Subunit

SLC40A1 Solute Carrier Family 40 Member 1

SNAP91 Synaptosome Associated Protein 91

SSH1 Slingshot Protein Phosphatase 1

SST Somatostatin

STMN2 Stathmin 2

TANK TRAF Family Member Associated NFKB Activator

TCF7 Transcription Factor 7

TH17 T-helper cell 17

TIV Trivalent inactivated vaccine

TLR Toll-Like Receptors

TNF Tumor Necrosis Factor

TRAF6 TNF Receptor Associated Factor 6

Treg Regulatory T cells

TukeyHSD Tukey Honest Significant Difference Test

UGCG UDP-Glucose Ceramide Glucosyltransferase

US United States

WTA Wall teichoic acid

ZAP70 Zeta Chain Of T Cell Receptor Associated Protein Kinase 70

CHAPTER 1

OVERVIEW AND AIMS OF DISSERTATION

1.1 Overview

Aging is a natural part of life, and it refers to the physiological changes within the human body from birth to death [6]. With age, the body gradually deteriorates and cell and organ function becomes impaired. These physiological changes include cellular senescence, mitochondrial dysfunction, genomic instability and altered intercellular communication [7]. More specifically, changes such as the decline in immune system function and also cognitive decline with age puts the elderly at a higher risk for contracting diseases. For example, neurodegenerative diseases such as Alzheimer's Disease (AD) and respiratory infections are highly prevalent in the elderly population [8, 9]. Furthermore, with the elderly population living longer, fully understanding how gene expression changes in response to diseases can assist in treating and reducing the disease burden for this population.

An underdeveloped immune system, seen in infants and young children, also causes an increased risk for infectious diseases such as pneumonia and influenza [10]. In the case of respiratory diseases, the literature indicates that there is an age-dependency in immune response to viral and bacterial respiratory infections [10, 11]. Understanding the role that age plays in host immune system activation is essential for better prognosis and treatment of diseases. In addition, the Centers for Disease Control (CDC) have specific vaccine recommendations for different age groups [12, 13]. These recommendations have been made based on the results of previous studies that indicate the existence of an age dependency in immune response to respiratory infections such as influenza and pneumonia.

Another respiratory condition caused by the exposure to toxic substances (tobacco smoke) is chronic obstructive pulmonary disease (COPD). The main risk factor for COPD is tobacco exposure, and the age of onset is around 40 years old[14]. Tobacco exposure causes inflammation and lung damage, and because of this, COPD sufferers are at higher risk for other respiratory infections.

Differentiating between normal changes in gene expression due to aging and age-dependent immune responses to diseases will help solidify our understanding of the relationship between age and immune system activation. This dissertation investigates age-dependent diseases and assesses

gene expression variation due to sample characteristics such as age and sex. We focus on the neurodegenerative disease, AD and also the respiratory diseases: pneumococcal disease, COPD and influenza.

1.2 Aims of Dissertation

In Chapter 2, we explore AD and the effects of age, sex and tissue type on gene expression. Chapter 2, describes a meta-analysis approach to highlight statistically significant differentially expressed genes in AD, determine biologically significant genes and also highlight disease genes that interact directly with age, sex and tissue type. Chapter 2 also explores pathways associated with statistically significant genes to determine what processes are affected by gene expression variation due to the study factors.

Chapter 3 combines the literature to assess pneumococcal diseases, host defenses and how aging impairs the host's ability to clear the pathogen. We also elaborate on current diagnostics methods, treatments and preventative methods available and how efficacy changes with age. In Chapter 4, we evaluate gene expression changes in COPD by looking at factors such as disease state, age, sex and smoking status. COPD being a respiratory disease mainly caused by tobacco exposure, assessing how gene expressing changes due to smoking status provides insight on disease pathology. In the last data chapter, Chapter 5, we investigate gene expression differences in influenza infection and vaccination. Exploring temporal patterns for these two disease states highlights genes that are unique to each disease state, and also what genes are in common. These results not only highlight how the immune response to influenza infection differs from influenza vaccination, but also looks at how aging and sex differences can also affect gene expression toward influenza.

Together, the chapters of the dissertation aim to improve the understanding of age-related/age-dependent diseases and how physiological changes in the immune system and within the brain can affect disease susceptibility. We identify gene expression differences across diseases including full age, sex, smoking status and tissue type considerations. Our findings provide potential gene and pathway associations that can be targeted to improve treatment and prevention of these diseases.

CHAPTER 2

DATA-DRIVEN ANALYSIS OF AGE, SEX, AND TISSUE EFFECTS ON GENE EXPRESSION VARIABILITY IN ALZHEIMER'S DISEASE

Work presented in this chapter has been published as Brooks LRK, Mias GI. Data-Driven Analysis of Age, Sex, and Tissue Effects on Gene Expression Variability in Alzheimer's Disease. *Frontiers in neuroscience*. 2019;13:392.

2.1 Abstract

Alzheimer's disease (AD) has been categorized by the Centers for Disease Control and Prevention (CDC) as the 6th leading cause of death in the United States. AD is a significant health-care burden because of its increased occurrence (specifically in the elderly population), and the lack of effective treatments and preventive methods. With an increase in life expectancy, the CDC expects AD cases to rise to 15 million by 2060. Aging has been previously associated with susceptibility to AD, and there are ongoing efforts to effectively differentiate between normal and AD age-related brain degeneration and memory loss. AD targets neuronal function and can cause neuronal loss due to the buildup of amyloid-beta plaques and intracellular neurofibrillary tangles.

Our study aims to identify temporal changes within gene expression profiles of healthy controls and AD subjects. We conducted a meta-analysis using publicly available microarray expression data from AD and healthy cohorts. For our meta-analysis, we selected datasets that reported donor age and gender, and used Affymetrix and Illumina microarray platforms (8 datasets, 2,088 samples). Raw microarray expression data were re-analyzed, and normalized across arrays. We then performed an analysis of variance, using a linear model that incorporated age, tissue type, sex, and disease state as effects, as well as study to account for batch effects, and included binary interactions between factors. Our results identified 3,735 statistically significant (Bonferroni adjusted $p < 0.05$) gene expression differences between AD and healthy controls, which we filtered for biological effect (10% two-tailed quantiles of mean differences between groups) to obtain 352 genes. Interesting pathways identified as enriched comprised of neurodegenerative diseases pathways (including AD), and also mitochondrial translation and dysfunction, synaptic vesicle cycle and GABAergic synapse, and gene ontology terms enrichment in neuronal system, transmission across chemical synapses and mitochondrial translation.

Overall our approach allowed us to effectively combine multiple available microarray datasets and identify gene expression differences between AD and healthy individuals including full age and tissue type considerations. Our findings provide potential gene and pathway associations that can be targeted to improve AD diagnostics and potentially treatment or prevention.

2.2 Introduction

Aging refers to the physiological changes that occur within the body overtime [7]. These changes are accompanied by deteriorating cell and organ function due to cellular and immune senescence and DNA and protein damage [7, 15, 16]. Aging causes an increased risk for diseases. Age-related diseases are becoming a public health concern due to an overall increase in the older population and the average human life span in developed countries [17, 18]. It is predicted that by the year 2050, the number of Americans over 85 years of age will triple from 2015 [19, 20]. Larger percentages of the elderly and their increased risk for diseases can affect the economy, and social and health care costs [21]. For instance, immune system dysfunction and cognitive decline due to aging increases the risk of neurodegenerative diseases such as Alzheimer's disease (AD) [22, 23]. Previous research explored brain aging and found notable changes in brain size, brain structure and function [24]. Changes in the brain as we age are also known as hallmarks of brain aging. These hallmarks include: mitochondrial dysfunction, damage to proteins and DNA due to oxidation, neuroinflammation due to immune system dysfunction, reduction in brain volume size and gray and white matter, and impaired regulation of neuronal Ca^{2+} [23, 24]. These alterations render the aging brain vulnerable to neurodegenerative diseases such as AD.

AD, the most common form of dementia, is currently the 6th leading cause of death [25] in the United States (US). In 2010, an estimate of 4.7 million people in the US had AD, and the number of AD patients is expected to increase to 13.8 million in 2050 and to 15 million by 2060 [26–28]. As with other age-related diseases, the risk of AD increases with age. AD is currently characterized by the accumulation of amyloid-beta ($\text{A}\beta$) plaques and neurofibrillary tangles due to tau protein modifications [29]. These two protein changes are the main pathological changes in AD [29]. $\text{A}\beta$ is formed when the amyloid precursor protein (APP) is cleaved by γ -secretases and β -secretases. Cleavage of APP forms fragments of $\text{A}\beta$ which aggregate and deposit on neurons as plaques, which causes neuronal death in conjunction with neurofibrillary tangles [29].

While AD's prevalence is on the rise due to increased life expectancy, there is still no treatment available and diagnosis of AD is challenging. How AD progresses is still not completely understood

[30]. New technologies are available such as positron-emission tomography (PET) imaging and monitoring levels of $A\beta$ and tau in cerebrospinal fluid [29]. Co-morbidities that can exist due to aging such as hippocampal sclerosis further complicate AD diagnosis [31]. Furthermore, questions have been raised regarding whether or not AD is simply an accelerated form of aging due to them both being associated with changes in cognition [31]. However, studies have identified clear neurocognitive differences in cognition, brain size and function in AD compared to healthy aged subjects. For example, AD patients have more grey matter loss compared to white matter, impaired verbal and semantic abilities and more intense memory dysfunction compared to healthy seniors [31].

Pathological changes within the brain are observed prior to clinical diagnosis of AD. In most cases AD cannot be confirmed until postmortem examination of the brain. Researchers are investigating novel biomarkers to detect for earlier diagnosis before diseased individuals become functionally impaired. Meta-analysis of microarray datasets is becoming more popular for it provides stronger power to studies due to larger sample sizes obtained through statistically combining multiple datasets. Microarray data are also available in large quantities on public online data repositories. In the case of AD, Winkler et al., performed a meta-analysis that compared neurons within the hippocampus of AD patients and healthy controls. They identified that processes such as apoptosis, and protein synthesis, were affected by AD and were regulated by androgen and estrogen receptors [32]. Researchers have also explored differences in gene expression in Parkinson's and AD subjects via a meta-analysis approach [33], and identified functionally enriched genes and pathways that showed overlap between the two diseases [33]. Most recently, Moradifard et al. identified differentially expressed microRNAs and genes when comparing AD to healthy controls via a meta-analysis approach. They also identified two key microRNAs that act as regulators in the AD gene network[34].

In our investigation, our goal was to identify age, sex, and tissue effects on gene expression variability in AD by comparing age-matched healthy controls to AD subjects via a meta-analysis approach. In this data-driven approach, we explored global gene expression changes in 2,088 total

samples (771 healthy, 868 AD , and 449 possible AD, curated from 8 studies) from 26 different tissues, to identify genes and pathways of interest in AD that can be affected by factors such as age, sex and tissue. Our findings provide potential gene and pathway associations that can be targeted to improve AD diagnostics and potentially treatment or prevention.

2.3 Methods

We conducted a meta-analysis using 8 publicly available microarray expression datasets (Table 2.1) from varying tissues and microarray platforms on AD. We developed a thorough computational pipeline (Figure 2.1A) that involved curating and downloading raw microarray expression data, pre-processing the raw expression data and conducting a linear model analysis of the gene expression profiles. Statistically different genes based on disease state were identified following analysis of variance (ANOVA) on the linear model which compared gene expression changes due to disease state, sex, age and tissue. These genes were further analyzed using a Tukey Honest Significant Difference (TukeyHSD) test to determine their biological significance [35]. In addition to the p-values, we also obtained the mean differences between binary comparisons of groups (also generated by the TukeyHSD), as a measure of biological effect size. We examined the TukeyHSD results by filtering by each factor, and identified up and down regulated genes. We then selected genes that showed statistically significant pairwise interactions between disease status and sex, age and tissue. Using these genes, we used R packages ReactomePA [36] and clusterProfiler [37] to conduct gene enrichment and pathway analyses of the differentially expressed genes (DEG). We used BINGO in Cytoscape v.3.7.0 for gene ontology (GO) analysis on each gene set for each factor [38, 39].

Database	Accession Number	Controls	AD	Possible AD	Platform	Citation
GEO	GSE84422	242	362	449	Affymetrix Human Genome U133A, B and Plus 2.0	[40]
GEO	GSE28146	8	22	-	Affymetrix Human Genome Plus 2.0	[41]
GEO	GSE48350	173	80	-	Affymetrix Human Genome Plus 2.0	[42]
GEO	GSE5281	74	85	-	Affymetrix Human Genome Plus 2.0	[43]
GEO	GSE63060	104	142	-	Illumina HumanHT-12 V3.0 expression beadchip	[44]
GEO	GSE63061	134	139	-	Illumina HumanHT-12 V4.0 expression beadchip	[44]
GEO	GSE29378	32	31	-	Illumina HumanHT-12 V3.0 expression beadchip	[45]
Array Express	E-MEXP-2280	5	7	-	Affymetrix Human Genome Plus 2.0	[46]

Table 2.1: Curated microarray datasets and the study description.

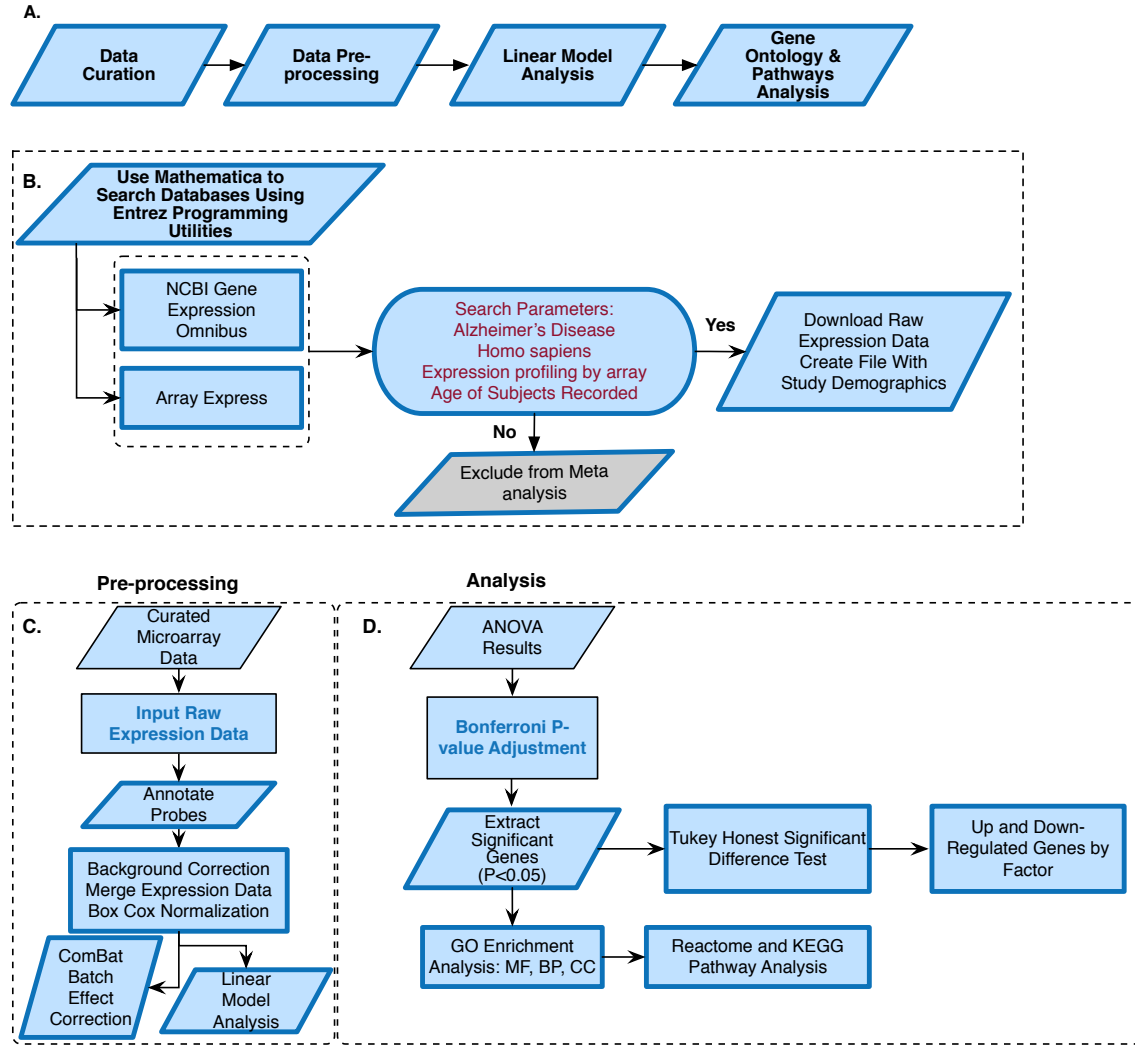


Figure 2.1: Alzheimer's disease meta-analysis framework. (A) Simplified workflow used for the meta-analysis, (B) Pipeline for curating microarray data, (C) Pipeline for pre-processing the microarray data, (D) Methods used for meta-analysis of raw expression microarray data.

2.3.1 Microarray Data Curation

We curated microarray expression data from two data repositories: National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [47] and Array Express [48] (Figure 2.1B). We searched these repositories by using entrez programming utilities in Mathematica [49, 50]. In this search, we used the following keywords: *Homo sapiens*, Alzheimer's Disease and expression profiling by array (Figure 2.1B). This search resulted in 105 datasets from GEO

and 8 from Array Express. We further filtered the search results by excluding data from cell lines, selecting for expression data from Illumina and Affymetrix microarray platforms, and focusing on datasets that provided the ages and sex of their samples (Figure 2.1B). After filtering through the databases, we found 7 datasets from GEO (GSE84422, GSE28146, GSE48350, GSE5281, GSE63060, GSE63061, GSE29378) and 1 dataset from Array Express (E-MEXP-2280) to conduct our meta-analysis of expression profiling to assess differences in gene expression due to disease state, sex, age and tissue (Table 2.1). The majority of samples from AD subjects were collected post-mortem, from a variety of brain banks, while the subjects from GSE63060 and GSE63061 voluntarily gave blood samples (Table A.1 of Appendix A). The criteria and guidelines followed for diagnosis and sampling varied across datasets (Table A.1 of Appendix A). Additionally, we downloaded the raw expression data from each dataset, and created a demographics file per study, which included characteristics about the samples (Table 2.2). Our demographics file included information about the subjects that was reported in all datasets. For example, some studies reported the type of AD diagnosis for their respective subjects, as well as the Braak stage and APOE genotype, whereas others did not (Table A.1 of Appendix A). Therefore, to ensure uniform annotation of the subjects, we re-annotated subject information provided from the databases: For GSE28146, we grouped the sub-types of AD, incipient, moderate and severe, as AD because we did not have such classification information for our other AD samples. We changed all the GSE29378 tissue types to hippocampus, relabeled the "probable AD" disease state to "possible AD" in GSE84422, only used AD and control subjects from the E-MEXP-2280 and GSM238944 with an age of >90 (not a definite age) was removed from GSE5281. We should note also that the 1,053 samples from the GSE84422 dataset included different tissues from the same subjects, which were treated independently - a paired-design was not incorporated in our downstream analysis.

2.3.2 Pre-processing and Data Normalization

We downloaded the raw expression data from the data repositories in Mathematica [50] and pre-processed each file in R [51] using the appropriate R packages based on the microarray platform.

Accession Number	Sex (M/F)	Age Range
GSE84422	302M/166F	60-103
GSE28146	12M/18F	65-101
GSE48350	124M/129F	20-99
GSE5281	102M/56F	63-102
GSE63060	88M/158F	52-88
GSE63061	107M/166F	59-95
GSE29378	38M/25F	61-90
E-MEXP-2280	7M/5F	68-82

Table 2.2: Patient characteristics for curated datasets.

The `affy` package was used to pre-process all the .CEL data files from Affymetrix [52], and the `limma` package for Illumina summary data files [53]. We performed background correction, normalization and annotated and summarized all probes (Figure 2.1C). For the Affymetrix expression data files, we used the `expresso` function with the following parameters: robust multi-array analysis (RMA) for background correction, perfect-match (PM) adjustment to correct the perfect match probes, and ‘`avdiff`’ for the summary method to compute expression values [52]. We also used the `avereps` function from `limma` to summarize probes and remove replicates [53]. For the Illumina expression data, we corrected the background using the `NormExp Background Correction (nec)` function from the `limma` package for datasets where the detection p-values were reported, we annotated and used the `aggregate` function from the `stats` package in base R to summarize probes [51, 53]. We merged all 8 datasets into one large matrix file via common gene symbols. After merging the datasets, we performed a BoxCox power transformation [54] using the `ApplyBoxCoxTransform` function and data standardization using the `StandardizeExtended` function from the `MathIOmica` package [49, 55] (Figure 2.1C and also see ST2 of online supplemental data (Appendix A)).

2.3.3 Visualizing Variation due to Batch Effects

Merging expression data from different studies, array platforms and tissues can introduce confounding factors and manipulate interpretation of results. To address this, and assess whether

batch effects were evident and could be accounted for, we used the ComBat function in the sva package in R [56, 57] to adjust data for known batch effects . In this study, the batch effect was the study (i.e. different experiments/research groups), and we also found that there was a one-to-one correspondence between study and platform. Using expression data from prior to and post ComBat corrections, we used principal component analysis (PCA) plots to visualize the variability in the data and the effectiveness of possible batch effect removal [58].

2.3.4 Analysis of Variance

We modeled the merged expression data (see model breakdown below) prior to running ANOVA (using the anova and aov functions from the stats package in base R) to analyze differences among the different study factors (Figure 2.1D) [59]. We defined age group, sex, disease state, study and tissue as factors.

$$x \sim \sum_i x_i + \sum_{i,j;j>i} x_i : x_j \quad (2.1)$$

where $x_i \in \{\text{age group, sex, tissue, disease status}\}$ and the factors have the following levels:

- disease status = {control, possible AD, AD}
- sex = {male, female}
- age group = {under 60, 60-65, 65-70, 70-75, 75-80, 80-85, 85-90, 90-95, over 95}
- tissue = {amygdala, anterior cingulate, blood, caudate nucleus, dorsolateral prefrontal cortex, entorhinal cortex, frontal pole, hippocampus, inferior frontal gyrus, inferior temporal gyrus, medial temporal lobe, middle temporal gyrus, nucleus accumbens, occipital visual cortex, parahippocampal gyrus, posterior cingulate cortex, precentral gyrus, prefrontal cortex, primary visual cortex, putamen, superior frontal gyrus, superior parietal lobule, superior temporal gyrus, temporal pole}
- study = {GSE84422, GSE28146, GSE48350, GSE5281, GSE63060, GSE63061, GSE29378, E-MEXP-2280}

The p-values following the ANOVA were adjusted using Bonferroni correction for multiple hypothesis testing [59]. Genes with p-values <0.05 were considered statistically significant. We found statistically significant disease genes by filtering on the disease status for p-values <0.05 . Additionally, we used the `enrichKEGG` function in the `clusterProfiler` package in R for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis on these genes [1, 37]. We also performed Reactome pathway analysis with the `enrichPathway` function in the `ReactomePA` package in R [36]. These packages adjust p-values using the Benjamini Hochberg method for False Discovery Rate (FDR) control. Enriched pathways with adjusted p-value <0.05 were considered statistically significant [36, 37] (see ST5 and ST6 of online supplemental data (Appendix A)).

2.3.5 Identifying Up and Down Regulated Genes by Factor

To identify which of the 3,735 genes that show biologically significant differences, we conducted a TukeyHSD (using the `TukeyHSD` function from the `stats` package in base R) to determine statistically significant up and down-regulated genes using the difference in the means of pairwise comparisons between the levels within each factor [35, 60]. We carried out TukeyHSD testing on the statistically significant disease genes we obtained from the ANOVA. To account for multiple hypothesis testing in the TukeyHSD results, we used <0.00013 ($0.05/\text{number of genes ran through TukeyHSD}$) as a Bonferroni adjusted cutoff for statistical significance.

We selected the TukeyHSD results from the disease status factor, and focused on the "Control-AD" pairwise comparison to assess statistically significant gene expression differences. To assess biological effect, and select an appropriate fold-change-like cutoff (as our results had already been transformed using a Box-Cox transformation), we calculated the quantiles based on the TukeyHSD difference of mean difference values (Table A.2 of Appendix A). We used a two-tailed 10% and 90% quantile to identify significantly up and down regulated genes (Table A.2 of Appendix A).

The DEG by disease status factor were subsequently used to determine whether or not there was a sex, age or tissue effect on them. For sex, we used the DEG to filter the TukeyHSD results for sex factor differences, identified statistically significant sex-relevant genes based on p-value cutoff, and

the computed 10% and 90% quantiles based on the difference of means between male and female groups. We repeated the above steps for age group, but focused only on the binary comparisons where all age groups were compared to the <60 age group, which was used as a baseline (i.e. computed the mean gene expression differences per group comparison, i -<60, where i stands for any age group). This was carried out to enable us to compare the progression with age, relative to a common reference across all age groups. As for tissue, we carried out the same steps as above to determine DEG based on comparisons both a hippocampus-based baseline, as well a blood-based baseline.

Following the identification of the DEG by disease status and sex, we visualized the raw expression data for these genes in heatmaps. In addition to this, we generated heatmaps using the difference of means values (TukeyHSD) for the identified DEG by age group (<60 baseline) and tissue (hippocampus and blood as baseline).

To further investigate the significance of pairwise interactions with disease status and the factors sex, age and tissue, we used the identified statistically significant (p-value <0.00013, two-tailed 10% and 90% quantile) genes from our post-hoc analysis for each factor, and filtered our ANOVA results for statistically significant interactions (Bonferroni corrected p-value < 0.05, see also ST4 of online supplemental data (Appendix A)).

2.3.6 Gene Ontology and Reactome Pathway Analysis

For the disease and sex DEG sets, we used the R package `ReactomePA` to find enriched pathways[36]. We also built networks to determine if genes overlapped across pathways. Additionally, we used BINGO in Cytoscape for GO analysis to determine the biological processes the genes were enriched in [38]. Results were considered statistically significant based on Benjamini-Hochberg adjusted p-value <0.05.

2.4 Results

With our data selection criteria outlined in Figure 2.1B we identified 8 datasets from GEO and Array Express to conduct our meta-analysis to assess differences in gene expression due to disease state, sex, age and tissue (Table 2.1). We merged the processed expression data by common gene names, which gave us a total of 2,088 samples and 16,257 genes. The 2,088 samples consisted of 771 healthy controls, 868 AD subjects, 449 subjects reported as possibly having AD, 1308 females and 780 males.

2.4.1 ComBat Batch Effect Visualization

Combining data from different platforms, tissues and different laboratories introduces batch effects. Batch effects are sources of non-biological variations that can affect conclusions. We used the ComBat algorithm in R which works by adjusting the data based on a known batch effect. For our analysis we classified the study variable as our batch (the study and type of platform are directly related). We used PCA to visualize variation in the merged expression data before and after ComBat. In Figure 2.2 before correcting for batch effects, the datasets separate into 4 main clusters with a variance of 54.3% in PC1 and 13% in PC2. Following ComBat, those main clusters appear to be removed, with an overall reduction in variation for both principal components. We also looked at how the data separated by factor. In Figure 2.2B, there are two clear groups and this separation is accounted for when we look at the separation in the data by tissue (Figure 2.3). In Figure 2.3, before correction the 4 groups observed in Figure 2.2 are still evident. Following ComBat, the tissues: amygdala and nucleus accumbens cluster together in one group while all other tissues are in another. Batch effect correction with ComBat was solely used for visualizing how the expression data separates before and after ComBat correction - i.e. the batch corrected expression data were not used in the downstream analysis. We instead used a linear model to account for confounding study effects. Visualizing and understanding the variation within the expression data following the merge confirmed the need to include the study as a factor in the linear model analysis.

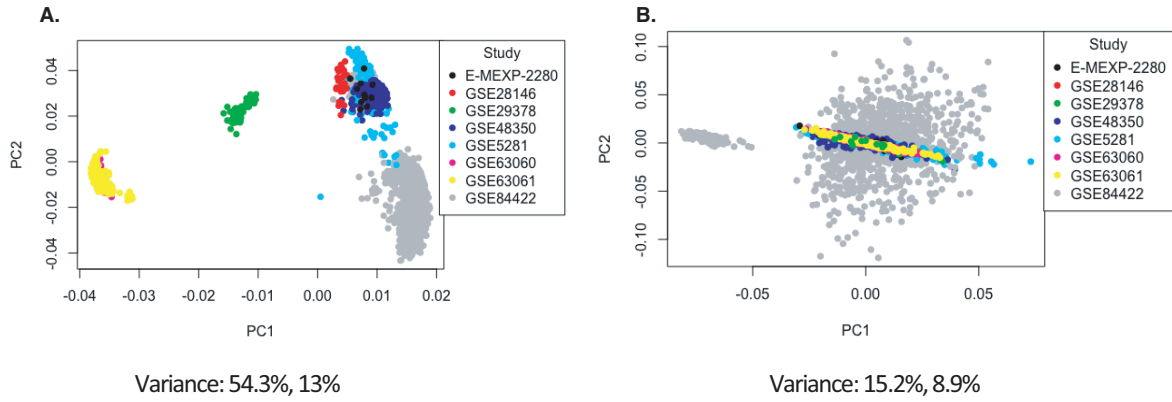


Figure 2.2: Principal component analysis of the study factor before and after batch correction with ComBat.

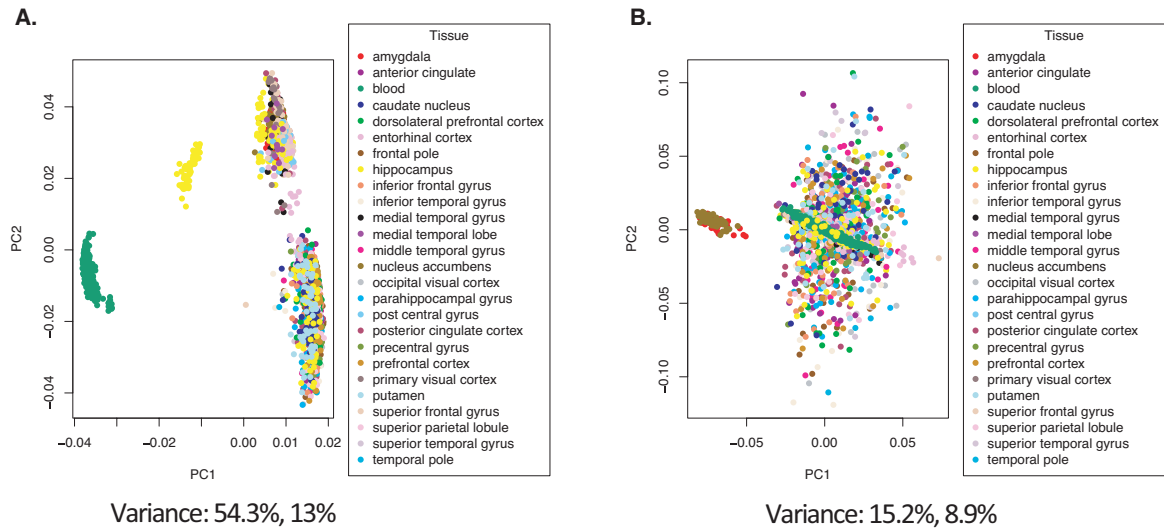


Figure 2.3: Principal component analysis of the tissue factor before and after batch correction with ComBat.

2.4.2 Analysis of Variance on Gene Expression By Disease State

Using ANOVA we assessed the variance in gene expression across the different factors in our linear model by including the following factors and their pairwise interactions: age group, study, tissue, sex

and disease state [59]. Statistically significant gene expression differences were determined using a Bonferroni([61] adjusted p-value was (<0.05) [59, 60]. With our focus on differences by disease status, we filtered genes based on the ANOVA adjusted p-values for the disease factor. Selecting for statistical significance by disease status we found 3,735 genes (see ST4 of online supplemental data (Appendix A)). We conducted GO and pathway analysis on these genes. The KEGG pathway analysis results are displayed in Table 2.3 (see ST5 of online supplemental data for full table (Appendix A)). The analysis showed that the genes are involved in Reactome pathways such as the Mitochondrial Translation Initiation (55 gene hits), Signaling by the B Cell Receptor (61 gene hits), Activation of NF-kappa β in B cells (40 gene hits), Transmission across Chemical Synapses (83 gene hits) and Neuronal System (119 gene hits) (see ST6 of online supplemental data (Appendix A)). The KEGG pathways that were enriched for this gene set included neurodegenerative disease pathways such as Alzheimer's (31 gene hits), Huntington's (76 gene hits) and Parkinson's (53 gene hits) (Table 2.3) Pathways. We also had genes enriched in synaptic pathways including Synaptic vesicle cycle (30 gene hits), Dopaminergic synapse (48 gene hits) and GABAergic synapse (34 gene hits) (Table 2.3). In addition to synapses and neurodegeneration, the long term potentiation (23 gene hits) pathway was associated with these genes (see ST5 for full KEGG pathway analysis results). To further explore the enriched genes in the KEGG AD pathway, we used the TukeyHSD results to determine whether genes were up- or down- regulated (see ST7 of online supplemental data (Appendix A)). To further assess the 73 gene hits identified in the enriched AD pathway we computed their mean differences between AD and control subjects, and used MathIOmica [55] tools to highlight them in the AD pathway (Figure 2.4) [1–3, 49] (see ST7 on online supplement data for full table with difference of means (Appendix A)). For instance, the APOE and LRP gene were both found to be up-regulated in AD subjects compared to healthy controls, and in the KEGG AD pathway these genes are involved in A β aggregation (Figure 2.4).

ID	Description	p-value	p-adjusted value	# of hits
hsa03050	Proteasome	1.55E-11	4.78E-09	31
hsa04723	Retrograde endocannabinoid signaling	3.46E-10	4.78E-08	66
hsa05010	Alzheimer's disease	4.64E-10	4.78E-08	73
hsa00190	Oxidative phosphorylation	3.85E-09	2.98E-07	59
hsa05016	Huntington's disease	1.60E-08	9.90E-07	76
hsa04714	Thermogenesis	2.54E-08	1.31E-06	86
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	2.98E-06	1.32E-04	57
hsa04721	Synaptic vesicle cycle	4.57E-06	1.77E-04	30
hsa05012	Parkinson's disease	1.51E-05	5.18E-04	53
hsa04728	Dopaminergic synapse	6.48E-05	0.002003299	48
hsa04724	Glutamatergic synapse	1.58E-04	0.004085366	42
hsa05169	Epstein-Barr virus infection	1.59E-04	0.004085366	66
hsa04720	Long-term potentiation	1.73E-04	0.004119762	28
hsa04727	GABAergic synapse	2.31E-04	0.00506623	34
hsa01200	Carbon metabolism	2.46E-04	0.00506623	42
hsa01521	EGFR tyrosine kinase inhibitor resistance	3.12E-04	0.006031187	31
hsa04725	Cholinergic synapse	4.73E-04	0.008596289	40
hsa00270	Cysteine and methionine metabolism	5.56E-04	0.009547497	20
hsa04911	Insulin secretion	5.99E-04	0.009738112	32
hsa04713	Circadian entrainment	6.78E-04	0.01048273	35
hsa05033	Nicotine addiction	8.70E-04	0.012730978	18
hsa00650	Butanoate metabolism	9.06E-04	0.012730978	14
hsa03010	Ribosome	0.0010736	0.014423588	50
hsa04510	Focal adhesion	0.001159439	0.014927779	62
hsa04390	Hippo signaling pathway	0.001260878	0.015584456	50

Table 2.3: Top 25 KEGG Pathways using differentially expressed genes.

2.4.3 Up and Down- Regulated Gene Expression in AD and Sex Specific Differences

We conducted a post-hoc analysis (TukeyHSD) on the 3,735 statistically significant disease genes to identify factorial differences and explore up- and down- regulation of genes. We were particularly interested in the control compared to AD gene expression differences, and how these could be further sub-categorized to explore effects by sex, age and tissue. We used a Bonferroni adjusted p-value cut off for significance (<0.000013) and the 10% two-tailed quantile to determine significantly up

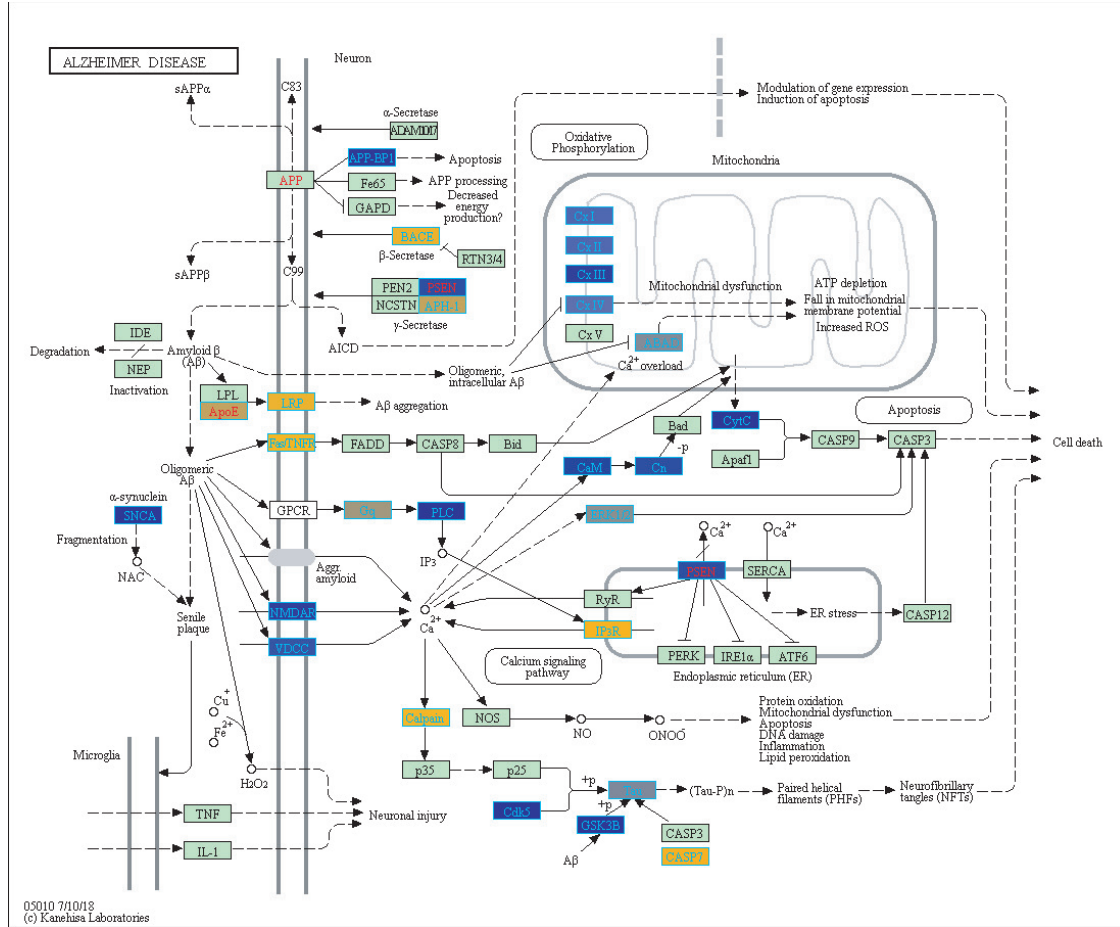


Figure 2.4: Enriched genes from the ANOVA statistically significant disease status gene list (p -value < 0.05) found in the KEGG Alzheimer's disease pathway (hsa05010) [[1–3]]. The yellow shading represents up-regulated and the blue shading represents down-regulated in AD samples. These genes were not yet filtered for biological significance.

and down regulated genes (Table A.2 of Appendix A). In the Control-AD TukeyHSD comparisons, we found 352 statistically significant genes that we classified as up-regulated (176 DEG) and down-regulated (176 DEG) in AD subjects (or correspondingly up or down-regulated in controls) if their mean differences were ≤ -0.0945 and ≥ 0.1196 respectively (Appendix A Table A.2, see also ST8 of online supplemental data (Appendix A)). The top 25 up- and down-regulated genes sorted by the TukeyHSD adjusted p -values are outlined in Table 2.4 (Figure A.4 of Appendix A and ST8 of online supplemental data (Appendix A)). After performing gene enrichment and pathway analysis with the ReactomePA R package [36] on the 352 genes we built pathway-gene networks for the statistically

significant Reactome pathways (Benjamini-Hochberg adjusted p-value < 0.05) (see ST13 and ST14 of online supplemental data (Appendix A)). Some of the top 10 enriched Reactome pathways from DEG down-regulated in AD include: Mitochondrial translation elongation, Mitochondrial translation, Transmission across chemical synapses, neuronal system (Figure 2.5 and Figure A.5 of Appendix A). The network in Figure 2.5 illustrates that some genes overlap across pathways - the difference of means from the TukeyHSD results of these genes are indicated by the color scale. The up-regulated genes in AD were enriched in pathways such as Extracellular matrix (ECM) organization and ECM proteoglycans, Non-integrin membrane-ECM interactions and potassium channel activation (Figure 2.6 and Figure A.6 of Appendix A). Additionally, we used BINGO for GO analysis on the 352 disease DEG to determine the biological processes they are involved in (Figure A.7 of Appendix A). Some examples of significant terms: Cell signaling development, nervous system development, neuron differentiation, cell proliferation, response to chemical stimulus, cell communication and brain and nervous system development (Figure A.7 of Appendix A).

Of the 352 DEG in the above disease analysis, 46 genes were differentially expressed by sex: 23 down- and 23 up- regulated in males compared to females based on mean differences (≤ -0.0864 and ≥ 0.2502 respectively (Table A.2 of Appendix A). We used the ReactomePA package to build a network of enriched genes and pathways with sex differences (Figure A.8 of Appendix A) [36]. We found 6 pathways that were enriched with the up-regulated gene list in males: Neuronal System, Transmission across chemical synapses, neurotransmitter receptors and post-synaptic signal transmission, and GABA A receptor activation (Figure A.8 of Appendix A and see also ST9 of online supplemental data (Appendix A)). Of these 46 genes that were differentially expressed by sex (Figure A.9 of Appendix A), we further filtered the ANOVA results to identify which of these genes showed statistically significant interactions with disease (sex:disease, Bonferroni corrected p-value < 0.05). We found one gene, chemokine receptor type 4 (CXCR4), to have a statistically significant pairwise interaction between disease status and sex (see ST4 of online supplemental data (Appendix A)).

Up-Regulated		Down-Regulated	
Gene	Difference of Means	Gene	Difference of Means
ITPKB	0.1709575	RPA3	-0.1781622
ARHGEF40	0.1574220	NME1	-0.1755078
CXCR4	0.1907433	LSM3	-0.1527917
PRELP	0.1319160	MRPL3	-0.1577078
SLC7A2	0.1568425	PTRH2	-0.1205413
AHNAK	0.1304494	RGS7	-0.1778522
NOTCH1	0.1014441	GLRX	-0.1622333
GFAP	0.1198343	RPH3A	-0.2168597
HVCN1	0.1151989	BEX4	-0.1416335
LDLRAD3	0.1627433	COX7B	-0.1726039
KANK1	0.0992824	NRN1	-0.1634702
HIPK2	0.1255059	PPEF1	-0.1430548
SLC6A12	0.1485253	PCSK1	-0.3127961
KLF4	0.1870071	ENY2	-0.1496523
ABCA1	0.1386346	CD200	-0.1537059
DDR2	0.1069751	NRXN3	-0.1203814
KLF2	0.1070143	GTF2B	-0.1508171
GNG12	0.1318200	MRPS18C	-0.1535766
POU3F2	0.1022426	NCALD	-0.1858802
AEBP1	0.1498719	C11orf1	-0.1448555
IQCA1	0.1134073	DCTN6	-0.1222108
ERBIN	0.1309312	SEM1	-0.1765024
LOC202181	0.1184466	APOO	-0.1384320
LPP	0.1072798	CCNH	-0.1394853
NOTCH2	0.1213843	RAD51C	-0.1280948

Table 2.4: Top 25 up- and down- regulated genes in Alzheimer's disease compared to healthy controls.

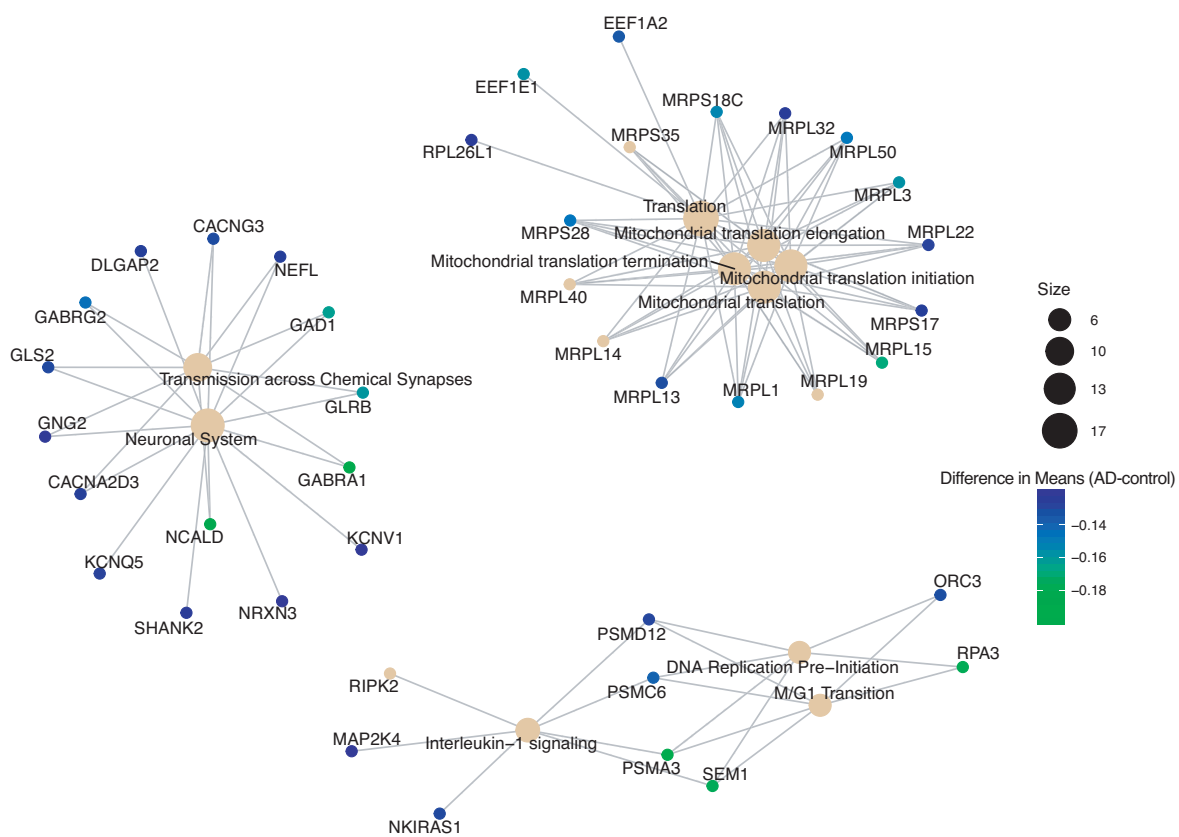


Figure 2.5: Pathway-gene network of top 10 enriched Reactome pathways from down-regulated genes in Alzheimer's disease patients.

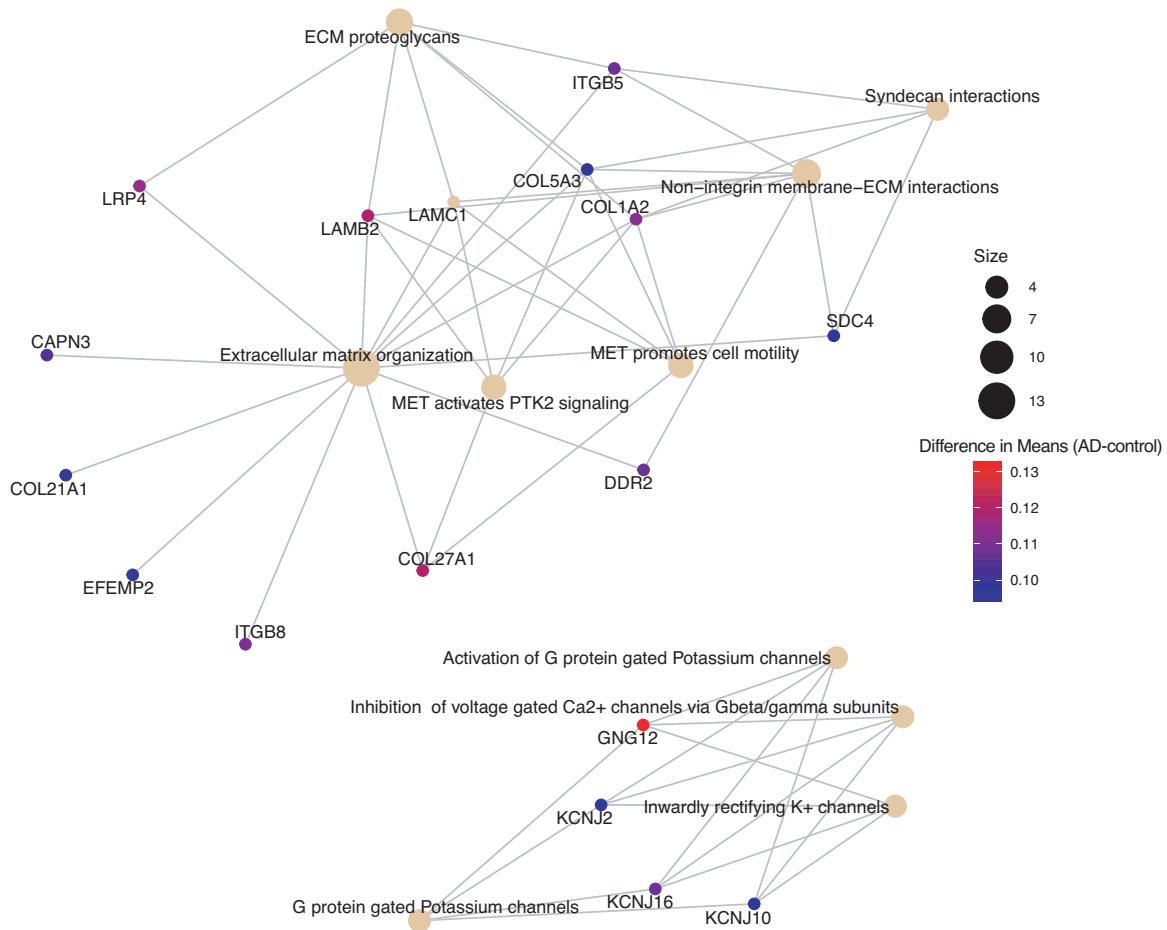


Figure 2.6: Pathway-gene network of top 10 enriched Reactome pathways from up-regulated genes in Alzheimer's disease patients.

2.4.4 Aging and Tissue Differences in AD Gene Expression

To determine if age or tissue had an effect on the DEG by disease status, we filtered the 352 DEG in disease results discussed above for age group and tissue comparisons. For age effects, we used our TukeyHSD results that compared age groups to <60 (served as the baseline). This allowed us to explore if genes associated with AD change with age by using a common reference group. We used the 352 DEG genes from disease status TukeyHSD results to find sizable age effects in this gene set by selecting for statistical significance and using the two-tailed 10% quantile filter(\leq

-1.0477827 and ≥ 0.330869) to find significant DEG per age-group pair comparison (Table A.2 of Appendix A). We found 396 significant comparisons of age differences in 141 genes (see ST10 of online supplemental data (Appendix A)). The 141 genes were plotted across all age comparisons where < 60 was the baseline to visualize expression changes and how the genes clustered (Figure A.10 of Appendix A), indicative of distinct differences in expression profiles due to aging. There is a cluster of genes down-regulated in older age groups, specifically ages 65-80 compared to those < 60 . There also appears to be an overall trend of genes associated with disease being up-regulated compared to < 60 . Of the 141 DEG by age group (Figure A.10 of Appendix A), we found 114 DEG that had a statistically significant interaction (Bonferroni corrected p-value < 0.05) between disease status and age (Figure 2.7). Changes in expression across each age group comparison (< 60 baseline) in the interacting genes were visualized, and the genes clustered into 3 clear groups based on similarities in expression patterns (Figure 2.7 and Figure A.10 of Appendix A).

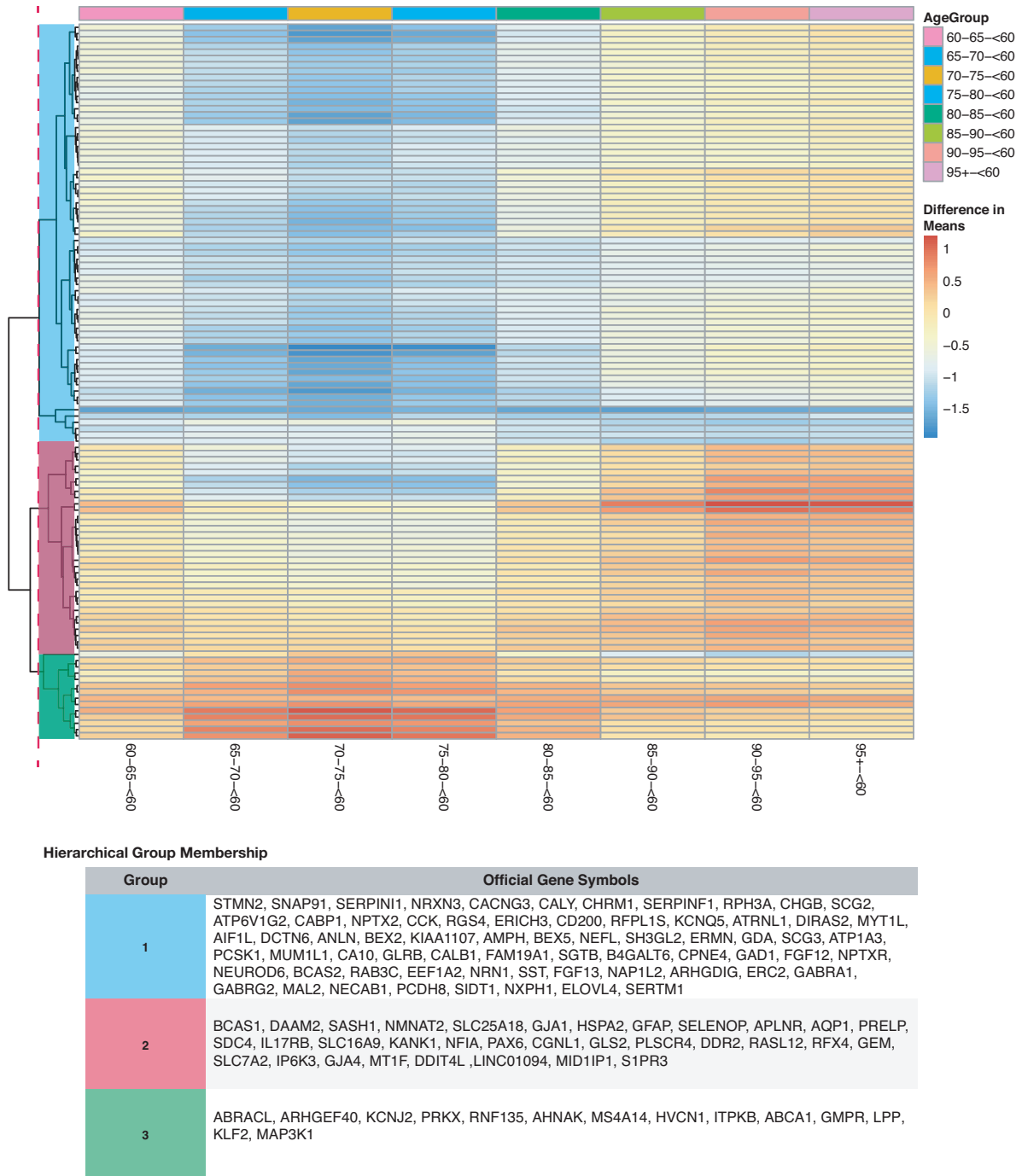


Figure 2.7: Heatmap with gene clustering to visualize age group effect (difference in means) on the differentially expressed disease (control-AD) gene list that have agegroup:disease status interaction.

For tissue effects, we used hippocampus as our baseline due to it being a known target of AD. In addition to filtering for significance, we used again a two-tailed 10% quantile filter ≤ -0.6359497 and ≥ 0.7932871 from the tissue-specific means differences between tissue types (Table A.2 of

Appendix A). We found 167 comparisons with tissue differences (see ST11 of online supplemental data (Appendix A)) from 125 genes. Our heatmap of these genes show that differences do exist across tissues when compared to hippocampus (Figure A.11 of Appendix A). For example, nucleus accumbens has higher expression of genes compared to the hippocampus, and putamen has genes that are down-regulated compared to hippocampus (Figure A.11 of Appendix A)). The majority of the expression differences appear to be found in nucleus accumbens and putamen (Figure A.11 of Appendix A), see also ST11 of online supplemental data (Appendix A)). From these 125 tissue specific (hippocampus) genes, we found 13 to have a statistically significant (Bonferroni corrected p -value < 0.05) interaction between disease and tissue (Figure 2.8A).

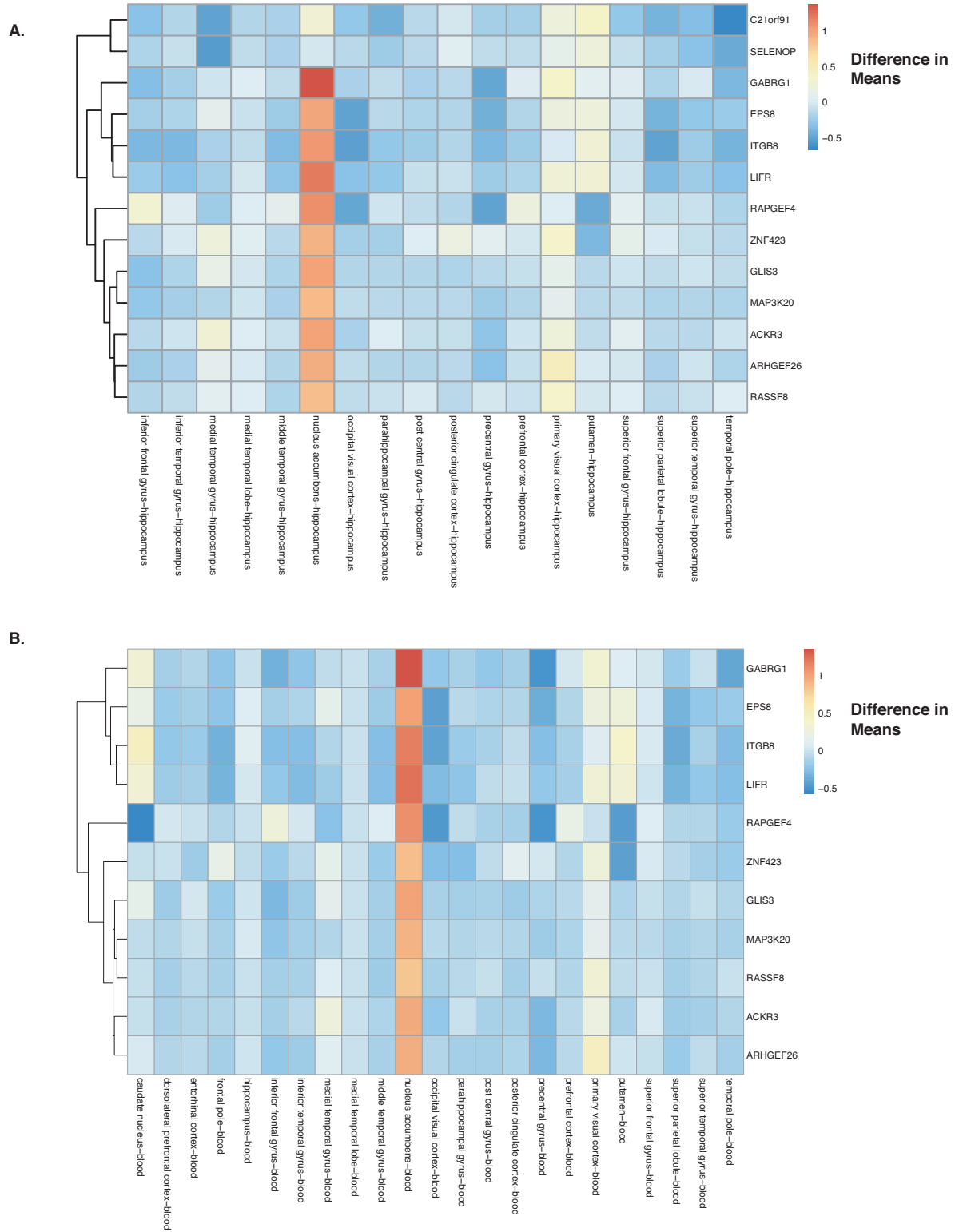


Figure 2.8: Heatmap with gene clustering to visualize tissue effect (difference in means) on the differentially expressed disease (control-AD) gene list that have tissue:disease status interaction.

We also assessed how gene expression changes in a given tissue compared to blood (10 %quantile filter: ≤ -0.6359497 and ≥ 0.7932871) (Table A.2 of Appendix A), identifying 152 significant tissue comparisons in 115 genes (see ST12 of online supplemental data (Appendix A)). These 115 gene expression profiles across tissues are visualized using the differences of means in Figure A.11 of Appendix A). We again noticed similar trends in the blood comparisons as had in the hippocampus comparisons, with nucleus accumbens showing higher gene expression and putamen lowered expression compared to blood (Figure A.12 of Appendix A). Finally, we found that 11 of these genes had a statistically significant (Bonferroni corrected p-value < 0.05) interaction between disease and tissue (Figure 2.8B).

2.5 Discussion

As debilitating as Alzheimer's disease (AD) is, there is still no cure available, and diagnosis is not confidently confirmed until death. There are ongoing research efforts to find biomarkers and gene targets for early detection and intervention in AD. In our study, we investigated changes at the transcript level by conducting a meta-analysis to analyze 8 microarray expression datasets for temporal changes in gene expression due to disease status. In addition to this, we determined if sex, age or tissue type had an effect on gene expression changes in Alzheimer's associated disease genes. We pre-processed the 8 datasets by background correction, data normalization, and probe annotation. Following this, the datasets were merged into a single dataset (by common gene name) for the meta-analysis. This is the first meta-analysis to explore over 20 different tissues and use a linear model to identify linear and binary effects on gene expression. Our linear model also adjusted batch effects by modeling for the study effect and included age in the model as a linear time series. Modeling with the study factor to account for batch effects was shown to be necessary after exploratory visualization of the expression data before and after combat batch effect correction using principal component analysis to remove variation within the data that was introduced due to different studies (Figures 2.2,2.3).

2.5.1 Significant Gene Expression Differences Due to Disease Status and Biological Significance

We first identified statistically significant disease genes (p-value <0.05; factor: disease status) from ANOVA (see ST4 of online supplemental data (Appendix A)), and these genes included: APOE, PSEN2, APOD, TREM2, CLU which all have been previously associated with AD. APOE and APOD are members of the apolipoprotein family that transport and metabolize lipids in the central nervous system and play a role in healthy brain function [62]. APOE is a strong, well documented, genetic risk factor for AD, and polymorphisms in APOE have been shown to affect age of AD onset [29]. APOD's mechanism is still not completely understood [62], PSEN2 encodes presenilin-2, an enzyme that cleaves APP, regulates production of A β , and mutations are associated with early onset [29]. Mutations in CLU lead to lower white matter and increases AD risk [29, 63] and TREM2 was identified by a genome-wide association study (GWAS) as a disease variant and risk factor for AD [29]. Our enrichment results of the 3,735 genes (from ANOVA) were interesting due to them having already been associated with AD in the literature (Table 2.3 and also see ST5 and ST6 in online supplemental data (Appendix A)). For instance, mitochondrial dysfunction has been previously associated with AD and characterized to cause A β deposition, higher production of reactive oxygen species and lowered ATP production [64–66]. Researchers have also suggested that the immune system plays a role in AD [67, 68]. As for adaptive immune cells, their role in AD is still not clear, however, adaptive immune cells have been shown to reduce AD pathology [69]. The loss of B cell production can exacerbate the disease [69]. Neurodegenerative diseases have also been described as having genes that overlap [33, 34]. Neurodegeneration is closely related to synaptic dysfunction and long term potentiation becomes impaired with age and synaptic dysfunction [70]. These results suggest that our meta-analysis is producing disease-related results (Table 2.3 and also see ST5 and ST6 in online supplemental data (Appendix A)).

We also identified the KEGG AD pathway as one of our enriched pathways based on the 3,735 statistically significant disease genes. To explore how these genes are regulated in the AD pathway, we used the difference of means (using the TukeyHSD) to create Figure 2.4 which highlights 73

of the 3,735 genes from our ANOVA analysis and their role in the KEGG AD pathway (see ST7 online supplement data (Appendix A)). NAE1, also known as amyloid precursor protein-binding protein 1 (APP-BP1), was down-regulated in AD subjects and is involved in neuronal apoptosis (Figure 2.4). The literature indicates that APP-BP1 is necessary for cell cycle progression and activates the neddylation pathway that drives apoptosis [71–76]. Down-regulation of APP-BP1 has been associated with increased APP while over expression of APP-BP1 leads to APP degradation [71–76]. TNFRSF6 was up-regulated in AD subjects (Figure 2.4, and this gene produces the Fas antigen which plays a role in mediating apoptosis. [77].

The KEGG AD pathway also highlights genes from our analysis that are involved in APP processing and cleavage (Figure 2.4). Specifically, BACE, PSEN and APH-1 are all involved in APP processing by coding for γ -secretase and β -secretase (Figure 2.4). BACE is a β -secretase, that we found to be up-regulated in AD subjects compared to controls (Figure 2.4). This finding also supports previous reports that BACE is over-expressed in AD brains, and plays a role in forming $A\beta$ [78, 79]. APH-1A and PSEN2 are a part of the γ -secretase complex that finalizes cleavage and release of APP to produce $A\beta$ [80–82]. As shown in Figure 2.4, in AD subjects there was a high production of APH-1 while PSEN2 was down-regulated. This indicates that while in a complex, the two genes may function differently. For example, mutations in PSEN2 can lead to memory loss and loss of synaptic plasticity [83]. A better understanding of the mechanistic behavior of the γ -secretase complex genes can aid in the potential development of targeted therapeutics for γ -secretase. Also in the AD pathway we found up-regulated expression of APOE and LRP1 in AD subjects compared to control subjects (Figure 2.4). These genes are both involved in $A\beta$ aggregation. LRP1 a known receptor of APOE and promotes $A\beta$ aggregation and migration across blood-brain barriers [84].

As discussed above, mitochondrial dysfunction is a key hallmark of AD. Genes from our meta-analysis that are in the AD pathway are involved in the respiratory electron chain transport complexes. For example, NDUFC2 (in CxI on Figure 2.4), SDHA (in CxII on Figure 2.4), and COX5B, COX6A1, COX6C (in CxIV) are all necessary for electron transport, but were down-

regulated in AD (Figure 2.4). In Figure 2.4, complexes I-IV of the electron chain transport were all down-regulated in AD. Previous work observed lower expression of 70% of genes that code for subunits of the electron transport chain [43]. Reduced mitochondrial translation and lowered mRNA levels for genes such as cytochrome oxidase (COX), can lead to increased oxidative stress, irregular calcium levels and decreased oxidative phosphorylation (OXPHOS) [43, 85–89]. Hence, changes due to mitochondrial dysfunction may affect the pathology of neurodegenerative diseases such as AD.

We also found ITPR3, a gene involved in the calcium signaling pathway, was up-regulated in AD (Figure 2.4). ITPR3 is necessary for the release of Ca^{2+} from the endoplasmic reticulum [90]. Increased expression of this gene and calcium concentrations can cause memory loss and neuron cell death (Figure 2.4) [90]. Additionally, we found genes involved in tau phosphorylation to be up-regulated in AD (Figure 2.4). Calpain (CAPN1,CAPN2) which is activated by elevated levels of cytosolic calcium is up-regulated as well as CASP7 [91]. Together these genes regulate tau phosphorylation and the formation of neurofibrillary tangles, which eventually leads to neuronal cell death (Figure 2.4).

In addition to enrichment in the AD pathway, our KEGG results on the 3,735 genes included enrichment in Parkinson's disease and Huntington's disease pathways. Because of this we investigated if the three neurodegenerative disease signaling pathways had any common genes in our gene list (Table 2.3). We determined that AD had 49 genes that overlapped with Huntington's and 47 with Parkinson's pathways respectively. We also found that GNAQ, GRIN1 and PLCB1 are in both Huntington's and AD but not in Parkinson's pathways, and SNCA is in both Parkinson's and AD but not Huntington's pathways. In filtering the statistically significant disease genes for biological effect size (post-hoc analysis), PSEN2, APOE, TREM, CLU and other apolipoproteins did not make the cutoff (based on their difference in means between the compared AD/healthy groups).

Focusing on the 352 DEG that had a sizable biological effect, the down-regulated genes in AD connect with the pathology of the disease (Figure 2.5). Specifically, genes in the Mitochondrial translation pathway that were down-regulated in AD included MRPL15, MPRL13 and MRPL1,

which are all mitochondrial ribosomal proteins necessary for protein synthesis [92]. These genes may also be related to down-regulation of the mitochondrial electron transport chain complexes [93] in the KEGG AD pathway (Figure 2.4). Translational elongation factors (EEF1E1 and EEF1A2) were also down-regulated (Figure 2.5). Previous findings have indicated a reduction in EEF1A expression in AD patients specifically in the hippocampus [94]. Genes down-regulated in the Neuronal System pathway and Transmission across Chemical Synapses included GABRA1, GABRG2, NCALD, GAD1 and NEFL (Figure 2.5). GABRA1 and GABRG2 are receptors in the gamma-aminobutyric acid (GABA) signaling system that bind to GABA (inhibitory neurotransmitter) and regulate chloride levels in the brain [95, 96]. In AD, the GABA signaling system is dysregulated with changes in GABA expression in the hippocampus citepcalvo2018gabaergic. NCALD is a calcium sensor that is involved in neuronal calcium signaling [92, 97]. NEFL makes the protein neurofilament light chain (Nfl), which has recently been investigated as a fluid biomarker for monitoring AD disease progression [98].

Our results also included down-regulated genes PSMA3, PSMC6 and SEM1 that are part of the proteasome complex (cell cycle progression and DNA damage repair) [92, 99, 100] and replication factor protein, RPA3 (needed to stabilize single stranded DNA during DNA replication) [92, 101], which are down-regulated in the DNA Replication Pre-Initiation and M/G1 Transition pathways. It has been reported that incomplete DNA replication and irregular cell cycle events such as abnormal cell cycle reentry by neurons have been observed in AD brains and lead to cell death [102]. Additionally, dysregulation of the proteasome complex in AD is supported by the literature [103–106]. However, the role of the proteasome complex in AD and how it is regulated is still not clearly understood [103], and merits further consideration.

Reactome pathway analysis on the up-regulated genes resulted in some interesting pathways such as Extracellular Matrix (ECM) Organization, ECM proteoglycans, Mesenchymal Epithelial Transition (MET) activates PTK2 signaling, MET promotes cell motility, Non-integrin Membrane-ECM interactions and Syndecan Interactions, which all had overlapping genes (Figure 2.5). CAPN3, COL21A1, EFEMP2 and ITGB8 were only in the ECM organization pathway (Figure 2.6).

COL21A1 has been described as being necessary for maintaining the integrity of the ECM, and has been previously found to be up-regulated in severe AD [107]. Additionally, changes in the ECM components and degradation with proteases have previously been found to be associated with plaque formation, which causes brain dysfunction [108–110]. The up-regulated genes in the potassium and Ca^{2+} channel pathways included GNG12, KCNJ2, KCNJ16 and KCNJ10. In general, as potassium channels open to increase potassium in the cells, calcium is decreased by inhibiting the Ca^{2+} gated channels [96]. Increased activity of the potassium channels, especially the voltage-gated channels have been associated with regulating microglia function and priming which in turn leads to increased ROS production in AD [111, 112].

We compared the 352 genes identified as differentially expressed and exhibiting a biological effect with respect to disease status to a recently published meta-analysis in which 1400 differentially expressed disease genes were identified [34]. We determined that 136 DEG from our gene list overlapped with Moradifard et. al's findings., and 216 of our DEG were not in their list [34]. Genes that were unique to our DEG list included GMPR, ABCA1, NOTCH1 and 2, GABRG1, HVCN1, CXCR4, HIP1, MRPS28, FOS.

The top up-regulated gene in AD from our meta-analysis, ITPKB, (Table 2.4) has previously been observed to have over-expression in AD subjects. In a mouse model, the gene was found to be over-expressed and connected to apoptosis, increased ($\text{A}\beta$) production and tau phosphorylation [113]. Additional DEG included CXCR4 (brain development and neuronal cell survival in the hippocampus) [92, 114], AHNAK (may have a role in development of neuronal cells)[92], NOTCH1, and NOTCH2 (signaling pathway may be involved in brain development) [92, 115] which were all up-regulated in AD subjects (Table 2.4). On the other hand, RPA3 (DNA replication), NME1 (neural development) [92, 116], and mitochondrial proteins MRPL3, MRPS18C (associated with mitochondrial dysfunction observed in AD) were down-regulated in AD samples (Table 2.4).

2.5.2 Sex, Age and Tissue Effect on Disease Status Biologically Significant Genes

For the sex factor, we determined that 46 of our DEG (23 up- and down- regulated in males compared to females) had a sex effect, with 1 of them (CXCR4) showing a statistically significant (p-value <0.05) interaction between disease status and sex. The enriched pathways from the up-regulated genes (prior to selecting for interacting genes) in males are highlighted in Figure A.8 of Appendix A. Furthermore, these genes involved in pathways such as Clathrin-mediated endocytosis (SNAP91, SH3GL2, and AMPH), Neuronal System, Neurotransmitter receptors postsynaptic transmission and Transmission across Chemical Synapses (GABRG2, GABRA1, GAD1 and NEFL) were down-regulated in females (Figure A.8 of Appendix A and Table A.3 of Appendix A). Down-regulation in genes such as GABRG2, GABRA1, GAD1 and NEFL) was previously discussed as being down-regulated in AD from our DEG list for disease status (Figure 2.5).

Additionally, the current literature indicates that women are at higher risk for AD [117–119]. This increased risk by sex is due to the loss of estrogen protection (due to menopause) against (A β)'s toxicity on the mitochondria [117, 118]. Older women produce more reactive oxygen species with the decline in estrogen levels [117, 118]. Estrogen replacement therapy is a treatment for AD, and it is being determined that estrogen works by increasing the expression of antioxidant genes [117, 118]. A recently published meta-analysis also explored sex effects on AD gene expression [34]. Moradifard et al., found male and female specific AD associated genes and genes that overlapped in both sexes [34]. Of the 46 disease associated genes we found to be affected by sex, 22 were found in both males and females, 9 only in males, and 5 only in females in Moradifard et al gene list. 10 of our sex impacted disease genes (CYBRD1, DIRAS2, FAM107B, FOS, GMPR, HVCN1, ITIH5, MAPK, RNF135, SLC40A1) did not overlap with their findings, and these genes have been previously associated with oxidative stress, cell signaling and transport, apoptosis and AD. For instance, GMPR was found to gradually increase as AD progressed [120]. It produces GMPR1 which is associated with the phosphorylation of tau [120].

Focusing on the statistically significant pairwise interaction between disease status and sex, we identified CXCR4 which was up-regulated in females (Table A.3 of Appendix A). CXCR4 was

also up-regulated in AD (Table 2.4). CXCR4 has been previously investigated for its role in AD and other neurodegenerative diseases [114, 121, 122]. CXCR4 is a chemokine receptor that binds to CXCL12, and together they are involved in signaling pathways for inflammation and neuronal system function [114, 121, 122]. CXCR4/CXCL12 together regulate synaptic plasticity, apoptosis, calcium levels, microglia to neuron communication, neuronal signaling and neuroinflammation [114, 121, 122]. Dysregulation of CXCR4 has been associated with neurodegenerative diseases [114, 121]. More specifically, up-regulation of CXCR4 in a mouse model led to abnormal signaling in microglia and tauopathy [121].

Aging trends on the differentially expressed disease genes were visualized in Figure A.10 of Appendix A and Figure 2.7. Subjects grouped as <60 were used as a baseline because on average, AD symptoms start at ages 65 and older [29]. We observed clear age-related patterns when looking at the difference of means between age cohorts (prior to selecting for interacting genes) for the disease gene list (Figure A.10 of Appendix A and see ST10 of online supplemental data (Appendix A)). Highlighting a few of the changes: SNAP91 which is involved in synaptic transmission and associated with late onset [123], STMN2 which is necessary for microtubule dynamics and neuronal growth [124, 125], and SST, a neuropeptide that interacts with ($A\beta$) and can influence how it aggregates [126, 127] were all up-regulated in <60 age group (Figure A.10 of Appendix A and see ST8 of online supplemental data (Appendix A)). Also, STMN2 and SST have both previously been associated with expression reduction due to age [92, 126]. ABCA1, GMPR, HVCN1, ITPKB, NOTCH1 all had higher expression in older age groups compared to the baseline.

Furthermore, visualizing the genes with a statistically significant interaction (p-value <0.05) between disease and age group, we observed three distinct groups of genes with similar patterns (Figure 2.7). Genes identified in group 1 in Figure 2.7 were down-regulated in ages 65 to 80 compared to the baseline (<60 years old). Group 1 genes also displayed a slight increase in relative expression from ages 85 and higher (Figure 2.7). Reactome pathway analysis on the group 1 genes identified 3 enriched pathways that were statistically significant (FDR <0.05): (i) MECP2 regulates transcription of genes involved in GABA signaling (GAD1) [128, 129], (ii) Muscarinic

acetylcholine receptors (CHRM1) [128, 130] and (iii) Neuronal System (CACNG3, GAD1, NEFL, GABRA1, GLRB, NRXN3, GABRG2, and KCNQ2) [128, 131]. Changes in GABA signaling in AD was previously characterized as age-dependent [132]. The ionic response to GABA, also reported as GABA currents, were reduced in AD, especially in younger subjects with AD [132]. We observe a similar pattern in our meta-analysis for the GABA receptor genes in group 1 (Figure 2.7). Genes within group 2 displayed a gradual increase in expression with age (Figure 2.7). Reactome pathway analysis did not identify statistically significant enrichment for these genes. However, genes in group 2 include DDR2 (regulates TREM2, microglia and neurotoxic proteins) [133], IP6K3 (Inositol phosphate metabolism) [134], and GJA1 (regulates known AD risk factor genes) [135]. Additionally, genes in group 3 exhibited significant up-regulation in gene expression for subjects 65 to 80 years with a gradual decrease in expression from ages 85 and older (Figure 2.7). These genes are associated with the statistically significant enriched pathway (FDR <0.05), TRAF6 mediated NF-kB activation (MAP3K1) [128, 136]. Our findings highlight genes previously associated with AD and their temporal trends, and also some additional genes that experience age-effects (Figure 2.7, Figure A.10 of Appendix A, and see ST10 of online supplemental data (Appendix A)).

To investigate tissue-specific effects (prior to selecting for statistically significant pairwise interactions between tissue and disease status), we used hippocampus (232 samples) as a baseline due to it being identified as one of the first regions to be affected by AD [29]. We also used blood (519 samples) as a baseline to explore an underdeveloped non-invasive approach to monitoring AD. In both analyses, we saw similar trends with the nucleus accumbens (51 samples) and putamen (52 samples) showing greater differences in expression (Figures A.11 and A.12 of Appendix A). Focusing on the genes that showed a statistically significant interaction between disease and tissue, we observed lower expression of genes in tissues compared to the hippocampus and blood with a slight increase in the primary visual cortex and the putamen (Figure 2.8). As for the nucleus accumbens we observed significantly higher expression for these interacting genes for both hippocampus and blood baseline comparisons (Figure 2.8). The statistically significant (p-value <0.05) interacting genes in Figure 2.8 include genes that are involved in development of dendritic

spines (C21orf91), normal brain function (SELENOP), GABA signaling (GABRG1), and structure of actin cytoskeleton (EPS8) [92, 95, 137–139]. In addition to the shrinking of the hippocampus, decreases in volumes for nucleus accumbens and the putamen have also been reported [140, 141]. The nucleus accumbens is important for reward processing, and in AD has been associated to impaired decision making and reduction in performance of rewarding behaviors [142]. AD is also associated with reduced dopamine levels and GABA signaling [143]. Finally, the putamen (motor behaviors) and primary visual cortex (visual processing) both have impaired functions in AD [144, 145].

The distribution of samples per tissue type was inconsistent with hippocampus and blood having larger number of samples compared to an average of around 55 samples per tissue in other categories. These results show the potential of blood and other tissues for monitoring gene expression changes in AD, but also the need for further focused mechanistic studies in different tissues.

2.5.3 Limitations of the Study

Using publicly available data introduced limitations to our research design. Lack of uniform annotation and missing information across datasets can make conducting a meta-analysis on multiple datasets challenging. For example the subclass of AD, details on cognitive status and APOE genotype were not uniformly reported across the datasets used (Table A.1 of Appendix A). The brain samples were from a variety of brain banks with varying institutional review boards and standards, protocols and criteria for AD diagnosis requirements (Table A.1 of Appendix A). Additionally, the number of datasets used in our meta-analysis was limited by poor annotations that could not meet our selection criteria, and this in turn placed bounds to our sample size and power of the study. Our analysis was also unbalanced: 2,088 samples made up of 771 healthy controls, 868 AD subjects, 449 subjects reported as possibly having AD, 1308 females and 780 males, and the breakdown of age groups is also somewhat uneven. One of our datasets (GSE84422) consisted of paired samples. However, as the the other datasets did not include paired samples, we did not incorporate a paired-

sample analysis in our study. The available public data used for our meta-analysis also lacked diversity in samples, because in most datasets race and ethnicity are not reported. This information would be helpful particularly since AD has been reported by the CDC to be more prevalent in African Americans [146, 147]. In addition, the use of micro-array expression data for meta-analysis is a limitation in terms of not being able to query the entire transcriptome or query novel genes. Also, in our merged dataset, large variability was introduced in data due to the large number of tissues (26) and methods used for extractions (study effect), which we attempted to correct for by utilizing both as factors in our model, and including binary interaction terms as well. An additional limitation of our study is that we included datasets that investigated gene expression changes in bulk tissue rather than on the cell-type-specific level. Cell-type-specific expression data that matched our inclusion criteria were not available to include in this meta-analysis. Furthermore, single-cell data is also only recently becoming available. A meta-analysis including single-cell analysis expression data from specific cell types such as neurons, astrocytes and microglia would allow an improved understanding of gene expression differences between AD and healthy controls [148, 149]. Finally, to our knowledge, there were also a limited number of RNA-sequencing (RNA-seq) datasets on GEO and Array Express (23), and only one that matched our selection criteria. Thus, we elected to carry out the analysis using the gene expression array data. We anticipate that more RNA-seq data, which can provide a more global view of the transcriptome, will become available in the future.

2.5.4 Future Directions and Recommendations

Our study provides gene lists by factor (disease status, sex, age and tissue) of differentially expressed genes. Our study is largely descriptive, but also yields new gene candidates which we may be studied further for their role in AD, including underlying mechanisms using model systems. To expand on this research, the use of RNA-seq data can reveal novel differentially expressed genes, biomarkers and gene targets for AD. As more RNA-seq data becomes available, a similar meta-analysis approach may be applied, if such data are annotated to include the necessary factors' metadata for the analysis. In addition to RNA-seq, implementing other omics technologies such

as proteomics and metabolomics can help to fully describe the pathology of AD, and identify additional biomarkers for early detection. To promote more meta-analyses, we recommend that future studies include more extensive, and structured standardized metadata in their submissions, that will enable use of data. Including data with racial diversity is also necessary. AD has higher prevalence in African Americans [147]. Due to reports of racial differences in AD, with an AD prevalence breakdown of: 14% of African American population compared to 12% in Hispanics and 10% in whites [146], including racial diversity in future studies would help identify this potential variability in susceptibility and identify if certain treatments might be better suited in some races than others. Improving the representation of races in clinical trials and molecular reports of AD can help with health disparities within the field. Exploring the use of easily accessible tissues, such as blood, to monitor changes in target genes/biomarkers might also prove helpful for early detection and provide a more systems-level understanding of AD. Determining the best or novel biomarkers to track for AD requires exploring also mechanistic aspects of the disease. For example, monitoring exosomes and autoantibodies which can be connected to the dysfunction of the immune system is one mode of action that is being associated with AD [150]. Lastly, as omics technologies advance, implementing personalized omics for early detection and treatment may prove useful in improving individual AD outcomes with the increase in the aging population.

CHAPTER 3

STREPTOCOCCUS PNEUMONIAE'S VIRULENCE AND HOST IMMUNITY: AGING, DIAGNOSTICS AND PREVENTION

Work presented in this chapter has been published as Brooks LRK, Mias GI. Streptococcus pneumoniae's virulence and host immunity: aging, diagnostics and prevention. *Frontiers in Immunology*. 2018;9:1366.

3.1 Abstract

Streptococcus pneumoniae is an infectious pathogen responsible for millions of deaths worldwide. Diseases caused by this bacterium are classified as pneumococcal diseases. This pathogen colonizes the nasopharynx of its host asymptomatically, but overtime can migrate to sterile tissues and organs and cause infections. Pneumonia is currently the most common pneumococcal disease. Pneumococcal pneumonia is a global health concern and vastly affects children under the age of five as well as the elderly and individuals with pre-existing health conditions. *S. pneumoniae* has a large selection of virulence factors that promote adherence, invasion of host tissues and allows it to escape host immune defenses. A clear understanding of *S. pneumoniae*'s virulence factors, host immune responses and examining the current techniques available for diagnosis, treatment and disease prevention will allow for better regulation of the pathogen and its diseases. In terms of disease prevention, other considerations must include the effects of age on responses to vaccines and vaccine efficacy. Ongoing work aims to improve on current vaccination paradigms by including the use of serotype independent vaccines, such as protein and whole cell vaccines. Extending our knowledge of the biology of, and associated host immune response to *S. pneumoniae* is paramount for our improvement of pneumococcal disease diagnosis, treatment and improvement of patient outlook

3.2 Introduction

Infectious diseases present a significant global burden affecting society [151, 152]. Most of these diseases are due to exposure to or the invasion of host cells and organs by microorganisms [151–153]. These pathogens disrupt the normal function of the human body by hindering immune responses and producing harmful toxins. Infectious diseases can easily spread from person-to-person via contact with body fluids, indirect contact or through animal vectors such as mosquitos and ticks [154]. Common widespread diseases of the respiratory system occur when microorganisms invade the respiratory tract. Infectious respiratory diseases are globally seen as a major health concern because they can rapidly become severe and lead to death. Respiratory diseases are categorized

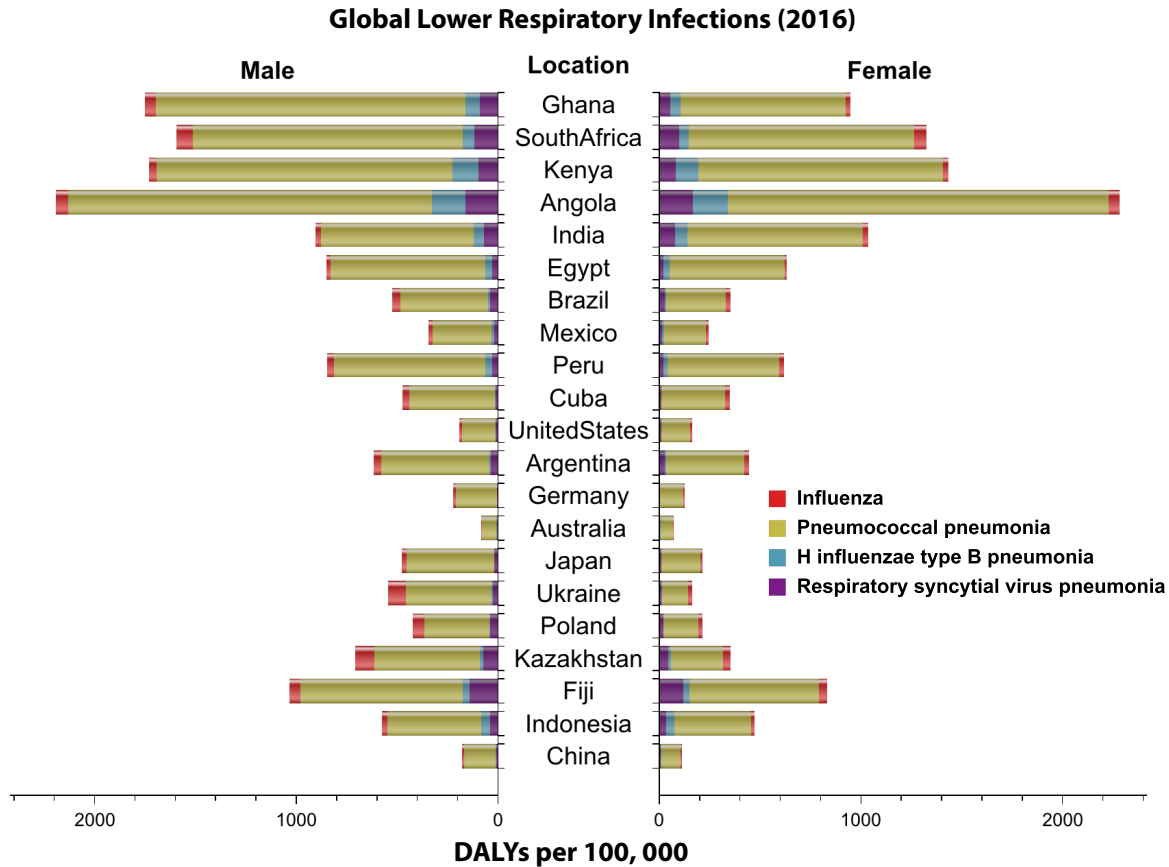


Figure 3.1: Global distribution of lower respiratory infections by sex. Highlighted in this figure is the distribution of the disability adjusted life year (DALY) per 100,000 (2016) for four major lower respiratory infections worldwide by sex. Data obtained from Institute for Health Metrics and Evaluation [4]

into upper and lower respiratory tract infections. Lower respiratory tract infections (LRIs) are more severe because pathogens infect sterile parts of the respiratory system such as: the lungs, trachea and bronchi [155]. In 2013, an estimated 2.6 million deaths worldwide were attributed to LRIs, while by 2015, this increased to 2.74 million [156]. Higher burden of LRIs is associated with low sociodemographic status, poor access to healthcare and nutrition (Figure 3.1) [4, 156].

Immune system function is important in a host's defense to pathogens. A host with a healthy and well-developed immune system is able to clear pathogens before they can become infectious and cause diseases [157–160]. The ability to clear pathogens before they can become infectious depends on the quality of the immune system and its effectiveness, which is linked strongly to age

[157, 161]. The immune system continues to develop from infancy to adulthood, while later in life a fully developed immune system begins to deteriorate with aging. Infants and the elderly are at higher risks for contracting infectious diseases due to their weakened immune system and the inability to clear the pathogens before they become pathogenic [157–160, 162–165]

Streptococcus pneumoniae is a bacterium that has been widely linked to causing respiratory infections in individuals with a weakened immune system [158, 161, 164]. *S. pneumoniae* is spread through airborne droplets, and it is estimated to cause about 4 million illnesses within the United States (US) and about 450,000 hospitalizations per year [166, 167]. Studies indicate that 10% of patients with invasive pneumococcal diseases die of their illnesses [168, 169]. *S. pneumoniae* invades its host by colonizing the nasopharynx asymptotically as it has been found to be part of the commensal microbiota of the upper respiratory tract [170, 171]. After colonization, if the bacterium is not cleared by the immune system, the bacterium is spread via horizontal dissemination into the lower airways and other organs and tissues, and becomes pathogenic [171]. A strong immune system and the balance between resident flora and invaders can help to clear *S. pneumoniae* before it becomes pathogenic. With poor defense mechanisms, the host becomes subject to frequent and long-lasting colonization of *S. pneumoniae*, which can later lead to diseases [172, 173].

The bacterium has several properties which allow it to go unnoticed by the host immune system, and defend against the resident flora within the nasopharynx that would try to clear it [165, 174, 175]. Thus, decreasing the burden of this bacterium and preventing further infections is very important to the healthcare field [175, 176]. Furthermore, *S. pneumoniae* is an opportunistic pathogen that takes advantage of hosts with underdeveloped, weakened and or deteriorating immune systems. Because of this, *S. pneumoniae* has greater incidence rates in children under the age of two, the immunocompromised and the elderly [177]. Figure 3.2 depicts that disease burden for major LRIs are highest in young children and the elderly [4, 168, 178–180]. Understanding how the immune system changes with age is important in providing appropriate treatments to hinder colonization of weaker hosts.

In this review, we provide a concise introduction to the expanding literature on *S. pneumoniae*,

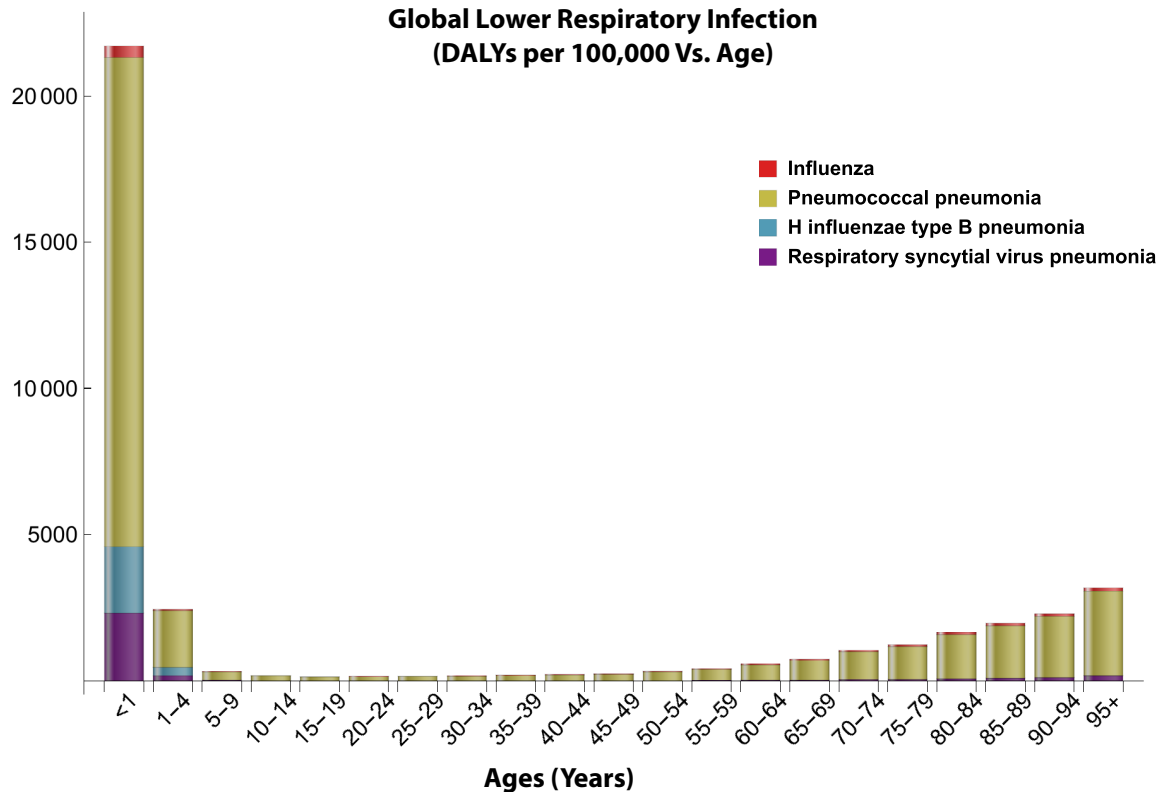


Figure 3.2: Global distribution of lower respiratory infections with age. This figure shows the age-dependent disease burden to lower respiratory infections especially pneumococcal pneumonia based on the disability adjusted life year (DALY) data from 2016. Data obtained from Institute for Health Metrics and Evaluation [4]

and focus on exploring the characteristics of *S. pneumoniae*, its pathogenesis, its virulence factors, and pathology. We will also delve into the general host immune response to *S. pneumoniae*, with a focus on pneumonia, and connect the severity of this disease to varying host immune responses with age. In addition, we will explore the medications available to prevent or treat pneumococcal diseases such as pneumonia, disease prognosis, and finally discuss what the future holds for pneumococcal diseases.

3.3 Pneumococcal Disease, Epidemiology and Transmission

Streptococcus pneumoniae, a Gram-positive bacterium (Figure 3.3) , also known as pneumococcus, can survive in both aerobic and anaerobic conditions [181]. It is a facultative anaerobe that

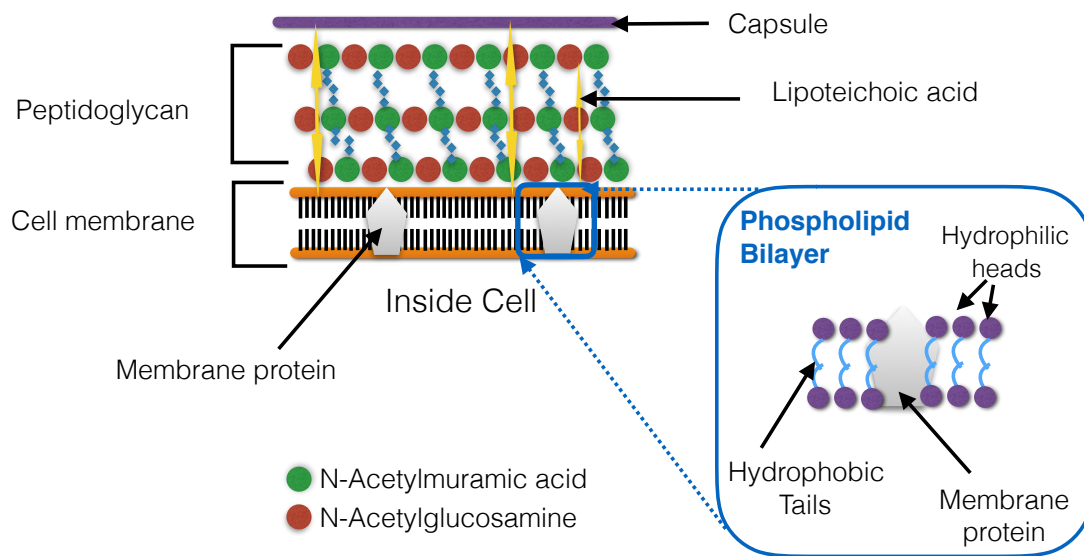


Figure 3.3: Schematic cross section of *Streptococcus pneumoniae* cell wall. The bacterial cell wall composes of teichoic acids, a thick peptidoglycan layer, and a phospholipid bilayer

is often found as diplococci [181]. Pasteur and Sternberg first isolated *S. pneumoniae* from saliva in 1881 [182–184]. Currently, there are varying reports on the number of identified serotypes of *S. pneumoniae* [173, 183, 185, 186]. However, there are at least 97 serotypes of *S. pneumoniae* that have been identified and characterized to date [183, 187]. All of these serotypes are independently recognized by the host [170, 173, 188–190].

Pneumococcal diseases occur worldwide [173, 175, 191] and are more prevalent in young children, the elderly and immunocompromised individuals (Table 3.1) [170, 171, 181, 190, 192, 193]. *S. pneumoniae* causes many pneumococcal diseases such as meningitis, bacteremia, pneumonia, acute otitis media and sinusitis [173]. *S. pneumoniae* causes about 40,000 fatal pneumococcal infections per year within the United States [172, 181, 194, 195].

S. pneumoniae colonizes the upper respiratory tract—specifically the nasopharynx [170, 196], and is able to asymptomatically reside in the upper respiratory tract—this is known as carriage [170]. Carriage is more prevalent in children (20–50%) compared to adults (5–20%) [196–198]. Carriage can lead to further transmission of *S. pneumoniae* within the community or can advance to pneumococcal diseases [170]. Biofilms form in the nasopharynx during colonization [199]. *S.*

pneumoniae has many virulence factors (Table 3.2; Figure 3.4) that allow for adherence to host cells, reduce the host's immune system's ability to clear the bacterium, and promote invasion of epithelial cells [165]. If the host is unable to clear *S. pneumoniae* immediately after colonization of the upper respiratory tract, the bacterium multiplies, disrupts the regular non-pathogenic flora of the respiratory system [171, 200], and is able to migrate to the tissues and organs and cause infections. The migration of *S. pneumoniae* to sterile tissues and organs is the main cause of all pneumococcal diseases. For example, when meninges, the protective membranes surrounding the spinal cord and brain, become inflamed due to *S. pneumoniae* infection, this is known as bacterial meningitis [173, 201]. Bacterial meningitis is predominantly seen in young children and is mostly caused by *S. pneumoniae* [202]. *S. pneumoniae* causes more than 50% of bacterial meningitis within the US [173, 202]. Bacteremia refers to infection of the blood by pneumococcus [173] which causes about 12,000 cases per year and usually accompanies other pneumococcal infections [173]. *S. pneumoniae* can also colonize the middle ear of infants and young children causing acute otitis media [173]. The Centers for Disease Control (CDC) estimates that approximately 60% of young children would have at least one ear infection [173]. Sinusitis occurs when *S. pneumoniae* infects fluid trapped in the sinuses [173].

Year	1997		2007		2012		2014		2015	
Age	Cases	Deaths	Cases	Deaths	Cases	Deaths	Cases	Deaths	Cases	Deaths
<1	142.9	4.02	40.51	0.9	15.7	0.24	15.9	0.48	18.4	0.24
1	178.7	0.9	32.39	0.23	13.6	0.24	10.3	0	12.9	0.24
2-4	31	0.15	13.03	0.08	5.9	0	6.3	0.08	5.1	0.16
5-17	4.8	0.14	2.91	0.14	1.9	0.14	1.4	0.05	1.3	0
18-34	9.3	0.52	4.19	0.22	2.8	0.1	2.7	0.18	2.5	0.08
35-49	18.9	1.65	11.89	0.98	7.5	0.6	6.6	0.7	6.7	0.5
50-64	23.5	2.72	20.59	2.33	15.9	1.53	15.1	1.64	15	1.53
65-74	61.7	11.02	39.26	6.37	29.6	4.24	19.1	2.41	18.2	2.3
75-84	-	-	-	-	-	-	28.2	3.46	29	4.5
>85	-	-	-	-	-	-	42.6	8.01	45.3	11.56

Table 3.1: Occurrence of pneumococcal diseases (Cases and Death Rates) from 1995 to 2015 as reported by the Centers for Disease Control. Rates are per 100,000 population for Active Bacterial Core surveillance (ABCs) areas

S. pneumoniae, which initially inhabits the mucosal surfaces of the nasopharynx in its hosts

Virulence Factor	Location on <i>S. pneumoniae</i>	Function	Refs
Polysaccharide Capsule	Layer of polysaccharides on cell wall	Allows the bacteria to escape the nasal mucus	[165, 171, 175, 176, 187, 200, 203–207]
		Inhibits phagocytosis by innate immune cells	
		Escapes neutrophil net traps	
		Inhibits complement and recognition by immunoglobulins	
Pneumolysin	Cytoplasmic toxin	Allows adherence and colonization of the nasopharynx	[171, 175, 176, 200, 205, 208–215]
		Binds to membranes with cholesterol	
		Forms pores which cause cell lysis	
		Induces inflammation	
Autolysin (LytA)	Intracellular enzyme produced by Gram positive bacteria	Drives host-to host transmission	[165, 176, 200, 216–219]
		Can activate complement, modulate chemokine/cytokine production	
		Cell lysis	
		Break down peptidoglycan	
PspA	Bound to the cell wall via PCho moiety	Exposes hosts cell to pneumolysin and teichoic acid	[165, 171, 187, 200, 209, 220–224]
		Aids with bacterial colonization	
		Protects against complement system of the host	
		Aids in colonization by adhering to epithelial cell membranes	
PspC	Bound to the cell wall via PCho moiety	Decreases the deposition of the complement	[165, 171, 176, 187, 200, 209, 220–226]
		- Protects against the complement system of the host	
		Binds to receptors such as the human polymeric immunoglobulin A during colonization and invasion the nasopharynx	
		Cell adhesion and colonization of nasopharynx	
PsaA	Surface of the cell wall	Transports magnesium and zinc into the cytoplasm of the bacteria	[165, 171, 187, 200, 209, 220–224]
		Aids in invasion of epithelial cells during nasopharynx colonization	
Other Choline Binding Proteins: LytB, LytC, CbpC, CbpG	Bound to the cell wall via PCho moiety	Promote bacterial colonization of the nasopharynx	[171, 176, 200, 207, 221, 225, 226]
		Modify proteins on cell surfaces and promotes binding to host cell receptors	
		Important for host cell recognition	
Non-Classical Surface Proteins	Surface of the cell wall	Act as adhesins	[227–229]
		Promote immune system evasion by inhibiting complement	
		Controls inflammation and affects cytokine production	
Pili	Cell surface	Promotes adherence and colonization of the epithelial cells within the nasopharynx	[171, 176, 200, 230, 231]
		Inhibits phagocytosis by immune cells	
Bacteriocin	Produced and secreted by the organism	Inhibits the growth of competing bacterial cells	[171, 176, 200]
Neuraminidase	Cell wall bound	Degrades mucus	[171, 176, 200]
		Promotes growth and survival	
Biofilm		Aids with cell adherence	[171, 176, 200]
		Helps to reduce bacterial recognition by the host immune system	
IgA protease	Secreted by the bacteria into the extracellular environment	Reduces the impact of antimicrobial agents on bacteria	[171, 176, 200, 232–234]
		Breaks down IgA	
Lipoteichoic acid	Membrane bound	Causes inflammation	[171, 176, 200]

Table 3.2: Selected virulence factors of *S. pneumoniae*, their location, and function.

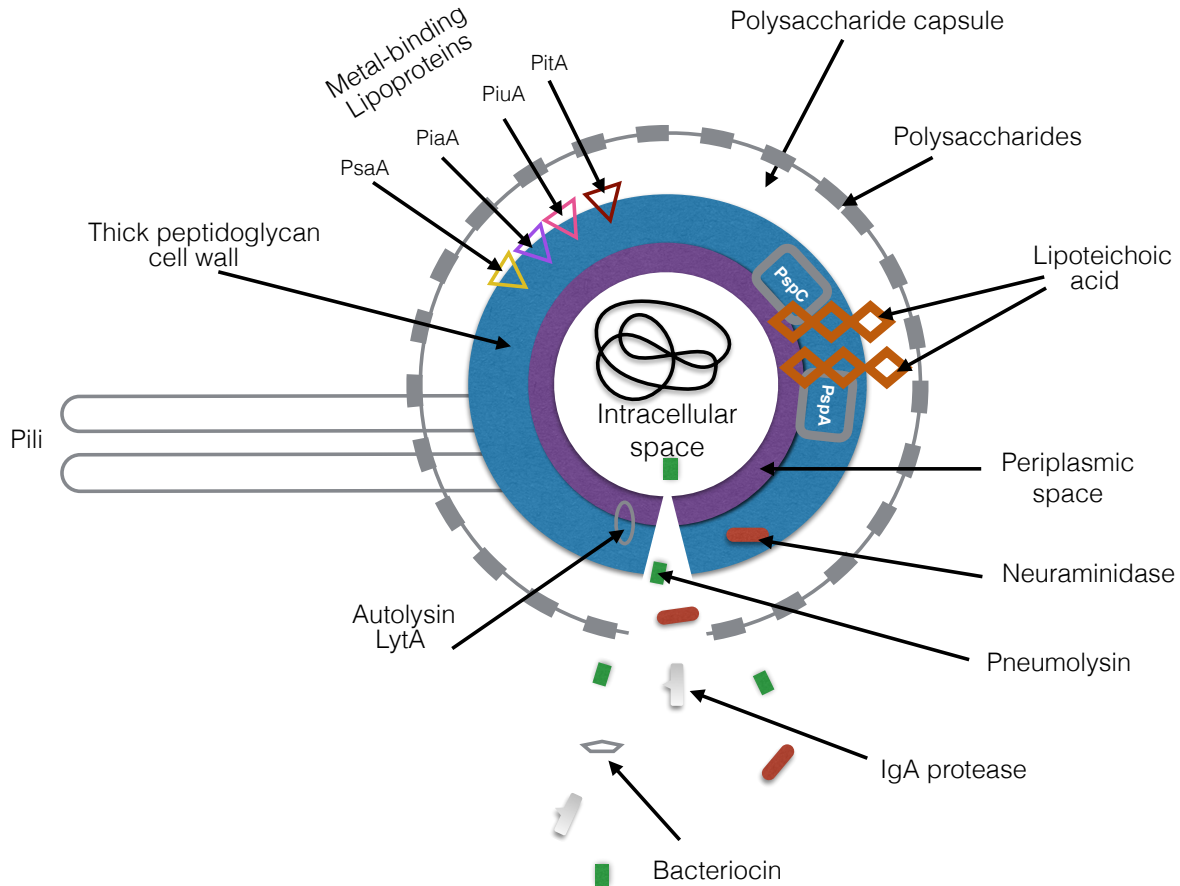


Figure 3.4: Virulence factors of *Streptococcus pneumoniae*. There are a variety of proteins and toxins that are expressed by *S. pneumoniae* that drive its pathogenesis. The major virulence factors are highlighted in the figure. Abbreviations: PsaA, pneumococcal surface adhesin A; PspA, pneumococcal surface protein A; PspC, pneumococcal surface protein C; PiaA, pneumococcal iron acquisition A; PiuA, pneumococcal iron uptake A; PitA, pneumococcal iron transporter.

[165], can migrate to the lungs, where it causes pneumococcal pneumonia [165]. This is an infection of the lungs that leads to inflammation of the air sacs causing them to fill with fluid, and making it difficult to breathe. Individuals who have pneumonia usually suffer with high heart rates, shortness of breath, frequent coughing and high fevers [235]. Thus, despite *S. pneumoniae*'s asymptomatic colonization of the nasopharynx, having a poor immune response and lack of clearance, may develop into pneumococcal pneumonia, which can be a serious health risk for those with reduced host defenses. Pneumococcal pneumonia dominates as the main type of pneumococcal disease within the US and worldwide [173] (Figure 3.5). Overall, pneumonia is the eighth leading cause of

death in the US [236], and is mainly caused by bacteria, but can also be caused by other pathogens such as viruses and fungi [171]. For example, *Haemophilus influenzae* type b, respiratory syncytial virus (RSV), and influenza can also cause pneumonia, but pneumococcal pneumonia is the most prevalent (Figures 3.1, 3.2 & 3.6) [4]. Over time the global disease burden of LRIs such as pneumonia has decreased, but they remain a healthcare concern for specific high-risk populations (Figures 3.2 & 3.6) [4]. Worldwide pneumonia is the leading cause of death in children under the age of five [180, 237]. The World Health Organization reported that a child dies from pneumonia every 20 seconds [238]. There are approximately 900,000 cases of pneumococcal pneumonia that occur annually within the US [182, 239]. In addition, United Nations Children Fund stated that in 2016, pneumonia accounted for 16% of the fatalities observed amongst young children under the age of five worldwide [240]. Pneumococcal pneumonia leads to about 300,000 - 600,000 elderly hospitalizations annually in the US, and the elderly have reduced survival rates [241, 242]. There are different types of pneumonia: community acquired pneumonia (CAP), atypical pneumonia, hospital acquired pneumonia and aspiration pneumonia [173]. These differ based on where someone contracts the infection and what bacteria causes the disease. Currently, the most common form of pneumonia is CAP (which is mostly pneumococcal). This type of pneumonia spreads via person-to-person contact in the community, but outside of healthcare facilities, by breathing in aerosol droplets from a carrier or infected person [200, 235]. Worldwide, CAP is currently the leading cause of death for young children who are under the age of five [178, 243]. In 2015, 920,136 children died from CAP [244]. Infants, young children, the elderly, smokers and immunocompromised individuals are all at a higher risk of developing pneumonia due to a weakened immune system [171]. CAP has a higher occurrence rate in the elderly compared to younger populations, and is also the 5th leading cause of death in the elderly population [241, 242].

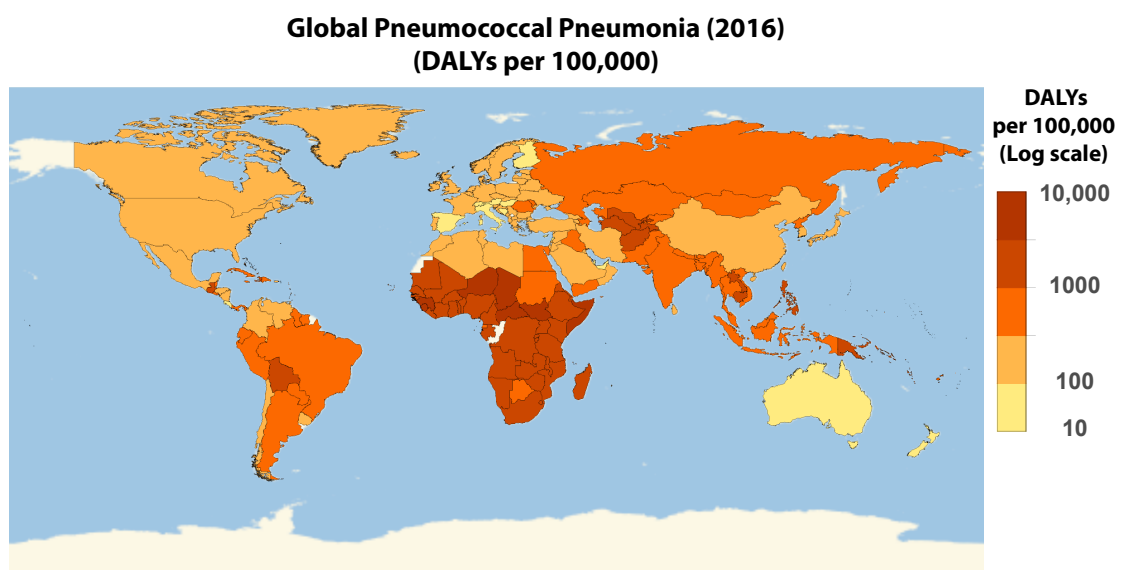


Figure 3.5: Worldwide disability adjusted life year (DALY) of pneumococcal pneumonia. Global distribution of pneumococcal pneumonia on a log₁₀ scale of the 2016 DALY per 100,000 pneumococcal pneumonia data obtained from Institute for Health Metrics and Evaluation [4]

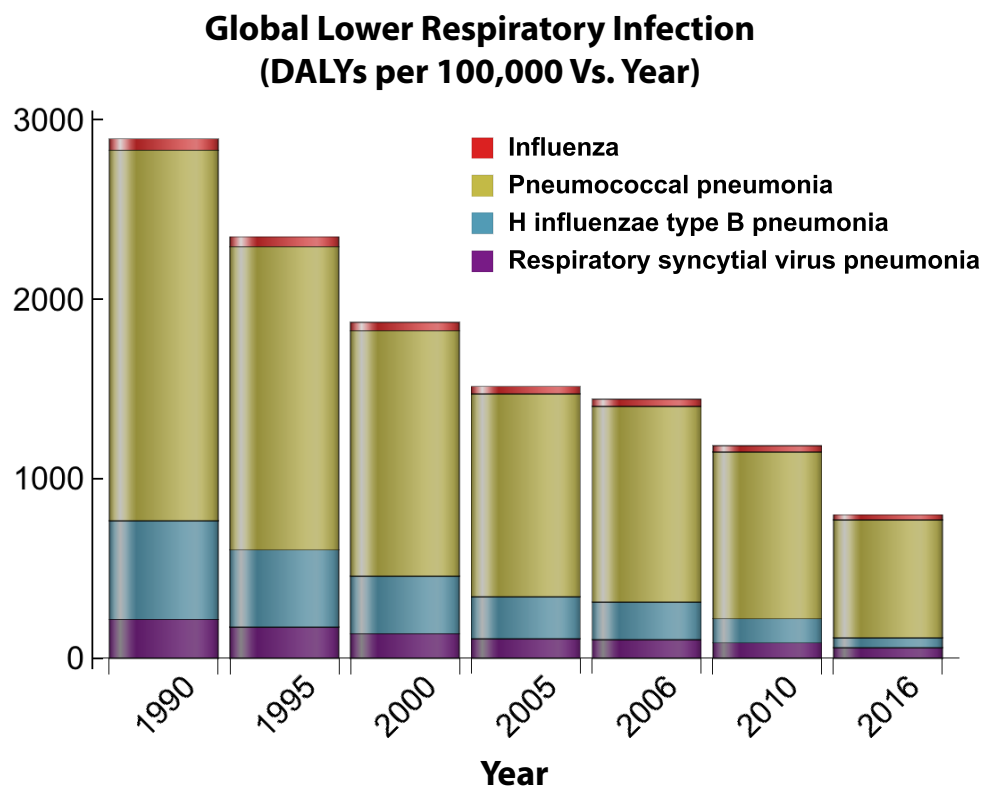


Figure 3.6: Global distribution of lower respiratory infections over time. This figure depicts how the burden for four major lower respiratory infections changes over time in response to the introduction of antibiotic treatments and vaccine implementation. Disability adjusted life year (DALY) data obtained from Institute for Health Metrics and Evaluation [4]

3.4 Transmission

The severity of pneumococcal diseases has led to multiple studies investigating how *S. pneumoniae* is transmitted. The nasopharynx has been classed as the main reservoir of *S. pneumoniae*. This is due to the nasopharynx of hosts being colonized without any symptoms [199]. Following colonization, the spreading of the disease depends on carriers coming into close contact with healthy individuals within the community. The CDC has declared that the main source of *S. pneumoniae* transmission is direct contact with secretions of the respiratory system of a carrier [173]. Le Polain de Waroux et al. [245] investigated transmission in 566 Ugandan subjects by studying nasopharyngeal samples, and determined that close interpersonal contact was necessary for the dissemination of *S. pneumoniae*. Who exactly the main carriers/reservoirs of *S. pneumoniae* are, is still heavily debated. There have been a variety of studies trying to pinpoint which age group acts mainly as carriers/reservoirs for *S. pneumoniae* [246–248]. Some researchers have suggested infants [170, 248], while others suggest that older children actually transmit the pathogen to infants [246, 247]. Lipsitch et al.’s longitudinal study suggests that infants are reservoirs due to the duration of carriage and colonization [248]. In this study, they also observed that the carriage time of *S. pneumoniae* decreases with age [246, 248]. On the other hand, a longitudinal study investigating transmission and colonization in a daycare setting showed that toddlers act as a reservoir for *S. pneumoniae* and spread to family members [246, 247]. Another contradicting study that used pre-existing data and mathematical modeling suggests that older children introduce the pathogen to their homes and transmit *S. pneumoniae* to younger children, siblings and adults [246]. Althouse et al. did confirm that there is higher colonization in infants, however, their results show that *S. pneumoniae*’s direction of transmission is instead from older siblings to infants as opposed to transmission from infants or parents to others in the household [246, 249]. The duration of carriage seems to affect how well *S. pneumoniae* is transmitted as well as close contact between carriers and healthy individuals [245, 246]. Althouse et al., concluded that despite the larger percentage of carriage being in infants, their role in transmission is minimal compared to that of toddlers and older children [246]. The differences between these findings suggest that the direction of transmission is still not yet fully

understood and further research is required. Another possibility would be that multiple age groups are acting as reservoirs rather than one specific group under different conditions. In addition to close contact with a *S. pneumoniae* carrier, the bacterium may also be transferred to healthy individuals via fomites [250]. Chronic carriers of *S. pneumoniae* can contaminate inanimate objects with biofilms [250]. *S. pneumoniae* biofilms are able to survive being in the environment because the biofilm's structure provides protection from drying out [196, 251]. *S. pneumoniae* was found in high concentration on items within a day care center following bacterial cultures [199, 250]. Pneumococcus can survive being in the environment for long periods of time (for example up to 4 weeks) [250, 252]. Because of this, fomites can serve as a reservoir. These findings indicate why it is important to improve hygiene and cleanliness in everyday-life, and at community-based facilities and daycare centers. *S. pneumoniae* also makes a toxin, pneumolysin, that promotes shedding and in turn enhances bacterial transmission [208]. Pneumolysin induces inflammation in hosts during colonization and this promotes bacterial shedding [208]. Zafar et al., conducted a shedding assay which suggests that *S. pneumoniae* may be using the host's inflammatory response as a signal for initiating its exit from the inhospitable host [208].

3.4.1 Transmission Via Coinfections:

Co-infection with *S. pneumoniae* is often seen during viral infections such as influenza, also the 8th cause of death within the US [236], and respiratory syncytial virus (RSV). Co-infections by pathogenic bacteria such as *S. pneumoniae* increase the severity and mortality rates of viral infections [253, 254]. For example, during the influenza pandemic of 1918, the analysis of lung samples from those infected indicated that a majority of the deaths were due to bacterial infections and not the influenza virus [254–256]. Co-infection is possible due to the pre-existing damage on the epithelia of the respiratory tract which promotes bacterial colonization [257–260]. More specifically, *S. pneumoniae*'s bacterial load increases during viral coinfections due to the bacteria's attachment to cells that are already infected by the virus [261]. Studies have also shown that colonization of *S. pneumoniae* is affected by flu vaccines, which also indicates that *S. pneumoniae*

benefits from colonizing hosts that are already compromised [254, 262]. Increased host colonization and bacterial cell density of *S. pneumoniae* during viral infections promote transmission [262]. Khan et al determined that there are higher risks of bacteremia, mortality and spread to other tissues during coinfections [257, 262]. Co-detection with *S. pneumoniae* has also been observed in RSV infections [263].

3.5 *S. pneumoniae*'s Virulence Factors

S. pneumoniae, like many other bacterial species, produces toxins that are harmful to its host, has several surface proteins and physical structures, which play a vital role in its pathogenesis [176]. These virulence factors (Figure 3.4, Table 3.2) work by hindering the host's immune system response, avoiding defense mechanisms, or by direct contact with host tissues and surface receptors, which in turn interferes with the host's immune system activation and bacterial clearance [176]. As discussed above, *S. pneumoniae* exploits hosts with weakened or compromised immune systems [162, 163, 264]. *S. pneumoniae*'s effectiveness in causing infections is directly related to the host immune system's developmental stage and possible deterioration with aging (see also Section 2.6)

S. pneumoniae's virulence thrives because of the bacteria's ability to acquire new genetic material via transformation and recombination [265]. Investigating the level of genetic variation within *S. pneumoniae* is important for not only thoroughly understanding its virulence, but also for developing effective treatments and vaccines. About 4,000 *S. pneumoniae* genomes have already been sequenced [265], with lengths ~ 2-2.2 million base pairs (bp) [203]. More than 2000 genes have been annotated, but novel genes are still regularly discovered as more sequences become available [265]. Variation in gene content and single genes plays a role in defining the virulence profile of some of *S. pneumoniae* strains [265]. Donati et al., describe genome diversification as *S. pneumoniae*'s ability to evolve in diverse host environments [265, 266]. Genetic variation has been observed within identical *S. pneumoniae* clones, due to changes in gene content of their dispensable genes [265, 267]. Dispensable genes are not needed for bacterial growth [265], but provide selective advantages to *S. pneumoniae* such as antibiotic resistance [268]. Additional variants are introduced

to the core genome of *S. pneumoniae* via allele replacement. This is because the bacteria lacks SOS genes and does not repair damaged DNA [269]. Carriage can also influence genetic variation. In 2017, Lees et al., developed a model to assess carriage duration and assembled those findings with data from whole genome sequencing. The results indicated that pneumococcal genetic variation accounts for the phenotypic variation compared to host's age and previous carriage(5%) [270].

The major virulence factors of *S. pneumoniae* that have been thoroughly characterized are summarized in Table 3.2. Below we further discuss virulence factors of particular interest:

1. **Polysaccharide Capsule:** *S. pneumoniae*'s extracellular polysaccharide capsule, the most important virulence factor [204], helps to initiate infection by allowing the bacterium to adhere to host cells and cause inflammation, while also providing protection from the host's immune system [203, 204]. The capsule inhibits phagocytosis by innate immune cells, prevents the recognition of the bacterium by host receptors and complement factors, and also avoids neutrophil traps [165, 176, 204, 205, 220, 271]. Many serotypes of *S. pneumoniae* are characterized by the polysaccharides that are on the outer coat of the capsule, and they are all pathogenic in their own unique manner – some more harmful than others [168, 205]. For example, serotype 1 has been found in invasive infections which have lower fatalities whereas, serotype 3 is associated with colonization of the nasopharynx and serious infections which can lead to fatalities [188, 191, 205, 272, 273]. The capsule manipulates how immunoglobulins recognize the bacteria [274] and inhibits the host's defenses such as mucus layers and cilia from removing the bacterium, and is vital for pneumococcal bacterial cells' colonization [206]. The roles of the capsule in pathogenesis have been described to be due to its charge [206, 275]. The capsule has a negative net charge which is in part due to the acidic polysaccharides and phosphates that make up this layer [206, 275]. The charge is important because it defines how interactions with other cells take place, specifically host cells [206, 275]. One explanation for *S. pneumoniae*'s ability to avoid being trapped by mucus layers and phagocytic cells is due to electrostatic repulsion [206, 275]. Negatively-charged mucus and phagocytic cells, such as macrophages, have led to a reduction in the clearance

of *S. pneumoniae* because of this electrostatic interaction [206, 275]. *S. pneumoniae*'s virulence via their polysaccharide capsule is enhanced by its ability to undergo capsule switching [265, 276, 277]. Mutations in the capsule polysaccharide synthesis genes (cps) promote serotype switching [265, 276, 277]. Serotype switching in strains is increasingly being observed and it is often via recombination or polymorphisms based on antibiotic and vaccine selective pressures (further discussed in Section 3.7.2) [265, 276, 277]. Currently, serotype switching is a healthcare concern as non-vaccine serotypes are being detected at higher rates compared to before vaccines were implemented [276]. Moreover, mutations in novel genes or a disruption of the cps loci can lead to *S. pneumoniae* strains without capsules [278]. Non-typeable *S. pneumoniae* cannot effectively colonize hosts, but novel genes such as pneumococcal surface protein K in the cps loci assist with adhesion [278]. Serotype switching and capsule-free strains of *S. pneumoniae* together will add to the burden on the high risk age groups (infants, and the elderly) [278], and because of this vaccines and treatments should be improved.

2. ***S. pneumoniae*'s Cell Wall Components:** *S. pneumoniae* is a gram-positive bacterium with a thick cell wall. The cell wall is important because it provides protection and shapes the cell [279]. Peptidoglycan, wall teichoic (WTA) and lipoteichoic acids (LTA) are the main components of *S. pneumoniae*'s cell wall [279]. WTA are covalently attached to peptidoglycan whereas LTA are non-covalently connected to the cytoplasmic membrane with a lipid anchor [279]. The capsular and cell-surface proteins are all linked to the peptidoglycan [279]. Alternating glycan chains of N-acetylglucosamine (GlcNac) and N-acetylmuramic (MurNac) acids crosslinked by peptides make up peptidoglycan [279, 280]. These glycan chains can undergo secondary modifications such as deacetylation of GlcNac and O-acetylation of MurNac [279, 280]. These modifications aid in *S. pneumoniae*'s virulence by making the cell resistant to lysozyme [280]. Cell wall components, WTA and LTA have phosphorylcholine (PCho) residues which serve as anchors for choline binding proteins. Choline-binding proteins are important for host-pathogen interactions such as

evasion of host immune responses (discussed later in this section) [279, 280]. PCho in bacterial cells is unusual and *S. pneumoniae* are currently the only bacteria known to require it for growth [280]. WTA, LTA and peptidoglycan are pathogen associated molecular patterns that can cause an inflammatory response in hosts. Peptide synthesis, peptidoglycan structure, WTA and LTA synthesis and modifications have been further discussed by Gisch et al. [279].

3. **Pneumolysin:** This toxin that is capable of forming pores in cell membranes [281] can be found in the cytoplasm of *S. pneumoniae* and other Gram-positive bacteria [176, 209, 281]. Pneumolysin is released as a result of cell lysis and is toxic to host cells [176, 210, 282]. Pneumolysin binds to membranes containing cholesterol [283], and forms pores which later lead to host cell lysis [176, 205, 211]. In addition to causing cell lysis, pneumolysin plays a role in promoting the formation of biofilms [212], it reduces mucus clearance of the bacterium, and it can interfere with the host's immune system [176, 209, 210, 282, 284]. Pneumolysin regulates the complement system [203] and reduces phagocytosis by innate immune cells. It is also a pro-inflammatory toxin which causes damage to host cells. It can regulate cytokine and chemokine production [171, 200]. This pro-inflammation has also been shown to assist with host-to-host transmission [208]. By increasing cell inflammation, there is an increase in shedding and thus a higher rate of transmission of the bacteria [208]. Studies have also shown that pneumolysin can cause DNA damage by inducing double stranded DNA breaks. One mechanism of DNA damage by pneumolysin was described by Rai et al. in 2016 [213]. They showed that the toxin can dysregulate the production of reactive oxygen species (ROS) intracellularly [213]. This is possible because of pneumolysin's pore-forming properties – it creates ion channels that disrupt cell calcium levels, which leads to overproduction of ROS, that then causes DNA damage [213]. Host DNA damage may lead to increased pneumolysin virulence in the elderly, who are already experiencing a compilation of DNA damage and telomere shortening due to aging [285]. Pneumolysin has different allelic forms that and can also affect the toxin's hemolytic activity [203, 286]. For example, genetic variation in allele 5 produces a non-hemolytic form of pneumolysin [286–289]. Previously, a cysteine residue

at amino acid position 428 in the conserved sequence was described to be important for the hemolytic activity of pneumolysin [290]. However, cysteine was later substituted by alanine without affecting the toxin's hemolytic activity [286, 291].

4. **Autolysin:** This enzyme is involved in autolysis of bacteria which results in the release of pneumolysin, teichoic acid and other components from within the cell [171, 200]. An example of this is LytA [216], a choline binding amidase [165] (see below) that degrades peptidoglycan and causes cell lysis [176, 217, 218, 292]. Autolysins promote colonization of nasopharyngeal cells due to the release of toxins such as pneumolysin during cell wall degradation [176].

5. **Pneumococcal Surface Proteins** *S. pneumoniae* has a large variety of surface-exposed proteins [165, 221] that aid in its pathogenesis by acting as adhesins to host cells and hindering the host's immune system, specifically the complement system [171, 176, 200, 293, 294]. Pneumococcal surface proteins are categorized into four groups: choline binding proteins (CBPs), lipoproteins, non-classical proteins, and proteins that have an LPXTG motif (X represents any amino acid) and can be covalently bound through sortase cleavage of the motif [165, 221].

a) **Choline Binding Proteins(CBPs):** Many of *S. pneumoniae*'s surface proteins are classed as choline binding proteins [176, 221, 225, 226]. These proteins are known for binding to phosphorylcholine on *S. pneumoniae*'s cell wall [176, 221, 225, 295], and are necessary for adhesion to host cells [176, 225, 295]. Choline binding proteins affect the host's complement system by blocking its activation and reducing the ability of immunoglobulins to eliminate the pathogen [176, 221]. Some of these choline binding proteins can also modify host cell surfaces to allow for binding interactions between to host cell receptors and *S. pneumoniae* [225]. *S. pneumoniae* has approximately 10 to 16 identified CBPs [165, 296–298] including pneumococcal surface protein A (PspA), pneumococcal surface protein C (PspC) and Lytic Amidase (LytA) which are discussed

below:

- i. **PspA** is very electronegative, and this characteristic can block complement binding, which prevents opsonization of *S. pneumoniae* [165, 222]. PspA can also bind to host lactoferrin [171, 176, 200], specifically apolactoferrin (iron-free), which in turn provides protection to *S. pneumoniae* against the bactericidal killing of apolactoferrin [299, 300].
- ii. **PspC**, also known as CbpA (highly polymorphic), promotes adherence by binding to the polymeric immunoglobulin receptor [221, 301]. It facilitates the colonization of *S. pneumoniae* into the nasopharynx and can prevent the formation of C3b (part of the complement system) by binding to factor H. This in turn interferes with opsonization of *S. pneumoniae* [165, 221, 302]. PspC exists in multiple allelic forms with most alleles containing a C-terminal cell wall choline binding motif. However, there are also 17 allelic variants that have the LPTXG motif (see LPXTG cell wall bound proteins in Section 2.5) [203, 303, 304]. Additionally, allelic variant PspC 4.4 was characterized as a ligand for complement inhibitor C4b-binding protein [203, 305], which leads to an allele-dependent form of protection from the complement [305].
- iii. **LytA**, an autolysin, was the first of 3 major lytic enzymes found in *S. pneumoniae* [225, 306]. LytA degrades peptidoglycan by cleaving the N-acetyl-muramoyl-L-alanine bond [165, 221]. This causes cell lysis and the release of pneumococcal antigens such as pneumolysin, peptidoglycan and teichoic acids which are all harmful to host cells [165, 221, 307]. The release of these harmful particles from *S. pneumoniae* cells is also capable of inhibiting cytokine (such as IL-12) production, which in turn blocks the activation of phagocytes [306, 308]. This is thought to be due to the fact that cells are already decomposed so phagocytic activity is no longer necessary [306], and acts as a form of immune system evasion by *S. pneumoniae* [308–310]. By blocking cell signaling via cytokine production,

LytA has also been shown to hinder complement activation [225]. How exactly this blockade might be happening still needs to be further researched.

- iv. **LytB and LytC** are two other lytic enzymes found in *S. pneumoniae*. Their roles in *S. pneumoniae*'s virulence is not as thoroughly understood as LytA. Studies have shown that LytB is necessary for separating daughter cells [311, 312]. LytC on the other hand, is described as a lysozyme. Ramos-Sevillano et al. have indicated that LytB and LytC interact and are both involved in adhesion of *S. pneumoniae* to epithelial cells within the nasopharynx of hosts [313]. Their results also suggest that LytC helps with evasion of the complement system via experiments with mutants [313]. LytC mutants had larger amounts of C3b deposition and LytB and LytC double mutants all had a reduction in their ability to adhere to host cells [313]. These findings shed light on the roles of LytB and LytC. They aid in virulence by playing a role in colonization and evasion of host immune responses [313, 314]. Additionally, LytC has also been described to play a role in cellular fratricide with LytA. These enzymes are released to lyse non-competent pneumococci in close proximity of competent cells [315]. This is important for transformation of *S. pneumoniae*. Competent cells are able to uptake and incorporate free DNA from the lysed cells [315, 316]. This promotes genetic exchange which in turn can improve bacterial survival. For example, the bacterium can take up genes for antibiotic resistance [315, 316]. LytC's activity is most active at 30 degrees Celsius which indicates it is probably most active in the upper respiratory tract [315, 316].
- v. **CbpF**, the most abundant protein on *S. pneumoniae*'s cell wall is capable of regulating LytC [226, 317]. CbpF regulates LytC's activity by blocking LytC's access to its substrate [226, 298, 317].
- vi. **Other Choline binding protein:** There are about 8 other choline binding proteins: CbpD, CbpG, CbpI, CbpJ, CbpK, CbpL, CbpM, CbpN. These have not been studied as extensively as the main choline binding proteins previously discussed. There

is not much known about their structure or function. CbpD has been shown to be involved in fratricide by working with LytA and LytC [298, 316, 318]. The CbpD is able to provide a substrate for LytC that is more accessible by binding to target cells and breaking down the cross-links of the peptidoglycan [298]. CbpG is necessary for adhesion and all others been reported to work as adhesins [226, 298].

b) **Lipoproteins:** These proteins are necessary for substrate transport. There are approximately 50 lipoproteins that have been characterized [221, 296, 298]. The four main lipoproteins are the pneumococcal surface antigen A (PsaA), pneumococcal iron acquisition A (PiaA), pneumococcal iron uptake A (PiuA), and pneumococcal iron transporter (PitA) [165, 221, 319, 320]. They are all metal-binding proteins that combine with ATP-binding cassette (ABC) transporter complexes. ABC transporters transport substrates across membranes by utilizing energy generated from ATP binding and hydrolysis.

i. **PsaA** is involved in transporting magnesium and zinc into the cell [176, 223, 321, 322]. Investigations have previously reported PsaA's role in cell adhesion and promoting cell invasion of *S. pneumoniae* [223, 323]. However, other studies on PsaA mutants have found that PsaA has no clear role in adhesion, but rather magnesium transport [322]. This particular characteristic of adhesion needs to be further investigated [175]. Also, genetic mutations can alter PsaA's function which may lead to impaired ability to acquire manganese which results in decreased resistance to oxidative stress [203].

ii. **PiaA, PiuA and PitA** are involved in regulating iron-uptake [324, 325]. In addition to this, PiaA and PiuA have been described to be needed for full virulence of *S. pneumoniae* in mice [324, 326]. Mutations in PiaA and PiuA affect growth and virulence of *S. pneumoniae* [327, 328]. This indicates the importance of iron in the environment for growth. Furthermore, Cheng et al., in 2013 crystalized PiaA and discovered that PiaA is capable of binding to ferrichrome [329–331] despite previous findings suggesting pneumococci do not produce siderophores

[332]. Cheng et al., concluded that *S. pneumoniae* is probably able to acquire holo-siderophores from other bacteria within the host [328, 329]. On the other hand, PiuA binds to both hemin and hemoglobin but has greater affinity for hemin [330, 333]. PitA was later discovered and characterized to bind to ferric irons [319, 320, 330]. A novel iron transport was discovered in 2016 by Yang et al., via proteomics [334]. In this study, they constructed a triple mutant by deleting PiaA, PiuA and PitA [334]. Using this mutant, they were able to identify potential iron transporters, such as putative protein SPD-1609, which functions similarly to PitA via translationalomics and proteomics [334]. These findings suggest that there are potentially more iron-binding proteins in *S. pneumoniae* to be discovered and that the bacteria have developed transport systems to ensure they have access to as much iron as possible for their survival.

- c) **LPXTG cell wall bound proteins** are recognized by the sortase of the cell wall [203, 297, 335]. Sortase recognizes the LPXTG sequence, cleaves at this site and anchors the proteins to the cell wall [203, 335, 336]. Mutating the sortase gene *srtA* caused a decrease in *S. pneumoniae*'s adhesion to host nasopharyngeal cells in vitro, and caused neuraminidase to be released from the cell wall into the media [337]. Neuraminidase is an example of a LPXTG cell wall bound protein and is known for cleaving sialic acid from glycoproteins. In the case of pathogenesis of *S. pneumoniae*, this activity can lead to the removal of sialic acid from lactoferrin which hinders lactoferrin's bactericidal effect. Neuraminidase is secreted from *S. pneumoniae* cells and targets host cells [203, 335]. It is also involved in colonization of the host and has been suggested to be involved with adhesion [337, 338].
- d) **Non-classical Surface Proteins (NCSPs)** are found on *S. pneumoniae*'s surface, but do not have a membrane-anchoring motif nor a leader peptide [221]. They are also known as moonlighting proteins for having multiple functions [221, 227, 296]. NCSPs function as adhesins that are able to bind to host molecules which promotes pneumococcal host

cell invasion [296]. There are two main NCSPs: pneumococcal adherence and virulence factor A (PavA) and glycolytic enzymes (enolase and GAPDH) [338]

i. **PavA** attaches to fibronectin and assists with adherence to host cells [297]. PavA also provides protection to pneumococci by controlling inflammation and inhibiting recognition by dendritic cells [339]. PavA mutants were more susceptible to recognition and phagocytosis by dendritic cells compared to wildtype [339]. In addition to this, when the dendritic cells encountered PavA mutants there was a reduction in cytokine production, which affected the adaptive immune response. These findings characterize PavA's potential function in immune system evasion by *S. pneumoniae* and cytokine production by dendritic cells [339].

ii. **Enolase & GAPDH** are both plasminogen binding proteins. Enolase is an anchorless protein found at the surface of *S. pneumoniae* [340]. It is important for proteolytic activity on the cell surface [341], which is necessary for the pathogenesis of *S. pneumoniae* [340]. Enolase also promotes complement system evasion by binding to the complement inhibitor C4b-binding protein [342]. Additionally, studies suggest that enolase may cause host tissue damage by inducing the production of neutrophil extracellular traps (NET) by binding to neutrophils [343]. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) can be found on the surface and in the cytoplasm of *S. pneumoniae* [344]. Although GAPDH binds to plasminogen, it has a higher affinity for plasmin [297, 344]. GAPDH is suggested to also play a role in iron acquisition due to its ability to bind to hemoglobin and heme [297]. Like enolase, GAPDH may also play a role in host cell invasion and evasion of the immune system. LytA has recently been identified to be involved in the delivery of GAPDH to *S. pneumoniae*'s cell surface [228].

6. **Pili:** These hair-like structures are located on the cell surface of *S. pneumoniae* and many other bacteria [176, 200, 230]. They assist with *S. pneumoniae*'s attachment and colonization of epithelial cells within the nasopharynx and lungs of hosts [200, 230, 231]. These pili

also help the bacteria avoid phagocytosis by host immune cells [171]. There are two main types of pili found on *S. pneumoniae*: Pilus-1 and pilus-2. Pilus-1 is found in 30% of clinical isolates [345] whereas pilus-2 is only in about 16% [346]. Studies have shown that piliated *S. pneumoniae* induce higher tumor necrosis factor (TNF) responses than the non-piliated during pneumococcal infection [230]. This suggests that pili are able to stimulate inflammatory responses of the host [230]. Pancotto et al.'s findings indicated that pilus-1's expression is regulated in vivo [231]. High expression of pilus-1 is observed at early stages of colonization and reduced expression during later stages of infection. This down regulation may be necessary for avoiding host immune response but this needs to be further investigated as it is not clear why this might be happening [231]. *S. pneumoniae*, like many other pathogenic bacteria have a type IV pilus that is necessary for transformation [347, 348]. This pilus is formed on the surface of the bacterial cell and contains the major pilin ComGC. The operon that codes for ComGC also encodes for an ATPase which is needed for powering the pilus assembly. The structure of ComGC was recently discovered by Muschiol et al., in 2017 [347].

7. **IgA1 protease:** This enzyme is produced by *S. pneumoniae* and it works by cleaving the human immunoglobulin A1 (IgA1) into fragments [232, 233]. The IgA1 represents an isotype of IgA which has two isotypes: IgA1 and IgA2 [349]. These two isotypes differ in hinge regions – IgA1 has an extended hinge region because of an insertion into this region of a set of duplicated amino acids [349]. IgA1 proteases reduce the binding IgA's effector region of the heavy chain and hinder killing of the bacterium by these antibodies [232, 234].
8. **Hydrogen peroxide:** *S. pneumoniae* secretes hydrogen peroxide (H₂O₂) which causes damage to host DNA [350]. However, this is only observed in strains with pyruvate oxidase activity (SpxB gene) [350, 351]. H₂O₂ production also has bactericidal effects. *S. pneumoniae* uses this to reduce the growth of bacteria it may be competing with [351]. Additionally, pneumococcal H₂O₂ induces an innate immune response by enhancing the release of pro-

inflammatory cytokines, and targets cellular stress responses [352]. As *S. pneumoniae* produces H₂O₂ via pyruvate oxidase, hydroxyl radicals form via the Fenton reaction [353]. These radicals are often harmful to bacteria but do not affect *S. pneumoniae*. This is because of *S. pneumoniae*'s ability to reduce reactive OH before it comes into contact with DNA [354], by sequestering Fe²⁺ away from DNA [354]. In addition to producing H₂O₂, the SpxB gene has also been found to increase resistance to H₂O₂ [355]. SpxB mutants produced no H₂O₂ and were less resistant [355]. Additionally, *S. pneumoniae* has a variety of defense proteins involved in detoxification, repair, regulation and cation homeostasis that provide protection against oxidative stress [354].

9. **Pathogenicity Islands:** These are parts of pathogenic bacterial genomes that were acquired via horizontal gene transfer [356]. The genes on pathogenicity islands (PAI) aid in the virulence of the pathogen [357]. PAIs can code for iron-uptake systems and proteins involved in cell attachment [357]. For example, the first PAI discovered in *S. pneumoniae*, pneumococcal pathogenicity island 1 codes for the PiaA iron transporter complex [327]. Additionally, the pilus-1 is encoded by another PAI, known as the rlrA islet [230]. However, this PAI is not found in all of the *S. pneumoniae* clinical isolates [230]. Pilus-2 is also encoded by a PAI, pilus islet 2 (PI-2) [346]. Another important adhesin, pneumococcal serine-rich repeat protein (PsrP), is also coded for by a PAI. PsrP is important for *S. pneumoniae*'s attachment to cells within the lungs [358]. High PsrP production is also linked to biofilm growth [359]. PAIs promote genetic variation in species, and this may affect current treatment and vaccine targets.
10. **Biofilms:** These are structured communities that consist of aggregated microbial cells surrounded by an extracellular matrix of polysaccharides that attach to surfaces [196, 360]. The extracellular matrix provides protection and enhances *S. pneumoniae*'s virulence [196, 360]. Biofilms are formed in response to stress and harsh conditions to promote bacterial survival [196, 360, 361]. To promote biofilm formation and competence, *S. pneumoniae* downregu-

lates expression of capsular proteins [362]. Within biofilms, horizontal gene transfer rates increase due to close cell proximity [196, 360, 363, 364]. Studies indicate that *S. pneumoniae* biofilms are not effectively cleared during antimicrobial treatments due to increased antimicrobial resistance [365]. In addition, *S. pneumoniae* biofilms are able to escape host immune responses such as mucociliary clearance [366].

3.6 Host Immune System Responses to *S. pneumoniae*

We have discussed above the virulence factors (Section 2.5) that aid in ensuring *S. pneumoniae* can evade the host's immune system. On the other hand, there are several host defenses that recognize *S. pneumoniae*, act rapidly, and clear the pathogen before it can cause pneumococcal diseases. Protection from *S. pneumoniae* is dependent on the state of the host's immune system. Age plays a role in how successful the immune system will be at clearing the infection by *S. pneumoniae*. Children under the age of five and the elderly are at higher risk for contracting pneumococcal diseases (Figure 3.1). This is due to infants having a naïve immune system, whereas the elderly are experiencing immunosenescence [177]. A variety of immune cells are involved in the innate (first-line of defense) and adaptive immune responses. The most important immune cellular and humoral components for defending against pneumococcal infections (Figure 3.7) are summarized in the following sections including how aging may affect their ability to defend the host.

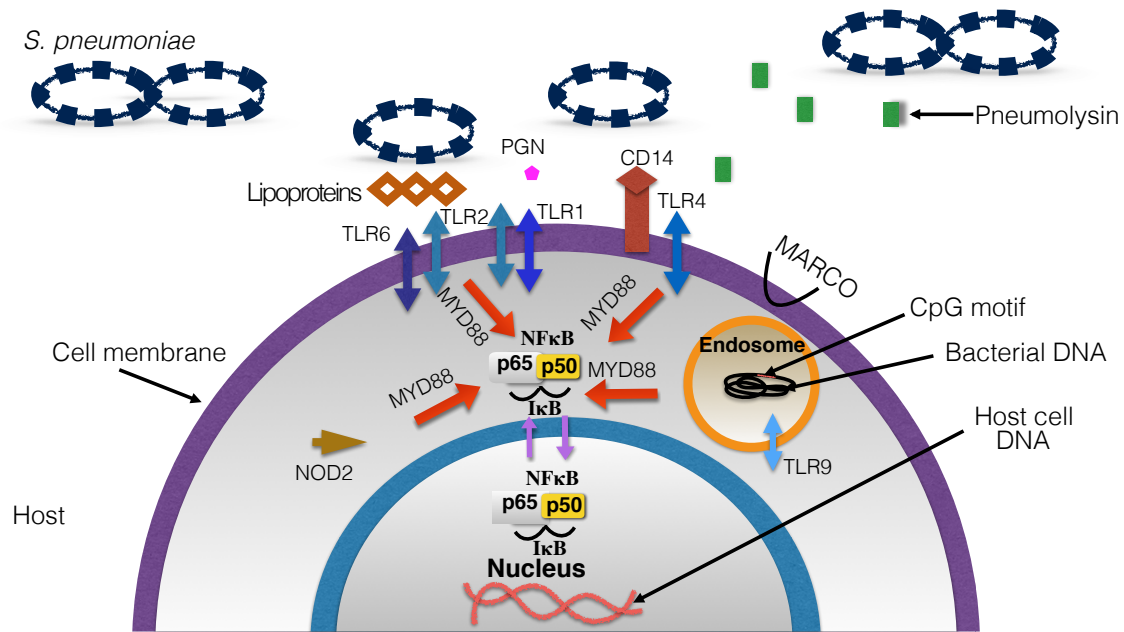


Figure 3.7: Host surface and intracellular receptors necessary for immune response to *Streptococcus pneumoniae*. Highlighted in this figure are the major pathogen recognition receptors necessary for binding to pneumococcal ligands and eliciting an immune response. Upon binding to the ligands, receptors and signaling pathways are activated, which leads to the overall production of inflammatory cytokines and recruitment of immune cells. There are 10 toll-like receptors (TLRs) that have been discovered in humans—TLRs involved in pneumococcal disease are depicted in the figure

3.6.1 Innate Immune Responses

Innate immunity involves nonspecific immune responses – cells and receptors recognize foreign particles and elicit immune responses to eliminate the invaders that can be harmful to the host [164, 367, 368]. Cell-related innate immune responses against pneumococcal infection include:

1. **Mucosa & Respiratory Epithelial Cells:** Epithelial cells provide a protective barrier for tissues and organs [369]. In this case, they line the respiratory tract and protect against pneumococcus [369]. There are epithelial cells known as goblet cells, which secrete mucus [370]. The negatively charged mucus is necessary for maintaining moisture and trapping foreign particles and pathogens. Additionally, ciliated epithelial cells function simultaneously with the mucus to clear pathogens. This process is known as mucociliary clearance [206].

Once the pathogen is trapped in the mucus, the cilia (hair-like structures) move together to direct the trapped pathogen and the mucus to the mouth, for expelling the pathogen via coughing or swallowing [371]. The respiratory epithelial cells can recruit other cells by producing and releasing cytokines and chemokines [171, 370]. They also can directly kill pneumococcus by secreting antimicrobial peptides such as defensins, human apolactoferrin and lysozyme (Figure 3.8) [171, 370, 372]. Human apolactoferrin sequesters iron and lyses cells. Lysozyme also lyses cells and acts as a bactericidal [372]. D-alanine in teichoic acids of *S. pneumoniae*'s cell wall help to evade killing by antimicrobial proteins (positively charged) by reducing the negative charge [373, 374]. The negatively charged capsule also promotes evasion of mucus via electrostatic repulsion [375]. Mouse model experiments showed that encapsulated *S. pneumoniae* were easily trapped in mucus and unable to migrate to the epithelial cells when compared to capsulated *S. pneumoniae* [206] This again was due to the negatively charged capsule. In addition to this, *S. pneumoniae*'s neuraminidase degrades mucin and reduces the negative charge by removing sialic acids [375–377]. As previously mentioned in Section 2.5, the structure of peptidoglycan can be modified. This modification promotes resistance of *S. pneumoniae* to lysis via the lysozyme [378]. Another impressive method of evasion by *S. pneumoniae* is its ability to undergo phase variation [379, 380]. *S. pneumoniae* is able to express a thick and a thin capsule under certain conditions [379, 380]. The thick capsule is necessary to avoid entrapment in mucus, and the thin capsule is necessary for binding directly to epithelial cells [379, 380]. Once the thin capsule is expressed, adhesins are exposed for binding to the glycoconjugates on epithelial cells [379, 380]. Infants and the elderly both are challenged with mucociliary clearance due to different reasons: In infants, immature submucosal glands, surface epithelial secretory cells and low numbers of ciliated epithelial cells can result in poor mucociliary clearance [381]. In the elderly, as the host ages there is a deterioration of mucociliary clearance, with reduced mucin and slower cilia beat frequencies [177, 382], which promotes dissemination of the bacteria [177]. As *S. pneumoniae* virulence factors can also degrade mucus and slow

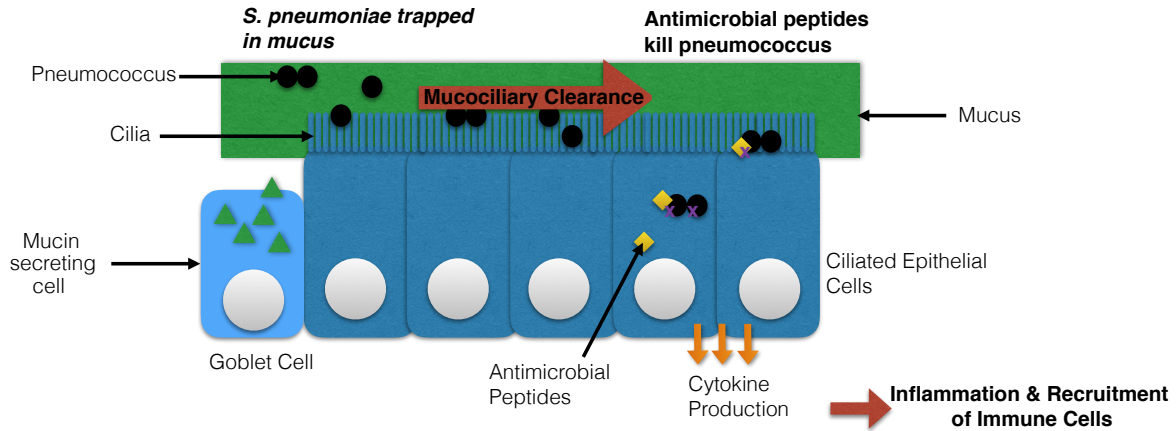


Figure 3.8: *Streptococcus pneumoniae*'s interaction with host epithelial cells. Two types of epithelial cells are depicted: goblet cells and ciliated epithelial cells. The cilia on the epithelial cells together with the mucus produced by goblet cells clear the pathogen via mucociliary clearance. Epithelial cells can also secrete antimicrobial peptides that directly kill *S. pneumoniae* or produce cytokines, which leads to a state of inflammation and the recruitment of immune cells.

down cilia [177], immaturity and deterioration of mucociliary clearance could cause disease exacerbation through increased colonization and recurrent infections.

2. Phagocytes:

- a) **Neutrophils:** These are found in larger concentrations compared to any other white blood cells, and they are generally the first to travel to the infection [383, 384]. Neutrophils are phagocytic cells [384] that also produce granules, which break down the cell walls of pathogens ultimately killing them [384]. There are two main types of granules produced by neutrophils: primary and secondary, which differ based on age/maturity of the neutrophil [369, 385]. Primary granules include defensins whereas secondary granules include enzymes necessary for digestion, such as lysosomes. Neutrophils can also trap *S. pneumoniae* extracellularly, by using extracellular fibers made up of DNA [386]. Neutrophil response changes with age: Infants experience minimal protection by neutrophils in their early days of life due to poor bactericidal function, impaired phagocytotic activity, low response to inflammatory signals, and reduced chemotaxis [387–389]. With age, neutrophil activity improves and strengthens in young adults but later starts to deteriorate. Elderly populations experience impaired chemotaxis, which may lead to the overproduction of proteases by neutrophils. This causes an increase in inflammation levels in older subjects [264, 390]. Neutrophil extracellular traps generation, phagocytosis and killing diminishes with age [264, 391].
- b) **Macrophages:** Macrophages are derived from monocytes [369] and function as phagocytic cells that engulf and directly kill *S. pneumoniae* [164, 367]. These cells can recruit other immune cells, such as neutrophils via cytokine signaling [392], and remove dead neutrophils [368, 393] and other cells via phagocytosis and apoptosis. Macrophages attack cells that have been opsonized by the complement system and Fc γ receptors [375]. The macrophage receptor with collagenous structure (MARCO), found on the surface of macrophages, aids with the phagocytosis of non-opsonized antigens [394]. Macrophage activation due to *S. pneumoniae*'s presence is dependent on pattern recognition receptors [375]. For example, Toll-Like receptors 2 and 4 work together to activate macrophages in the presence of pneumococci [375]. At birth, macrophage

levels are low with impaired phagocytosis, cell signaling and Toll-like receptor 4 (discussed in Section 3.6 3a) expression[387, 395]. Within days post birth, macrophage levels and function improve to reach adult levels [264]. In contrast, with old age alveolar macrophage concentrations are depleted, cytokine production and phagocytotic activity are reduced, and lowered expression of MARCO contributes to poor killing of *S. pneumoniae* [177, 264, 387].

3. Pattern Recognition Receptors (PRRs): These receptors can be found on host cell surfaces that recognize pathogen associated molecular patterns (PAMPs) [164, 369], PRRs can also be located intracellularly or be secreted [164]. PAMPs are structures found in bacteria and viruses. Many of these are necessary for virulence in pathogens. There are two main types of receptors that participate in the host's immune response to pneumococcus: Toll-like receptors (TLRs) and nucleotide-binding oligomerization domain (NOD) –like receptors (NLRs) as described below:

a) Toll-Like Receptors (TLRs): TLRs are mostly found on cell surfaces as membrane-bound molecules that recognize PAMPs [396]. Recognition of PAMPs activates TLR signaling pathways that cause the recruitment of immune cells and cytokines production [397]. There are currently 10 identified TLRs in humans [398]. The main TLRs involved in pneumococcal infections are TLR2, TLR4 and TLR9. TLR2 is necessary in pneumococcal infection because it recognizes bacterial cell wall constituents. Former findings suggested that TLR2 recognized lipoteichoic acids [164, 399, 400]. However, TLR2 is now found to be binding to pneumococcal lipoproteins and peptidoglycan[375, 401, 402]. TLR2 also has a role in transmission of pneumococci. Mouse models with deficient TLR2 had increased inflammation and shedding [403]. TLR2 forms dimers with TLR1 and TLR 6 which assist in the recognition of microbial antigens [396]. TLR4 was the first TLR to be characterized and is needed for recognition of pneumococcal pneumolysin [200, 399, 404]. On the other hand, TLR9 is intracellular and senses

bacterial DNA within endosomes. TLR9 binds to CpG motifs [396] on the DNA, and when activated it also has signaling pathways which result in the release of cytokines [405, 406]. TLR1,2,4,6 and 9 work in a myeloid differentiation primary response 88 (MYD-88)-dependent manner. MYD88 is an intracellular protein necessary for signal transduction and activation of TLR signaling pathways [396, 407]. In addition to cytokine production, the activation of these TLRs facilitates the secretion of co-stimulatory molecules [396] which are necessary for activating T cells [408]. Thus, the functions of these TLRs also play a role in adaptive immunity (Figure 3.9) [396, 405, 406]. Aging greatly affects TLR function. Expression of TLR1 is reduced with age [177]. TLR4 expression appears to remain normal but experiences a reduction of function [177, 264]. This association has been made in mice, due to macrophages having a lowered production of pro-interleukin-1B [177, 264, 409, 410]. This also indicates TLR4's inability to respond to pneumococcal cell wall components [411, 412]. Overall TLR cell signaling impairment causes a reduction in cytokines produced, leading to poor defense against *S. pneumoniae* [177].

- b) **Nod-Like Receptors (NLRs):** NLRs are intracellular proteins that can stimulate nuclear factor-kappa B (NF- κ B) [413], control inflammation, and activate inflammasome formation [205, 414]. NOD2's role in pneumococcal infections has been thoroughly investigated [413–415]. This NLR recognizes muramyl-dipeptide which is a fragment of bacterial peptidoglycan in the cytosol [164, 171]. It promotes the production of cytokines and activation of nucleotide-binding domain and leucine-rich-repeat-containing protein 3 (NLRP3) genes [205]. For example, when NOD2 senses peptidoglycan, CCL2 is made and that recruits macrophages and monocytes to the infection [416]. This is dependent on the lysozyme producing these peptidoglycan fragments [416]. NLRs expression decreases with age and responses to *S. pneumoniae*'s PAMPs are weakened [411]. Lack of NLR expression may contribute to the chronic low pro-inflammatory state observed in the elderly (discussed in Section 3.6.3).

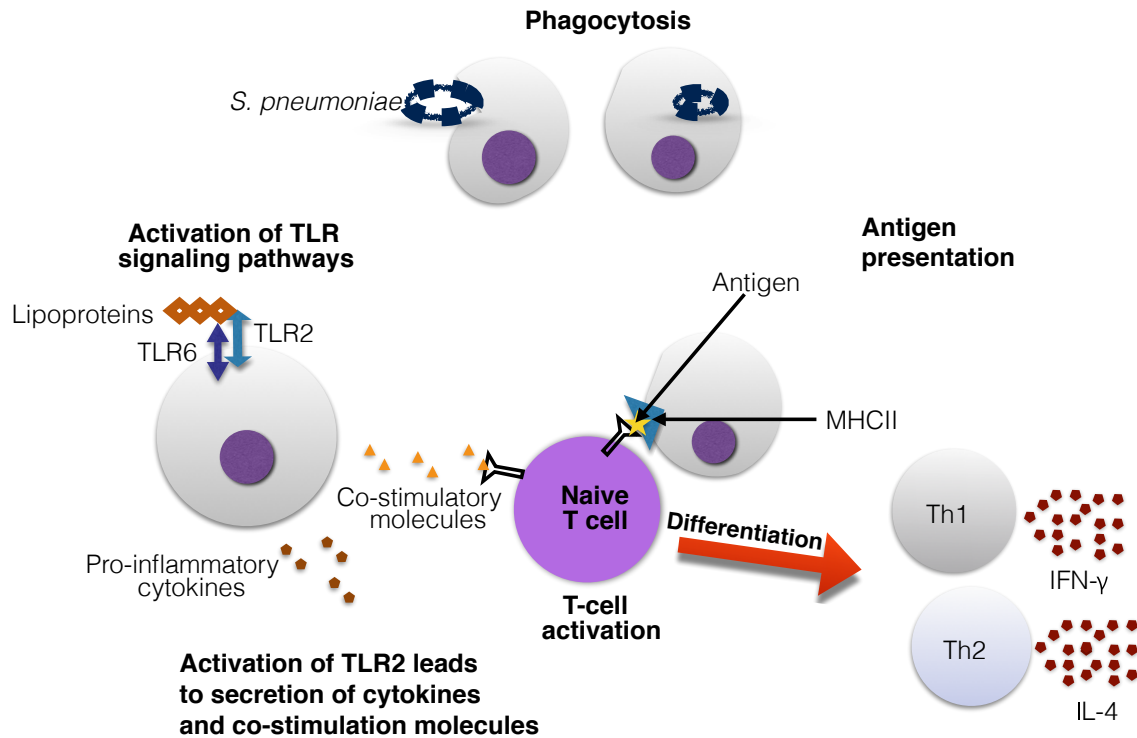


Figure 3.9: Toll-like receptors (TLRs) assist in the activation of adaptive immune cells. In this figure, TLR2 recognizes the *Streptococcus pneumoniae*'s lipoproteins. Upon activation, TLR2 secretes cytokines and co-stimulatory molecules. These co-stimulatory molecules are essential for co-stimulation and activation of T cells. The T cell is presented an antigen with major histocompatibility complex (MHC)II and antigen-presenting cell. The recognition of the antigen–MHCII complex and the co-stimulatory molecules activates the T cell and leads downstream to differentiation into Th1 and Th2 cells, that can release various cytokines such as interferon- gamma (IFN)- γ and interleukin (IL)-4

- c) **CD14:** This has been characterized as a PRR as it recognizes lipoteichoic acid and other cell wall components [397, 417]. CD14 works by interacting with other PRRs such as TLR4 for signal transduction [417]. It has been reported that, in the case of pneumococcal infections, CD14 promotes growth and dissemination of the bacteria [171, 417]. Previous studies have found CD14 to be beneficial and protective to hosts against Gram-negative infections, but as for Gram-positive pathogens such as pneumococci it instead enhances the pathogenesis of the bacteria and facilitates infection [417].

3.6.2 Adaptive Immune Responses (B and T Cells)

Adaptive immune responses transpire a few to several days post infection. The cells involved in adaptive immune responses respond to specific antigens from pathogens. Adaptive immunity can also be broken down into two types of responses: humoral and cell-mediated [369, 418]. Humoral immunity involves B cells that are activated by antigens, and production of antibodies that are specific to antigens. Cell-mediated immunity also involves T cells, including T cell activation and T cell mediated recruitment, which involves the activation of other immune cells that can directly kill pathogenic cells [369].

These immune cells are formed in the bone marrow – B cells mature in the bone marrow into plasma cells that make antigen-specific antibodies [369, 418]. Infections at mucosal sites are controlled by the pneumococcal specific immunoglobulin A (IgA) antibody. IgA is observed in mucosal areas of the nose and saliva following *S. pneumoniae* colonization [419]. Secretory IgA is important for opsonizing *S. pneumoniae* and promoting phagocytosis [380]. *S. pneumoniae* on the other hand, possess an IgA1 protease the cleaves the IgA (discussed in Section 3.5). This blocks opsonization. Following cleaving, the remaining Fab fragment binds to the cell wall [380, 420]. This exposes CBPs, decreases negative charge of the capsule and increases cell adhesion [380, 420]. Studies suggest that the Fab neutralizes the negative charge of the capsule and instead promotes cell adhesion [380, 420]. Furthermore, the complement (C3) activates B cells. Following antigen stimulation, the naïve B cells differentiate into IgM⁺ memory B cells. Class switching produces other immunoglobulins needed for clearing the infection [419]. T cells instead migrate to the thymus for maturity into mature helper (CD4⁺) and cytotoxic (CD8⁺) T cells [418]. Antigen presenting cells (APCs) paired with the major histocompatibility complex (MHC) proteins present antigens (specifically, peptides) to T cells to stimulate an immune response [418]. In pneumococcal infection, CD4⁺ T cells are stimulated via co-stimulatory molecules and APCs. Upon activation helper T cells differentiate into Th1 and Th2 cells (Figure 3.9). Th1 helper cells stimulate a cellular-mediated immune response by producing cytokines such as interferon-gamma (IFN- γ), that activate and recruit other immune cells such as macrophages [421]. Th2 helper cells release IL-4 cytokines,

and are geared towards facilitating a humoral immune response [421] by interacting with B cells and aiding in antibody production [369, 418]. Cytotoxic T cells directly kill infected cells [369, 418]. Furthermore, upon activation of T and B cells, they can differentiate into memory B and T cells that can provide quicker clearance in reoccurring infection [369, 418]. Similarly, natural killer T-cells are also important for clearance of pneumococci [164, 171]. More specifically, CD4⁺ T cells have been found to provide protection to *S. pneumoniae* in an antibody-dependent manner [422]. T-helper 17 (Th17) and Regulatory T cells (Treg) are also very important for pneumococcal infections. Th17 cells release the cytokine interleukin-17 (IL-17) which is pro-inflammatory. IL-17 is needed for recruiting and activating macrophages, monocytes and neutrophils to sites of infection and promotes clearance of *S. pneumoniae* [423]. Increased production of IL-17 has been connected to reduced *S. pneumoniae* density in the nasopharynx of mice and children [423]. Tregs are necessary for regulating Th17's production of IL-17. Imbalance between Tregs and Th17 cells can lead to autoimmune disease due to over inflammation [423]. Infants experience poor T cell responses to foreign antigens because their exposure to non-maternal antigens was restricted prior to birth [387]. Infants also display a skewed Th2 response to foreign antigens. To compensate for this, infants have a population of $\gamma\delta$ T cells that generate IFN- γ to provide a Th1 type response [387]. As for B cells, in infants there is a limited response to antigens due to low expression of co-receptors [387]. Infants also experience incomplete class switching for immunoglobulins and lower somatic hypermutations compared to adults [387]. Immunoglobulin protection against *S. pneumoniae*'s capsular polysaccharides is developmentally regulated. At birth, maternal IgG antibodies protect infants until 27 days of age (based on the half-life of IgG) [177, 424]. Once maternal antibodies have been depleted, the infant's ability to protect itself via steady antibody generation experiences a delay until age two [177]. In contrast, IgM has been detected in infants following *S. pneumoniae* infection and carriage [177, 425, 426]. Encountering the pathogen again, also promotes antibody production similar to booster effects in vaccines (discussed below) [427]. With development, the adaptive immune cells mature, develop memory, and the incidence of *S. pneumoniae* infections decrease. In elderly populations, the efficacy of the adaptive immune cells diminishes. Aging leads

to reduced production of antibodies, immunoglobulin class switching and cell maturation, which promotes *S. pneumoniae*'s colonization [425]. Antibodies specific for capsular polysaccharides decrease with age [425]. Additionally, there is an overall reduction in naïve T cells with age due to thymus involution [428]. Previously Th17 levels were described to increase in elderly populations whereas most recently, in 2014, Geest et al., showed lower Th17 concentrations and increased concentrations of memory T-regs [264, 429]. The ratio between CD4⁺ and T-reg cell populations was also reported to increase towards more Tregs [264, 429]. Diminished responses from the adaptive immune cells explain the higher incidence rates of pneumococcal diseases in these high risk age groups.

3.6.3 Additional Immune Response Considerations

1. **Chemokines and Cytokines:** These are signaling molecules released by innate and adaptive immune cells and receptors to direct other immune cells to the infected tissues [392, 418]. Chemokines are examples of cytokines that attract cells to the infected site. In addition to recruiting cells, they promote inflammation [392, 418]. Tumor necrosis factor- α (TNF- α), a well-studied pro-inflammatory cytokine, inhibits growth and dissemination of pneumococci [171]. Together TNF- α and IFN- γ can enhance clearance of pathogens by activating phagocytes. T cells, monocytes and macrophages produce TNF- α [307]. The phagocyte-activating cytokines are suggested to be inhibited by autolysin activity in pneumococci [307]. The elderly experience chronic low-grade age-associated inflammation (Inflammaging) [177]. This involves constant low levels of pro-inflammatory cytokines such as: TNF- α and IL-6. The inflammatory state of the elderly is worsened due to increased NF- κ B activation and the secretion of pro-inflammatory cytokines such as TNF- α from senescent cells [177]. High concentrations of TNF- α have been correlated with higher disease incidences [177]. Inflammaging induces the expression of host proteins which enhances *S. pneumoniae* adhesion, and is often accompanied by other morbidities that increase risk of *S. pneumoniae* infections [177].

2. **Inflammasome:** This is a protein complex that consists of a sensor protein, caspase 1 and an apoptosis-associated Speck-like protein with a caspase recruitment domain (ASC) [430]. The inflammasome is used by the host for indirectly recognizing bacterial or pathogenic molecules and DNA [205]. Upon recognition, the inflammasome regulates cytokine production [205]. NLRP3 plays a role in identifying pneumococcal infection, activating macrophages and has been shown to directly interact with pneumolysin during pneumococcal infection [375]. Pneumolysin can directly activate NLRP3 [431], and when activated, inflammasomes secrete IL-1 β and IL-18 [432]. Although inflammasomes may aid in the recognition of pathogens, the activation of inflammasomes promote inflammation and this can be harmful to the host [205]. Cho et al studied the effects of aging on NLRP3's activation in mice [411], and reported that enhancement in ER stress with age leads to decreased NLRP3 inflammasome activation in *S. pneumoniae* infection. Ensuring that NLRP3 is activated appropriately in the elderly population will promote stronger immune defenses against *S. pneumoniae*.
3. **Complement System:** This is comprised of a set of small proteins that enhance the ability of antibodies and phagocytic cells to clear microbes and damaged cells [418]. These proteins can mark antigens and cells by coating them with opsonins [164, 369, 418]. Complement activation involves three cascade pathways: classical, mannose-lectin and alternative pathways. In the classical pathway the complement proteins bind to an antibody-antigen complex [368], whereas the alternate and mannose-lectin pathways, bind directly to PAMPs and cell surface components [369, 418]. Which pathway plays the main role in response to pneumococcus has been debated. Brown et al. in 2002 stated that the classical pathway is most important in response to pneumococcal infection following investigations of complement pathways deficient mice [433]. On the other hand, in 2012 Ali et al. stated that the mannose-lectin pathway is more important, after following the use of mannose-lectin pathway in deficient mice that could still use the classical and alternate pathways showed susceptibility to pneumococcal infection [434]. The importance of the different complement pathways' role in pneumococcal infections needs to be further investigated. However, irrespective of the specific pathway,

the complement proteins also help to fight infections by pathogens such as pneumococcus by promoting inflammation, attacking pathogens and rupturing their cell walls [164, 171, 367]. For example, mice deficient in complement C3 infected with *S. pneumoniae* were unable to clear the infection and had short survival times in comparison to mice with complement C3 [435]. *S. pneumoniae* can evade the host complement system in many ways – most of which were previously discussed in Section 3.5. Pneumolysin is able to divert the complement system away from *S. pneumoniae* by directly activating the classical complement pathway [435]. PspA inhibits C1q binding and polyhistidine triad (Pht) proteins are suggested to degrade C3. *S. pneumoniae*'s complement evasion has been detailed by Dockrell and Brown [375]. The effects of aging on the complement system are complex. Previous studies suggest that complement levels are low in infants [436, 437]. In 2014, Grumach et al., also showed that in newborns complement activity is low with C1, factor H and C3a levels being lower than adult levels [437]. Studies have also indicated that complement activity is greater in the elderly compared to young adults [264, 391, 438].

4. **Acute phase serum proteins:** These proteins increase in concentration within the blood during an acute inflammatory infection [439]. The three main proteins that have been investigated and associated with pneumococcal infection include: C-reactive protein (CRP), serum amyloid P (SAP) and mannose binding lectin (MBL) [439]. These proteins work to alleviate infections and can recognize and bind to bacterial surfaces [439]. Acute phase proteins are made as a result of cytokine production from innate cells such as macrophages [439]. For example, CRP production by the liver is increased in response to IL-6 [439]. CRP and SAP bind to phosphorylcholine which is part of the *S. pneumoniae*'s cell wall. Once bound to the phosphorylcholine, CRP and SAP activate the complement deposition on the bacteria via the classical pathway [440]. As for MBL there are conflicting reports about its role in pneumococcus infection as discussed above in the description of the complement system. It has been shown to recognize and attach to sugars on the cell surface of *S. pneumoniae* [441], but more verification is needed on MBL's role in pneumococcal infection.

3.7 Diagnosis, Age-dependent Response, Prevention and Disease Prognosis

3.7.1 Diagnosis

Currently, there are several methods utilized in pneumococcal disease diagnostics. Traditionally, diagnosis begins with physicians performing a physical exam. For example, in the case of an ear infection an otoscope [173] is used to confirm infection, whereas for pneumonia physicians monitor breathing for cracking sounds and wheezing [151, 442, 443]. More specifically, for pneumonia, based on the results of the physical exam, physicians can conduct a chest X-ray to examine the lungs and monitor inflammation to confirm the presence of infection [442, 443]. This X-ray is also performed following signs of respiratory distress [443, 444]. Blood oxygen levels are also measured via pulse oximetry in both children and adults to assess the severity of the infection [151, 442, 443]. Pulse oximetry at the primary care level should be the future, and future technological developments might add respiratory rate and work of breathing to the parameters measured by oximetry [445]. Body fluids are also processed to assess whether or not pneumococcus is present, and to confirm its identity [151, 442–444, 446]. These fluids include: blood, urine, cerebrospinal fluid and saliva [151, 442–444, 446]. The blood test allows physicians to examine complete blood cell count. This test confirms whether or not an infection is present by giving an estimate of the percentage of white blood cells that are circulating [151, 442–444, 446]. A large concentration of white blood cells (WBC) is indicative of an infection [151, 442–444, 446]. This is expected in infection response. However, Gardner et al., in 2017 indicated that upon the time of admission, about 25% of subjects with pneumococcal pneumonia and roughly 38% with CAP actually have normal WBC counts [447]. Studies have also shown that poor prognosis has been associated with low WBC [447]. New finding associate low WBC rather than high WBC with poor prognosis [447]. These conflicting results indicate that WBC count alone should not be used to diagnose pneumonia and should be better investigated as key indicator of pneumonia. Bacterial cultures and Gram-staining tests using body fluids are important for determining the strain of bacteria and confirming its identity [151, 442–444, 446]. Currently, physicians are investigating other means of diagnosing

pneumococcal infections due to the poor yield and quality of sample when conducting cultures. This process is also dependent on bacterial growth which can be time consuming. One useful tool that is being developed is the urinary antigen detection test, which is only currently used in adults [448]. This test monitors the levels of the C-polysaccharide antigen of pneumococcus in the urine. It appears to be quicker, can allow for targeted treatment with better results than culture-based methods of diagnosis [151, 442, 443, 448]. In addition to testing for pneumococcus, physicians also test for other bacteria which may be causing the infection, and other viruses such as influenza which can co-infect patients [443]. Once all the tests confirm the presence of an infection, the cause of the infection and the severity of the disease patients are treated accordingly. Currently, thoracic ultrasounds are being investigated as a method for diagnosing CAP [449]. When compared to chest x-rays, thoracic ultrasounds identified 73.5% of the lung consolidations confirmed by chest x-rays, with about 27% false negative results. D'Amato et al., suggest using ultrasound as a monitoring tool [449]. Lung ultrasound has been tested for its diagnostic potential and it was found to be a sensitive tool for confirming CAP in children [450]. 96% children with pneumonia were detected, however given the small sample size further investigation is necessary. Chest computed tomography is not used for children due to radiation [450]. Recently, a computer-aided differential diagnosis system was tested for distinguishing types of pneumonia, using high-resolution computed tomography. This method was compared to radiologists' classification of interstitial and non-specific interstitial pneumonia, and was concluded to be a robust method for diagnostics [451]. Additionally, researchers have proposed combining clinical signs and laboratory markers to assess an individual's risk of contracting pneumonia. For example, high levels of C-reactive protein and procalcitonin accompanied by unilateral hyperventilation and grunting were associated with pneumonia [452]. On the other hand, children with no clinical signs of pneumonia and low C-reactive protein results were at a lower risk for pneumonia. The use of PCR for diagnosis is also being developed. A positive blood pneumococcal PCR can more accurately confirm the diagnosis of pneumonia [452]. PCR has been used to detect pneumolysin in whole blood samples [453]. The sensitivity of PCR tests varied from 68 to 100 percent and had poor specificity [453]. In contrast,

assessment of quantitative real time PCR indicated that it is more successful in achieving greater speed, specificity and sensitivity compared to multiplex PCR [454].

3.7.2 Prevention, Antibiotic Resistance and Age-Dependent Immune Responses

The two main modes of preventing pneumococcal infections are using antibiotics and vaccinations against pneumococcus [173]. Antibiotics are essential in reducing bacterial load [455]. Such treatment can work by killing the bacteria or hindering their growth [455]. The first antibiotic to be created was penicillin which was discovered in 1928 by Alexander Fleming [456], and antibiotics have been used widely since. However, misuse of antibiotics can cause bacteria to become resistant [188, 455, 457]. Resistant bacteria are then able to survive post antibiotic treatment and they can grow, multiply and share antibiotic resistant genes with each other. Pneumococcal strains that were penicillin-resistant were first recorded in the 1970s [171]. Currently, penicillin-resistant strains have spread worldwide with pneumococcus also being resistant to other types of antibiotics: erythromycin, tetracycline and chloramphenicol [455]. *S. pneumoniae* acquire multiple antibiotic resistance genes via transformation and evolution with the increase in antibiotic use [458]. Mutations in penicillin-binding proteins (pbp) affect binding of penicillin which acts by blocking cell wall synthesis [459]. Erythromycin resistance gene *erm*(B) blocks the binding of macrolides (antibiotics targeting protein synthesis) and *mefA* and *mefE* genes produce an efflux pump which regulates entry of the antibiotics [458–462]. Resistant *S. pneumoniae* strains have rapidly spread, and infections are harder to treat. In 2013, the CDC estimated that about 30% of pneumococcal cases were due to antibiotic resistance to one or more antibiotics [457]. This resistance increases the number of doctor visits and hospitalizations [457]. For example, the CDC reports that the resistance can lead to 1,200,000 more illness and 7,000 deaths annually [457]. This reduction in ability to treat and clear the pathogen led to the development of vaccines that would provide protection prior to infection and thus reduce the need for antibiotics [455]. Currently, there are two types of inactivated vaccinations that protect against *S. pneumoniae* [12, 463, 464]. The pneumococcal polysaccharide vaccine 23 (PPSV23) [465] uses purified capsular polysaccharides and is routinely

given to adults who are 65 and older [12, 463, 464]. It protects against 23 serotypes of *S. pneumoniae* and is effective in 50-70% of cases in adults [274]. This vaccine works in a T-cell independent manner. The polysaccharide antigens are recognized by B cells which differentiate into plasma cells that produce antibodies specific for the polysaccharide antigens [466]. PPSV23 provides T-cell independent immunity and requires revaccination five years after the first vaccination because the immunity is transient [465, 467]. The pneumococcal conjugate vaccine (PCV) [468] was developed after noticing the low efficacy and poor immunogenicity of PPSV23 in infants and young children [469, 470]. In the conjugate vaccine, the purified polysaccharides covalently conjugated to a carrier protein, specifically CRM197 [467, 468]. The current FDA approved conjugate vaccine is PCV13 which protects against 13 serotypes of the *S. pneumoniae* [468]. PCV13 replaced PCV7 in 2010 and protects against 6 additional serotypes [471]. This elicits a T-cell dependent response which provides mucosal immunity, and immunologic memory in children [274]. PCV13 provides long lasting immunity by causing B and T cells to interact [466]. B cells recognize and process the carrier protein [466]. The MHC II needed for antigen-presentation to T cells, binds to the peptide produced following B cell breakdown of the carrier protein [466]. The peptide is presented to the T cells by MHC II at the surface of the APC, providing co-stimulation necessary for producing plasma cells and memory B cells [466]. The use of this vaccine has led to a decrease in pneumonia cases in young children by more than 90%, and is most effective in children younger than five [274].

When it comes to high-risk individuals, the CDC recommends the prime boost method of vaccination. This involves priming the immune system to a specific antigen, and enhancing this antigen-specific immune response by re-administering the antigen[472]. The prime boost strategy increases immunity to antigens and is recommended for high risk individuals [473, 474]. There are two ways to prime and boost the immune system: homologous, in which the same vaccine is received twice, and heterologous, which utilizes different types of vaccines [472]. The heterologous method has been shown to be more immunogenic [475]. Currently, children and adults who are at high risk for pneumococcal disease and have pre-existing conditions undergo the prime-boost strategy prevention by receiving the PCV13 followed by the PPSV23 [476, 477]. This is also due

to the poor immunological response seen in HIV patients who receive the PPSV23. Prime-boost vaccinated HIV-infected groups have been shown to be more likely to display a 2-fold increase in IgG geometric mean concentrations [478]. PCV13 provides a longer and stronger level of protection against *S. pneumoniae* [473]. Within 4-8 weeks PCV13's IgG levels can equal or exceed PPSV23 in high risk individuals [473, 474].

Despite the availability of pneumococcal vaccines, it is important to note that these vaccinations are both serotype and age dependent [12, 172, 463, 464, 469, 470]. Understanding the role that age plays in host immune system activation is essential for better prognosis and treatment of diseases. As stated previously, young children and the elderly are at higher risk for contracting pneumococcal diseases [387]. This is due to immunosenescence within the elderly population, whereas for infants, it is due to their underdeveloped immune systems [387]. In addition to age recommendations, the CDC also recommends the use of either vaccine in high-risk individuals with pre-existing health conditions. For example, both vaccines are recommended in young children and adults ages 19-64 with pre-existing health conditions [172, 173, 469]. PPSV23 is also recommended by the CDC for use in adults that smoke or have asthma [173]. These vaccine recommendations are re-evaluated regularly based on vaccine efficacy and changes within the bacteria serotypes [13, 477]

Vaccines have drastically reduced invasive pneumococcal diseases, especially CAP in young children and adults (Table 2) [166]. However, these vaccines have pitfalls. Firstly, there have been at least 97 serotypes identified but these vaccines protect against 14-25% of these. The current vaccines only protect against *S. pneumoniae* serotypes that are mainly associated with causing the disease. Some studies suggest that there is little evidence that PPSV23 protects against non-invasive pneumococcal diseases, which are more prevalent in adults [479]. The CDC also confirms that PPSV23's efficacy in non-bacteremic pneumonia has led to contradicting findings, but nevertheless, it has shown sufficient efficacy in invasive pneumococcal diseases [480]. Weinberger et al., discuss the challenges of vaccinating the elderly with PCV13 and PPSV23 [479]. These researchers argue that PPSV23 does not show a real benefit to the elderly [479]. As for, PCV13 they argue that it is already used in children and thus adults should be partially protected from serotypes in PCV13 due

to herd immunity [479]. They also state that herd immunity should provide partial protection and thus will lead to reduction of efficacy of PCV13 [479]. Other studies also discuss herd immunity from PCV13 due to infants and toddlers being vaccinated [481, 482]. Due to PCV13, disease serotypes rates within this vaccine will decrease by 50% [481, 482]. This becomes a problem because of serotype replacement. The serotypes that are not in the vaccine, can colonize young children and spread to adults [445, 483]. Additionally, with serotype vaccines, the serotypes that are popular and commonly cause CAP and other diseases may not necessarily do so in the future and so these vaccines would need to be reevaluated. PCV13's replacement of PCV7 was a prime example of changes to serotypes that cause pneumococcal diseases. Recently Merck and Dohme completed a phase 1 clinical trial (NCT01215175) investigating a new conjugate vaccine, PCV15, immunogenic and safety properties compared to PCV13 [484]. This contains two extra serotypes (22F and 33F), which were previously identified for the cause of approximately 10% of invasive pneumococcal diseases in adults in 2007 [485]. Another concern for current vaccines is that 3-19 percent of pneumococcal diseases are due to non-encapsulated *S. pneumoniae* [187]. Current vaccines are ineffective against non-encapsulated *S. pneumoniae* due to serotype specificity [187]. Further developments of vaccinations are vital for eliminating the burden of *S. pneumoniae* and reduce the number of infections.

3.7.3 Post-infection Prognosis

Following pneumococcal diseases such as pneumonia, high risk individuals may experience longer recovery times and complications due to the disease [151, 442, 443]. About 1.6 million deaths from pneumococcal diseases occur worldwide [185]. According to the CDC, there were over 50,000 deaths within the US during 2014 and the majority of these deaths were seen in the elderly [178]. Older adults have lower survival rates than other age groups [242, 486]. The elderly may recover from pneumococcal diseases such as CAP, but they face higher death rates due to the high possibility of developing other health problems and the reoccurrence of the disease [151, 242, 486]. Infants and young children that recover from CAP have an increased risk for developing respiratory

problems [487]. For example, research indicates that young children face a greater risk for reduced lung function and developing Chronic Obstructive Pulmonary Disease (COPD) [193, 487]. In some cases, increased death rates and complications are due to delays in diagnosis. Such delays in turn hamper timely treatment, which also increases the severity of the disease. For example, meningitis can progress quickly and cause permanent disabilities such as brain damage, hearing loss and seizures [173, 201, 202]. Timely treatment can reduce the risk of neurological damage and death due to this infection [202]. Additionally, ear and sinus infections can lead to hearing loss and respiratory problems respectively [173]. The environment also plays a role in affecting recovery rates and reoccurrence of the disease especially for smokers and those residing in nursing homes and crowded areas [151]. Furthermore, Tiewsoh et al. study investigating the outcome of children with severe pneumonia showed that children who were not breastfed, had a low birth weight and were within crowded homes had longer hospital stays and the initial antibiotics were not helpful and required new antibiotics [488]. Nutrition also plays a vital role in how well someone will recover from these diseases [489]. Some complications due to pneumococcal pneumonia include respiratory failure, lowered oxygen levels and collapsed lungs [154]. It is also possible for the lungs to fill with fluid and this fluid can become infected. *S. pneumoniae* may also migrate to the blood [154, 173]. This is called bacteremia which is the most common complication for pneumonia [154, 173]. Pneumonia and other pneumococcal diseases are classified as invasive if the bacteria migrate to the blood. Additionally, individuals with this disease can develop pericarditis which is inflammation of the sac around the heart, lung abscess, empyema and blockage of airways [154, 173]. It is also highly probable for co-infections to occur when suffering with pneumonia. An example of this is influenza – 66% of CAP cases also present co-infection with influenza [260]. Most of these health complications are seen in elderly subjects, and this also points to the increasing importance of improved diagnostics, treatments and vaccinations for this age group.

3.8 Discussion

We have discussed the host defenses against *S. pneumoniae*, and how individuals with weakened immune systems may experience a harder time clearing the pathogen. We have also indicated that young children, elders and individuals who are immunocompromised all have an increased risk for contracting pneumococcal diseases. The majority of previous efforts have provided an extensive characterization of *S. pneumoniae* features and began probing the interactions of the bacteria with the host in the context of pneumococcal disorders. However, in terms of treatment and prevention there remain substantial open questions that need to be addressed as discussed below. There are a variety of methods available for pneumococcal disease diagnostics. Many of the current tests needed to confirm *S. pneumoniae*'s identity are culture-dependent [151, 442, 443]. Culture-independent methods that take advantage of the latest technologies are being developed, such as the use of a lung ultrasound to assess pneumonia [490]. Chavez et al. and Long et al. discuss the possibility of lung ultrasound use in pneumonia diagnosis [491, 492] indicating high diagnostic accuracy, while at the same time providing a radiation free method of examining the lungs [491, 492]. Similarly, the use of mass spectrometry to examine metabolites from the saliva [493], breath [494] and urine [448, 495] of patients being tested for pneumococcal diseases is under development. The urine antigen test discussed above also provides rapid results that will allow for quicker diagnosis and treatment once *S. pneumoniae* antigens are detected in the urine [448, 495]. With diagnostic methods improving, pneumococcal disease treatments are also being updated. Antibiotics are available to reduce the colonization of *S. pneumoniae*, however, the efficacy of antibiotics is being reduced due to the increase in antibiotic resistance [455, 457]. Broad-spectrum antibiotics are no longer as effective [455, 457]. Inhaled therapeutics are underdeveloped but can be beneficial for treating pneumonia and other pneumococcal diseases. This method can provide a mode of delivering antibiotics and antimicrobials [496] in a more targeted manner, improve mucociliary clearance via hypertonic saline solutions and inhalation of cytokines to stimulate the immune system [496]. On the other hand, to also reduce the effect of antibiotic resistance, *S. pneumoniae* strains may also be studied via RNA-sequencing and other high throughput technologies to detect antibiotic resistance

genes and thoroughly characterize serotypes. Treatment and prevention of pneumonia and other pneumococcal diseases are of major concern for the clinical field due to the high death rates and low efficacy of current vaccines due to aging differences and serotype replacement. Some alternative vaccination methods have been proposed and are also being developed. For instance, Weinberger et al., propose the use of a conjugate vaccine that is specific for elderly subjects, which targets the serotypes not in current vaccines but other serotypes that are mostly seen in elderly patients with pneumococcal diseases [479]. Some researchers have proposed creating a conjugate vaccine that targets all or more of the identified serotypes of *S. pneumoniae* [479, 497]. However, the impact on the immune system and immunogenicity of this vaccine would need to be thoroughly investigated [479]. This vaccine would also need to demonstrate better efficacy than existing vaccines [479]. In addition to this, conjugate vaccines are expensive (currently, the PCV13 costs about \$160 per dose [498], and a true benefit will need to be clearly identified. Additionally, observing how pneumococcal disease incidence rates are changing as more and more people are getting vaccinated will lead to accurate assessment of pneumococcal disease burden and vaccine efficacy [489]. Vaccination policies and cost-effect analyses can benefit from information on vaccine disease reduction [489]. Serotype-independent vaccines are also being investigated. These include protein, protein and polysaccharide combination, and whole cell vaccines [479, 497, 499, 500]. Protein vaccines would contain surface proteins that are highly conserved in *S. pneumoniae* [501, 502]. For example, PspA and inactivated pneumolysin have been tested in phase 1 clinical trials as protein antigens [502]. They both demonstrated safety [502], but PspA antigen's immunogenicity was low [503] whereas the inactivated pneumolysin was found to be immunogenic and effective in eliciting protective immune response [504]. PspA is considered an ideal protein candidate because reports indicate that PspA family 2 is commonly found in *S. pneumoniae* strains [505]. For example, in Pakistan most strains of pneumococci have PspA genes [505]. These protein vaccines can provide an extra preventative method once developed and will require thorough analysis of regulation and what regulatory issues may be faced [506]. Additionally, as a form of combination therapy, a vaccine with protein antigens as well as conjugated polysaccharide antigens may also provide a

broader range of protection against pneumococcal diseases [501, 502]. On the other hand, whole cell vaccinations would introduce a dead *S. pneumoniae* cell to hosts with the potential to provide broader protection to *S. pneumoniae* [507, 508]. HogenEsch et al., investigated the use of whole cell vaccines in mice by using a capsule deficient and autolysin mutant cell [507]. This exposed the host to multiple parts of *S. pneumoniae*. They found that the vaccine led to the productions of antibodies and IL-17 which defend against *S. pneumoniae* colonization of the nasopharynx in mice [507]. Researchers have also started developing live attenuated pneumococcal vaccines [509, 510]. The SPY-1 strain is a live attenuated strain of pneumococci that does not have a capsule [510]. Xiuyu et al in 2015 experimented with delivering this vaccine intranasally in mice and observed that it elicited a humoral response [510]. More recently, Xinyuan et al., added a mineralized shell to SPY-1 to improve its stability and test if it can elicit a stronger immune response [509]. The modified strain (SPY1 δ lytA) also did not have autolysin activity [509]. This modified SPY1 vaccine led to higher stability, more production of IgG, and an overall increase in protection when compared to the SPY-1 vaccine [509]. Additional concerns of serotype-independent vaccines include determining if the vaccines will be immunogenic in all ages, whether or not the vaccines would elicit a strong immune response, and ensuring that they can induce a pro-inflammatory state while not leading to an over activation of the immune system. All of these novel methods show great promise, but they require further assessments. Overall, there has been progress in our understanding of pneumococcal diseases over the last three decades, however, the diseases still constitute a big burden on health care. There has been a great decrease in pneumococcal diseases since the implementation of purified polysaccharide and polysaccharide conjugate vaccines, but over time due to serotype replacement, antibiotic resistance, and changes in immunity with age, the treatments and vaccines in place may prove ineffective. Therefore, ongoing research to improve vaccinations and treatments must continue towards alleviating the ill effects of *S. pneumoniae*.

CHAPTER 4

META-ANALYSIS OF GENE EXPRESSION MICROARRAY DATASETS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE.

Work presented in this chapter has been submitted to the PLOS One journal. A pre-print is available on bioRxiv: Rogers LRK, Verlinde M, Mias GI. Meta-analysis of Gene Expression Microarray Datasets in Chronic Obstructive Pulmonary Disease. bioRxiv 671206; doi: <https://doi.org/10.1101/671206>

4.1 Abstract

Chronic obstructive pulmonary disease (COPD) was classified by the Centers for Disease Control and Prevention in 2014 as the 3rd leading cause of death in the United States (US). The main cause of COPD is exposure to tobacco smoke and air pollutants. Problems associated with COPD include under-diagnosis of the disease and an increase in the number of smokers worldwide. The goal of our study is to identify disease variability in the gene expression profiles of COPD subjects compared to controls. We used pre-existing, publicly available microarray expression datasets to conduct a meta-analysis. Our inclusion criteria for microarray datasets selected for smoking status, age and sex of blood donors reported. Our datasets used Affymetrix, Agilent microarray platforms (7 datasets, 1,262 samples). We re-analyzed the curated raw microarray expression data using R packages, and used Box-Cox power transformations to normalize datasets. To identify significant differentially expressed genes we ran an analysis of variance with a linear model with disease state, age, sex, smoking status and study as effects that also included binary interactions. We found 1,513 statistically significant (Benjamini-Hochberg-adjusted p-value <0.05) differentially expressed genes with respect to disease state (COPD or control). We further filtered these genes for biological effect using results from a Tukey test post-hoc analysis (Benjamini-Hochberg-adjusted p-value <0.05 and 10% two-tailed quantiles of mean differences between COPD and control), to identify 304 genes. Through analysis of disease, sex, age, and also smoking status and disease interactions we identified differentially expressed genes involved in a variety of immune responses and cell processes in COPD. We also trained a logistic regression model using the 304 genes as features, which enabled prediction of disease status with 84% accuracy. Our results give potential for improving the diagnosis of COPD through blood and highlight novel gene expression disease signatures.

4.2 Introduction

Chronic obstructive pulmonary disease (COPD) impairs lung function and reduces lung capacity. In COPD there is inflammation of the bronchial tubes (chronic bronchitis) [511] and

destruction of the air sacs (emphysema) [512] within the lungs [14, 513–515]. Chronic bronchitis and emphysema often occur together and are grouped under COPD [511, 512]. Furthermore, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) describes COPD as a common and preventable disease that is caused by exposure to harmful particles and gases that affect the airways and alveolar of the lungs [516, 517]. Individuals with COPD experience shortness of breath due to lowered concentrations of oxygen in the blood and a chronic cough accompanied by mucus production [14, 511–514]. COPD progresses with time and the damage caused to the lungs is irreversible [517, 518]. However, there are treatments available to control disease progression [517, 518].

COPD, the 3rd leading cause of death in the United States (US), is expected to rise in 15 years to the leading cause of death [517–519]. Globally, there were over 250 million cases of COPD reported in 2016 and in 2015 3.17 million individuals died from the disease [515]. COPD is prevalent in low- and middle-income countries with over 90% of COPD cases occurring in these areas [515, 519]. The disease is mainly caused by tobacco exposure through smoking cigarettes or second-hand exposure to smoke [517, 518]. In addition to this, continuous exposure to other irritants such as burning fuels, chemicals, polluted air and dust can lead to COPD [515]. Cigarette smoke exposes the lungs to large amounts of oxidants that induce inflammation of the airways. Previous research on bronchial biopsies highlighted the presence of increased concentrations of inflammatory cells throughout the lungs [520, 521]. Studies have also suggested that COPD acts like an autoimmune disease due to persistent inflammation even after smoking has ceased [521–523]. In addition to environmental pollutants, there is also also a genetic deficiency, alpha-1 antitrypsin deficiency (AATD), that is associated with COPD [517]. AATD protects the lungs, and without it the lungs become vulnerable to COPD. The prevalence of COPD is expected to rise due to increasing smoking rates and larger populations of elderly individuals in many countries[515].

COPD is often underdiagnosed and despite tobacco exposure being the highest risk factor, not all smokers get COPD, and non-smokers can also develop COPD. Previous work has been done to identify biomarkers for earlier diagnosis of COPD in blood, a non-invasive approach. Bahr et

al., compared expression profiles of smokers with COPD and smokers without COPD [524]. They used multiple linear regression to identify candidate genes and pathways. Their results highlighted pathways involved in the immune system and inflammatory response [524]. Another study of blood gene expression in COPD explored using pre-existing gene interaction networks to perform unsupervised clustering to identify COPD disease sub-types [525]. More recently, Reinhold et al., took a different approach by conducting a meta-analysis that identified groups of genes associated with COPD by using consensus modules of gene co-expression. They built networks of genes that were co-expressed and associated with COPD phenotypes [5].

In our meta-analysis, the objective was to identify the effects of age, sex, and smoking status on gene expression in COPD. We investigated gene expression changes in blood for 1,262 samples (574 healthy samples and 688 COPD samples) to identify genes and their associated pathways in COPD (Figure 4.1-4.2). Our study is the largest meta-analysis on blood expression for COPD to date, to the best of our knowledge, and our results offer prospective gene and pathway associations that may be targeted for improving COPD diagnosis and treatment. Our meta-analysis also highlighted disease genes that interact with smoking status, and these genes can be used to further characterize the effects of smoking on COPD development.

4.3 Materials and Methods

We used seven publicly available COPD microarray gene expression datasets in our meta-analysis to evaluate variation in gene expression across samples due to disease status, sex, age and smoking status (Table 4.1). The 7 expression datasets were from 3 different microarray platforms: Affymetrix GeneChip Human Genome U133 Plus 2.0, Affymetrix Human Gene 1.1 ST Array and Agilent Whole Human Genome Microarray 4x44K. Our current meta-analysis pipeline (similar to Brooks et al.[526]), included 5 main steps (Figure 4.2): (1) data curation; (2) pre-processing of raw expression data; (3) analysis of variance (ANOVA) on our linear model which compared gene expression changes due to disease state, smoking status, sex and age group; (4) post-hoc analysis using Tukey Honest Significance Difference test (TukeyHSD) for biological significance;

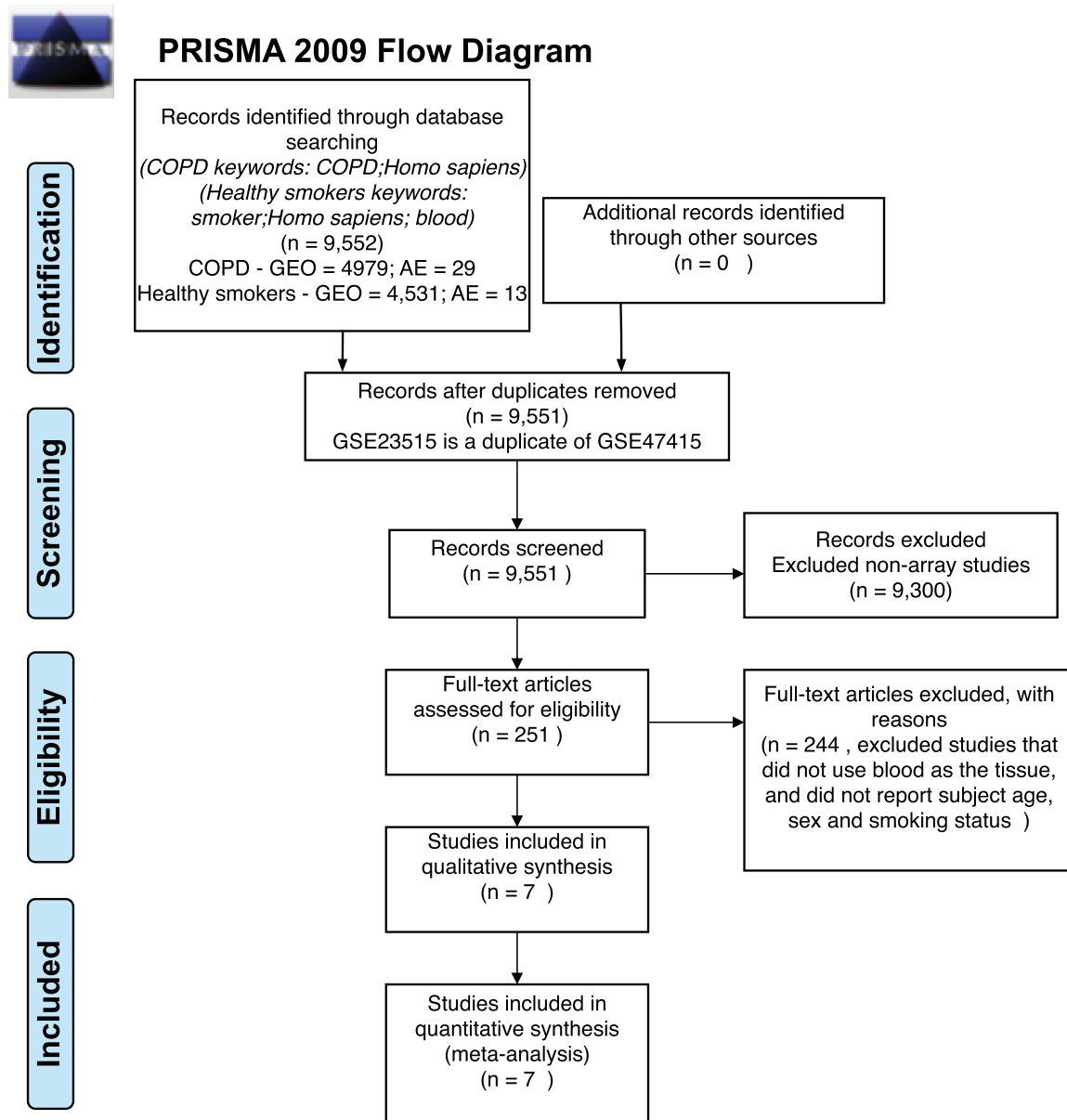


Figure 4.1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. Data were curated from Gene Expression Omnibus (GEO) and Array Express (AE). The PRISMA flow diagram shows the identification, screening, eligibility and inclusion of samples in our analysis.

and (5) Gene ontology (GO) and pathway enrichment analysis of the differentially expressed and biologically significant genes.

Database Repository	Dataset Accession	Control	COPD	Platform
Array Express	E-MTAB-5278	181	53	Affymetrix Human Genome Plus 2.0
Array Express	E-MTAB-5279	89	0	Affymetrix Human Genome Plus 2.0
GEO	GSE42057	42	94	Affymetrix Human Genome Plus 2.0
GEO	GSE47415	48	0	Agilent-014850 Whole Human Genome Microarray 4x44K
GEO	GSE54837	90	136	Affymetrix Human Genome Plus 2.0
GEO	GSE71220	44	405	Affymetrix Human Gene 1.1 ST Array
GEO	GSE87072	80	0	Affymetrix Human Genome Plus 2.0

Table 4.1: Description of datasets used in the meta-analysis

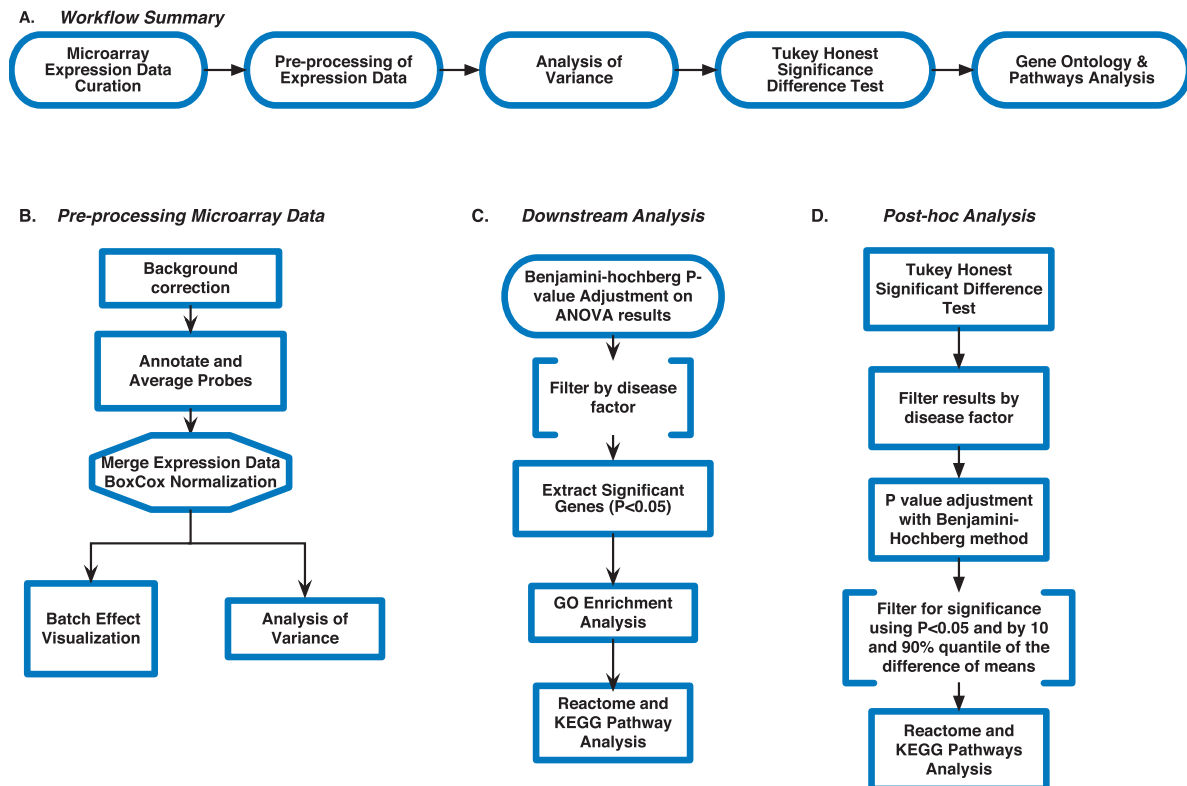


Figure 4.2: Meta-analysis pipeline for Chronic Obstructive Pulmonary Disease. (A) Summary of workflow used for the meta-analysis, (B) Pre-processing steps used on the microarray data, (C) Data analysis post ANOVA, (D) post-hoc analysis steps using ANOVA results.

4.3.1 Microarray Data Curation from Gene Expression Omnibus and Array Express

To gather the datasets for our meta-analysis, we searched the National Center for Biotechnology Information (NCBI)'s data repository, Gene Expression Omnibus (GEO) [47], and the European Bioinformatics Institute (EMBL-EBI)'s data repository, Array Express (AE) [48] for microarray expression data. We used the following keywords to search the repositories: COPD, *Homo*

sapiens, blood (whole blood and peripheral blood mononuclear cells) and expression profiling by array (Figure 4.1). The search results were further filtered to include datasets where the age, sex and smoking status of the samples were reported (Figure 4.1). We found 3 datasets from GEO (GSE42057 [527], GSE71220 [528], GSE54837 [529]) and 1 from AE (E-MTAB-5278 [530]) that met our search criteria (Table 4.1 and Figure 4.1). We conducted an additional search on GEO and AE to find healthy subjects with their smoking history reported to balance our control subjects with our COPD subjects. The search keywords included: *Homo sapiens*, blood, smoking and expression profiling by array. We also filtered these search results for datasets that reported the age, sex and smoking status of subjects. With this additional search, we added 3 more datasets: GSE87072 [531], GSE47415 [532], and E-MTAB-5279 [530] which helped improve the balance between COPD and control subjects (Table 4.1 and DF1 of online supplementary data files (Appendix B)).

After selecting the datasets for our meta-analysis, we retrieved the raw microarray expression data for each dataset, and created a demographics file per study, which included sample characteristics using e-utils in Mathematica [49] (Table 4.2). The demographics files were further filtered to eliminate samples that did not fit our inclusion criteria. For example, GSE71220 included subjects that were using statin drugs [528], and hence we excluded all samples that were receiving treatment from our analysis. For GSE87072, we removed the samples that were moist snuff consumers [531] and only used smokers and non-smokers in our analysis. In our additional search for controls with smoking status reported, we filtered the selected datasets (GSE87072, GSE47415 and E-MTAB-5279) and only used the healthy samples for our analysis. In addition to this, we excluded the subjects in GSE23515 [533] from our analysis because 22 of the 24 samples are duplicates from GSE47415 [532]. Our demographics files were created to include variables that were reported across all samples (see merged Demographics file DF1 of online supplementary data files (Appendix B)) because study annotations had not been uniformly reported in the databases (Appendix B).

Dataset Accession	Sex(M/F)	Smoking Status (S/NS/FS)*	Age Range
E-MTAB-5279	46/43	30/29/30	24 - 65
EMTAB5278	136/98	114/60/60	41 - 70
GSE42057	74/62	35/2/99	45 - 80
GSE47415	24/24	24/24/0	20 - 64
GSE54837	148/78	84/6/136	40 - 75
GSE71220	285/165	91/22/336	49 - 75
GSE87072	80/0	40/40/0	35 - 60

**S=smoker, NS=non-smoker, FS= former smoker*

Table 4.2: Sample Characteristics By Dataset

4.3.2 Microarray Pre-processing and BoxCox Normalization

To download the raw microarray expression for each dataset we used Mathematica [50]. All raw expression data files were pre-processed in R [51] using R packages specific to each microarray platform (Figure 4.2B). For the datasets from the Affymetrix Human Genome Plus 2.0 platform, we used the *affy* package [52] for pre-processing all of the .CEL files. The *oligo* [534] and *affycoretools* [535] packages were used to pre-process the data files from the Affymetrix Human Gene 1.1 ST microarray platform, while the *limma* package [53] was used for the data files from the Agilent Whole Human Genome microarray platform. We performed background correction, normalization, and all probes were annotated and summarized (Figure 4.2B). For the Affymetrix Human Genome Plus 2.0 expression data files, the *expresso* function was used to pre-process the files with the following parameters: background correction with robust multi-array analysis (RMA), correcting the perfect-match (PM) probes, and ‘avdiff’ to calculate expression values [52]. Subsequently, the *avereps* function from *limma* was used to summarize the probes and remove replicates [53]. The Affymetrix Human Gene 1.1 ST data files were also background corrected using RMA, and the probes were summarized and replicates removed using the *avereps* function. As for the Agilent data files, background correction was performed using the *backgroundCorrect* function with NormExp Background Correction as the method from the *limma* package [536]. The probes for both Affymetrix Human Gene 1.1 ST and Agilent were also summarized and replicates were removed using the *avereps* function from *limma*. Once pre-processing was

completed, the 8 datasets (Table 4.1) were merged by common gene symbols into a single matrix file. Using the `ApplyBoxCoxTransform` function and the `StandardizeExtended` function from the `MathIOmica` (version 1.1.3) package [49, 55] in Mathematica, we performed a Box-Cox power transformation and data standardization on the merged expression file [54] (Figure 4.2B and DF2 of online supplementary data files (Appendix B)).

4.3.3 Identifying and Visualizing Batch Effects

Conducting meta-analyses by combining expression datasets across different microarray platforms and research labs/studies introduces batch effects/confounding factors to the data. The batch effects can introduce non-biological variation in the data, which affects the interpretation of the results. In order to visualize variation in the expression data across factors, we conducted principal component analysis (PCA) on the expression data and generated PCA plots (Figure 4.3 and Figure B.1 of Appendix B). As we also previously described[526], the study factor is directly related to the microarray platform type. To address this, the `ComBat` function in the `sva` package was used to correct for variation in the data due to the study factor [56, 57]. PCA plots were used to visualize variation in expression data before and after batch correction with `ComBat` [58] (Figure 4.3 and Figure B.1 of Appendix B), confirming the main batch effect removal by adjusting for study.

4.3.4 Analysis of Variance to Identify Differentially Expressed Genes by Factor

To determine if the factors of disease status, sex, study, and smoking status had an impact on gene expression in COPD, we modeled (see linear model below) our merged expression matrix (DF2 of online supplementary data files (Appendix B)) and then ran ANOVA to identify differentially expressed genes (Figure 4.2B) using `aov` and `anova` from base R's `stats` package (as previously described[526]). Schematically our linear model formula for gene expression, g , per each gene included main effects and interactions:

$$g \sim \sum_i x_i + \sum_{i,j;j>i} x_i : x_j \quad (4.1)$$

where $x_i \in \{\text{age group, sex, smoker, disease status}\}$ and the factors have the following levels:

- disease status = {control, COPD}
- sex = {male, female}
- age group = {under 50, 50-55, 55-60, 60-70, over 70}
- smoker = {non-smoker, former smoker, smoker}
- study = {GSE42057, GSE47415, GSE54837, GSE71220, GSE87072, E-MTAB-5278, E-MTAB-5279}

ANOVA p-values were adjusted using the Benjamini-Hochberg (BH) correction method for multiple hypothesis testing [59, 60, 537]. Genes were considered statistically significant if their BH-adjusted p-values were <0.05 . We focused on the ANOVA results for the disease factor, and filtered them for BH-adjusted p-values <0.05 . These filtered genes were then identified as statistically significant disease genes. We used this gene list to identify what GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways they were enriched in. We used the `GOAnalysis` and `KEGGAnalysis` functions from the `MathIOmica` package for GO and KEGG pathway enrichment. Additionally, we used the `enrichPathway` function from the `ReactomePA` package in R [36]. All functions for enrichment analysis used the BH p-value correction method and GO terms, KEGG and Reactome pathways with a BH-adjusted p-value <0.05 were considered statistically significant (see DF5-DF7 of online supplementary data files (Appendix B)).

To determine the biological effect of the ANOVA statistically significant genes (disease status factor) and calculate relative expression (difference in means) to determine up- or down- regulation of genes, we conducted a post-hoc analysis with `TukeyHSD` function in the `stats` package in base R using our linear model outlined above. We added an additional column to the `TukeyHSD` results which contained BH-adjusted `TukeyHSD` p-values, and all GO terms and pathways with a BH-adjusted p-value <0.05 were considered significant. To find genes that were significantly up-

and down-regulated, we further filtered the gene list by difference in means by using the two-tailed 10 and 90% quantile. With these results we carried out GO and pathway enrichment to identify which biological processes and pathways the genes were enriched. We used the disease genes and explored sex, smoking status and aging effects on their gene expression.

4.3.5 Machine Learning with COPD

Machine learning classification was carried out in Mathematica using the `Classify` function [538], with the `Method` parameter set to “LogisticRegression”. We first trained on all 1262 samples, using the statistically significant disease genes, filtered with a two tailed 10 and 90% quantile selection for effect size as features (304 genes). We also randomized the dataset, and created 10 sets for training and testing, with 90% of the samples used for training, and 10% of the samples used for testing, where the 10 testing sets were mutually exclusive (10-fold cross-validation).

4.4 Results

Our meta-analysis selection criteria for data curation (Figure 4.1) resulted in 8 datasets from GEO and AE (Table 4.1). After pre-processing the data, we combined all datasets into a large matrix by merging by common gene names. This data merge resulted in 1,262 samples (574 controls and 688 COPD subjects) and 16,237 genes. Our 1,262 samples consists of 792 males and 470 females, and also 661 former smokers, 418 current smokers and 183 non-smokers.

4.4.1 Visualizing Batch Effects and Batch Effect Correction

Prior to designing our linear model, we wanted to visualize variation introduced into the data due to batch effects, and how the variation changes when the data is adjusted with ComBat for batch effects. We used ComBat in R to adjust for the study effect on the data and generated PCA plots before and after batch correction (Figure 4.3). In Figure 4.3A, before running ComBat, the data separates into four major clusters with a variance of 49.9% in PC1 and 15.7% in PC2. After running ComBat, the clustering of the data is removed, and variance reduced to 17.7% in PC1 and 4.4%

in PC2 (Figure 4.3B). We also plotted the PCAs for the other factors (Figure B.1 of Appendix B) before and after using ComBat for batch effect correction. The ComBat batch effect corrected expression data was only used to visualize changes in variation due to removal of batch and to confirm the inclusion of study as an effect factor in our linear model.

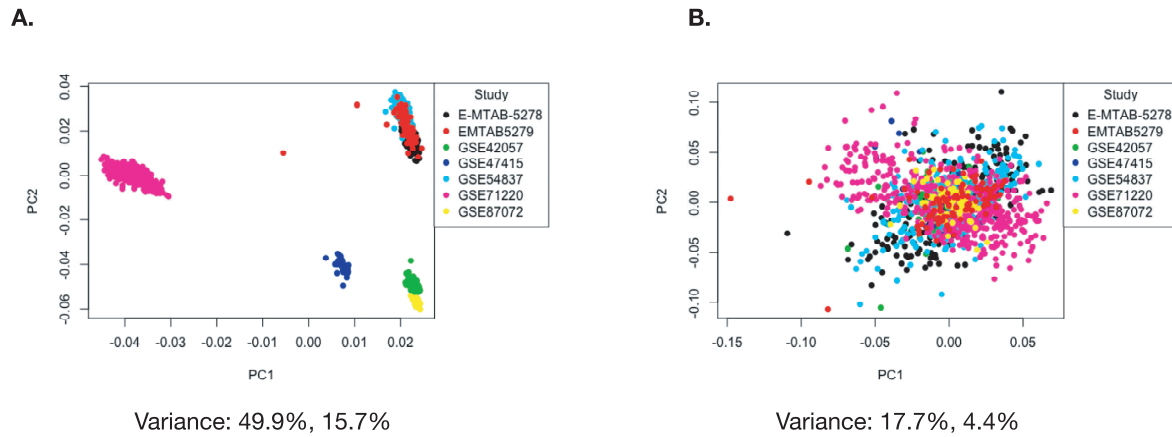


Figure 4.3: Visualizing batch effects introduced by using multiple studies in our meta-analysis. (A) PCA before and (B) PCA after batch effect correction with ComBat.

4.4.2 Variance in Gene Expression Due to Disease Status

With our ANOVA results, we were able to evaluate variance in gene expression introduced by each factor and their pair-wise interactions [59]. To determine which genes from our ANOVA results were statistically significant by the disease status factor, we filtered the genes by using BH-adjusted p-value <0.05 . We found 1,513 statistically significant disease genes (see DF4 of online supplementary data files (Appendix B)). We performed GO and pathway enrichment analysis on the 1,513 genes. Our enriched GO terms included: innate immune response (57 gene hits), inflammatory response (48 gene hits), apoptotic process (58 gene hits), adaptive immune response (24 gene hits) and response to drug (40 gene hits) (see DF7 of online supplementary data files (Appendix B) for full table). We found 7 enriched KEGG pathways (Table 4.3 and DF5 of online supplementary data files (Appendix B)). The enriched KEGG pathway analysis results include: Ribosome (29 gene hits), Primary immunodeficiency (11 gene hits), lysosome (22

gene hits), and cytokine-cytokine receptor interaction (35 gene hits) (Table 4.3 and DF5 of online supplementary data files (Appendix B)). The 1,513 genes are involved in Reactome pathways such as Neutrophil degranulation (103 gene hits), Eukaryotic Translation Elongation (31 gene hits), Signaling by Interleukins (66 gene hits) Diseases of Immune System (8 gene hits) Fc epsilon receptor (FCERI) signaling (24 gene hits) and Signaling by the B Cell Receptor (BCR) (21 gene hits) (see DF6 of online supplementary data files (Appendix B) for full table). We also used the KEGGPathwayVisual function in the MathIOmica package to highlight whether our gene hits for the enriched KEGG pathways were up- or down- regulated in the pathway (based on TukeyHSD calculated differences in means) (Figure 4.4 and Figures B.2-B.6 of Appendix B). For example, Figure 4.4 depicts the Primary Immunodeficiency KEGG pathway and highlights our gene hits (with yellow: up-regulated, and blue: down-regulated gene expression). In this pathway, Figure 4.4, our results indicate that Ig α is down-regulated in COPD compared to controls (involved in differentiating from a Pro-B Cell to a Pre-B cell 1), and also BTK9 is up-regulated in COPD (involved in differentiating from Pre-B1 cell to Pre-B2 cell).

Of the 1,513 disease genes we further filtered our ANOVA results (see DF4 of online supplementary data files (Appendix B)) to identify genes with statistically significant interactions with smoking status (disease:smoking status, BH-adjusted p-value < 0.05). We found 39 genes that had a statistically significant pairwise interaction between disease status and smoking status (see DF14 of online supplementary data files (Appendix B)). Using the 39 interacting genes, we calculated the row means across the different pairings of smoking status and disease status to compare expression (Figure 4.8). We used the row means of the non-smoking controls as our baseline to calculate the difference in means for the different disease and smoking groups. In Figure 4.8 the data clusters by disease state (COPD together and controls together), and smokers and former smokers across both disease states have similar expression profiles. There are subset of genes that are over expressed in COPD smokers compared to control non-smokers as well as a subset of genes that are down-regulated. Finally, control smokers and former smokers have similar expression profiles with GGT6 being an outlier (Figure 4.8).

4.4.3 Up and Down- Regulated Gene Expression in COPD

To assess biological effect and determine factorial differences in gene expression we ran TukeyHSD on our 1,513 statistically significant disease genes. We first focused on COPD and control gene expression differences and used BH-adjusted p-value <0.05 to determine significance. We also filtered further by using a 10% two-tailed quantile cutoff to identify significantly up- and down- regulated genes. Once we filtered by p-value, we calculated to 10 and 90% quantiles using differences in group means. For the COPD-control TukeyHSD comparisons we found 304 statistically significant genes that we classified as down-regulated (mean differences ≤ -0.0260) and up-regulated (mean differences ≥ 0.0338) in our COPD subjects. Of the 304 differentially expressed genes (DEG), 152 genes were down-regulated and 152 genes were up-regulated (DF9 of online supplementary data files (Appendix B)). The top 25 up- and down- regulated genes are displayed in Table 4.4. KEGG enrichment analysis on the 152 down-regulated disease genes resulted in two significantly enriched pathways: Hematopoietic cell lineage (5 Gene hits: CD2, CD3E, CD7, FLT3LG and MS4A1) and Cytokine-cytokine receptor interaction (8 gene hits: CCL5, CCR6, CD27, CXCR3, CXCR6, FLT3LG, IL2RB, and IL2RG). For the Reactome enrichment analysis on 152 up-regulated genes, they were enriched in Neutrophil degranulation (30 gene hits) Figure 4.6, while the down-regulated genes were enriched in the Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell pathway (8 gene hits) (Figure 4.7).

Up-Regulated		Down-Regulated	
Gene	Difference of Means	Gene	Difference of Means
GPR15	0.123	LBH	-0.032
HK3	0.046	CD3E	-0.039
CLEC4D	0.075	DUSP7	-0.030
F5	0.050	TCF7	-0.031
DOCK4	0.063	RRAS2	-0.045
GPR55	0.045	ST6GAL1	-0.029
STAB1	0.043	PYHIN1	-0.042
ASGR2	0.046	CCNK	-0.026
ARG1	0.087	CD79A	-0.053
MPO	0.063	GORASP2	-0.028
NRG1	0.064	IL2RG	-0.029
HP	0.086	LRIG1	-0.028
PLD1	0.037	PURA	-0.027
CLEC4E	0.041	LPAR5	-0.033
PLSCR1	0.048	FAM102A	-0.030
FCGR1B	0.051	UCP2	-0.027
FKBP5	0.035	SPON1	-0.035
ANG	0.042	B3GNT7	-0.042
DSC2	0.063	CAMK2N1	-0.040
OSBPL1A	0.035	DENND2D	-0.027
TLR5	0.040	IGFBP4	-0.036
FLVCR2	0.035	IL2RB	-0.036
NLRC4	0.037	CD74	-0.031
AHRR	0.063	CBLB	-0.031
CAPNS2	0.043	DCXR	-0.037

Table 4.4: Top 25 up and down regulated differentially expressed genes in COPD based on effect size

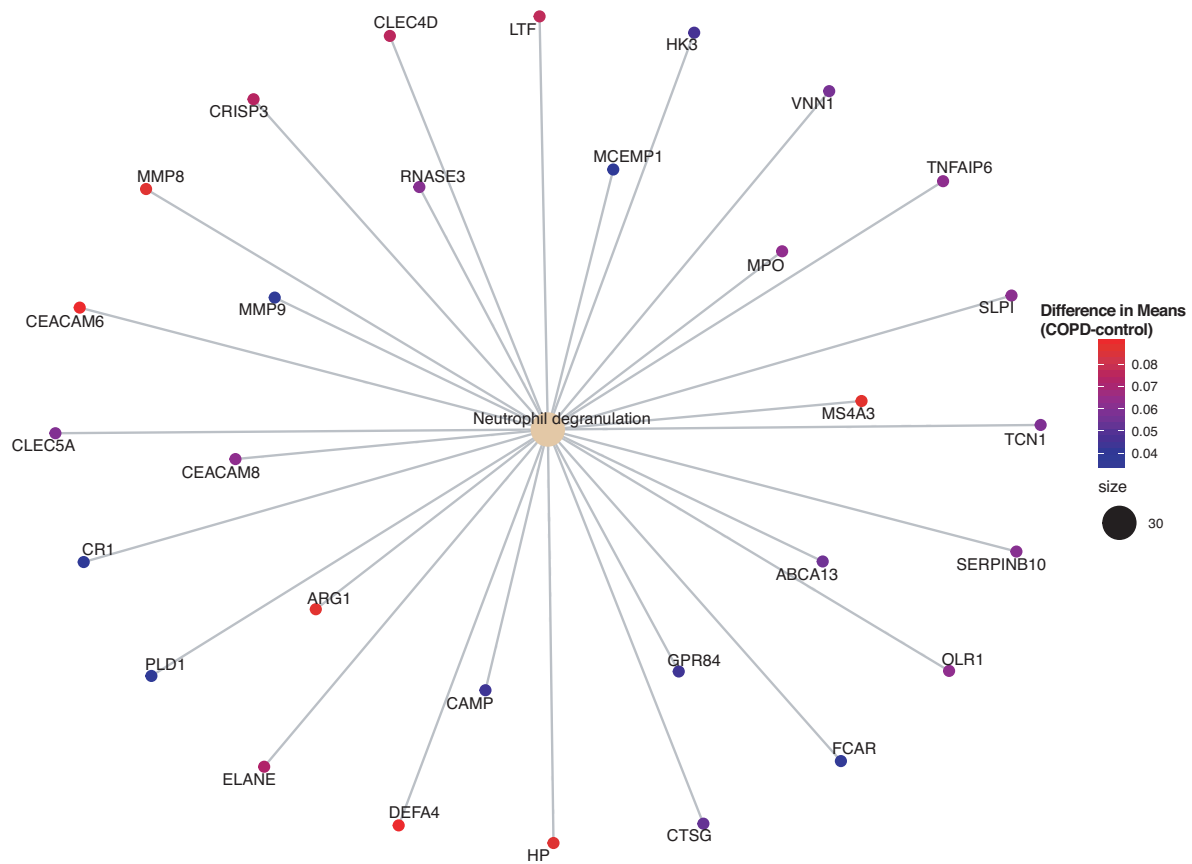


Figure 4.6: Enriched Reactome pathway-gene network from up-regulated disease genes in COPD subjects. The enrichment analysis was based on the 304 statistically significant differentially expressed genes filtered for effect size.



Figure 4.7: Enriched Reactome pathway-gene network from down-regulated disease genes in COPD subjects. The enrichment analysis was based on the 304 statistically significant differentially expressed genes filtered for effect size.

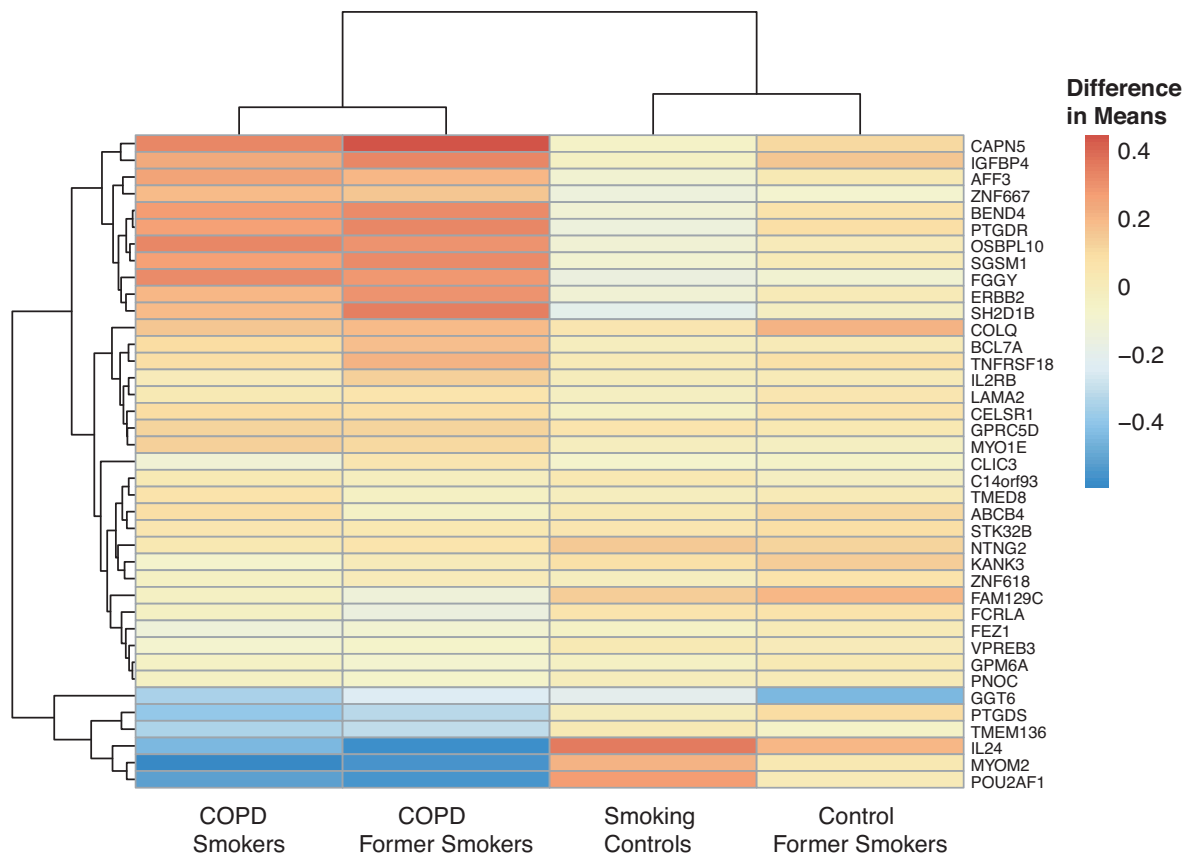


Figure 4.8: Heatmap of statistically significant interacting genes across disease states and smoking statuses. Difference in means calculated using control non-smokers as the baseline.

4.4.4 Sex and Age on COPD Expression

We further analyzed the 304 DEG found to have a biological effect by disease status to identify sex and aging effects on gene expression. We found 44 genes that were differentially expressed by sex: 22 up- and 22 down-regulated in males compared to females by filtering the mean differences using two-tailed 10% quantiles, ≤ -0.0957 (down-regulated) and ≥ 0.0908 (up-regulated). With the 44 genes we performed pathway enrichment analysis using the ReactomePA package. There were 7 enriched Reactome pathways (BH-corrected p-value <0.05) that were all up-regulated in males (see DF12 of online supplementary data files (Appendix B) and Figure B.7 of Appendix B). These pathways include: Neutrophil degranulation (13 gene hits), Antimicrobial peptides (5 gene hits), Extracellular matrix organization (6 gene hits), Activation of matrix metalloproteinases (3 gene hits) and degradation of the extracellular matrix (4 gene hits). We did not find any statistically significant interacting genes between disease status and sex from our ANOVA results.

To determine the age effect on our DEG associated with COPD (304 genes), we focused on our TukeyHSD results where the age group <50 was the baseline. We selected for significance (BH-adjusted p-value <0.05) and two-tailed 10% (up-regulated ≥ 0.421 and down-regulated ≤ -0.193) on the difference in means results to find significant age-group effects. We identified 304 significant age-group comparisons across 95 unique genes (see DF13 of online supplementary data files (Appendix B)). We plotted the relative expression (difference in means) across all age comparisons with <50 as the baseline. We identified two clear clusters of the genes by expression which indicated that there are significant differences in expression profiles due to aging (Figure 4.9). However, we did not find any statistically significant genes with an interaction between disease status and age from our ANOVA results.

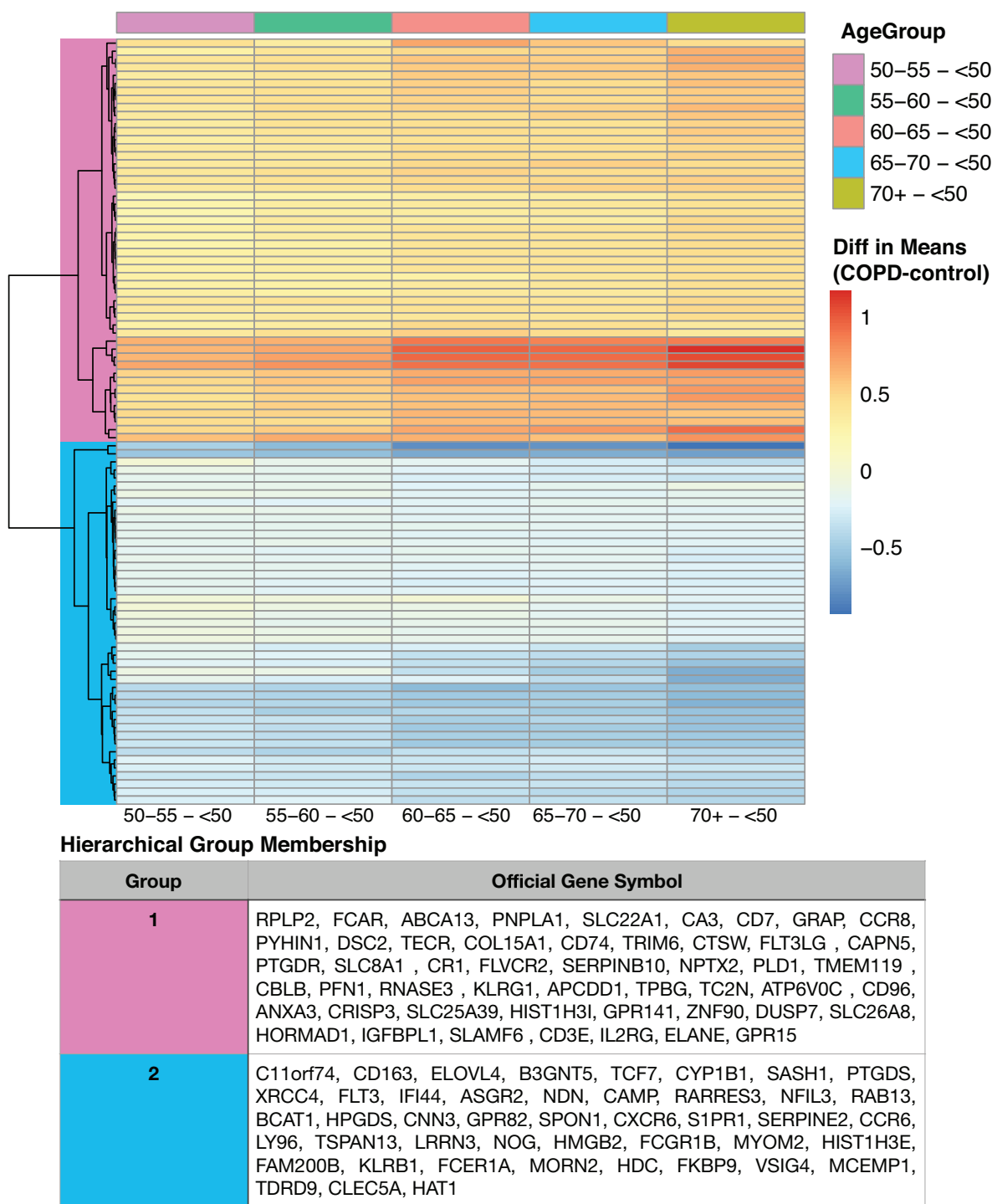


Figure 4.9: Heatmap of age effect on the statistically significant disease gene list. The enrichment analysis was based on the 304 statistically significant differentially expressed genes filtered for effect size. The clustered groups are color-coded, with the corresponding genes in each group listed in the table.

Machine Learning with COPD Data

Using the gene expression from the top 304 statistically significant for disease genes, and with 10% two-tailed highest effect size we trained a logistic regression model in Mathematica for predicting whether a profile belongs to the control or COPD group. Training with all samples achieved an accuracy of $87.0 \pm 3.0\%$, (Fig 4.10A). The corresponding confusion matrix and receiver operating characteristic (ROC) curves are shown in Fig 4.10 respectively, with an ROC area under the curve (AUC) of 0.979. Furthermore, we decided to carry out a 10-fold cross-validation analysis of randomized order samples, where we trained on 90% of the data each time and tested on the remaining 10%. On average the model had an accuracy of 84.2% (standard deviation of 3.1%), and ROC AUC of 0.921 (standard deviation of 0.022). An example of the worst performing realization from the cross-validation is shown in Fig. 4.10D-F, where 48/57 controls and 42/69 COPD samples were classified correctly, whereas 9/57 controls were mis-classified as COPD, and 17/69 COPD were misclassified as controls. Equivalently, the false positive rates were on average 0.17 (control) and 0.14 (COPD), and the false discovery rates were on average 0.19 (control) and 0.12 (COPD).

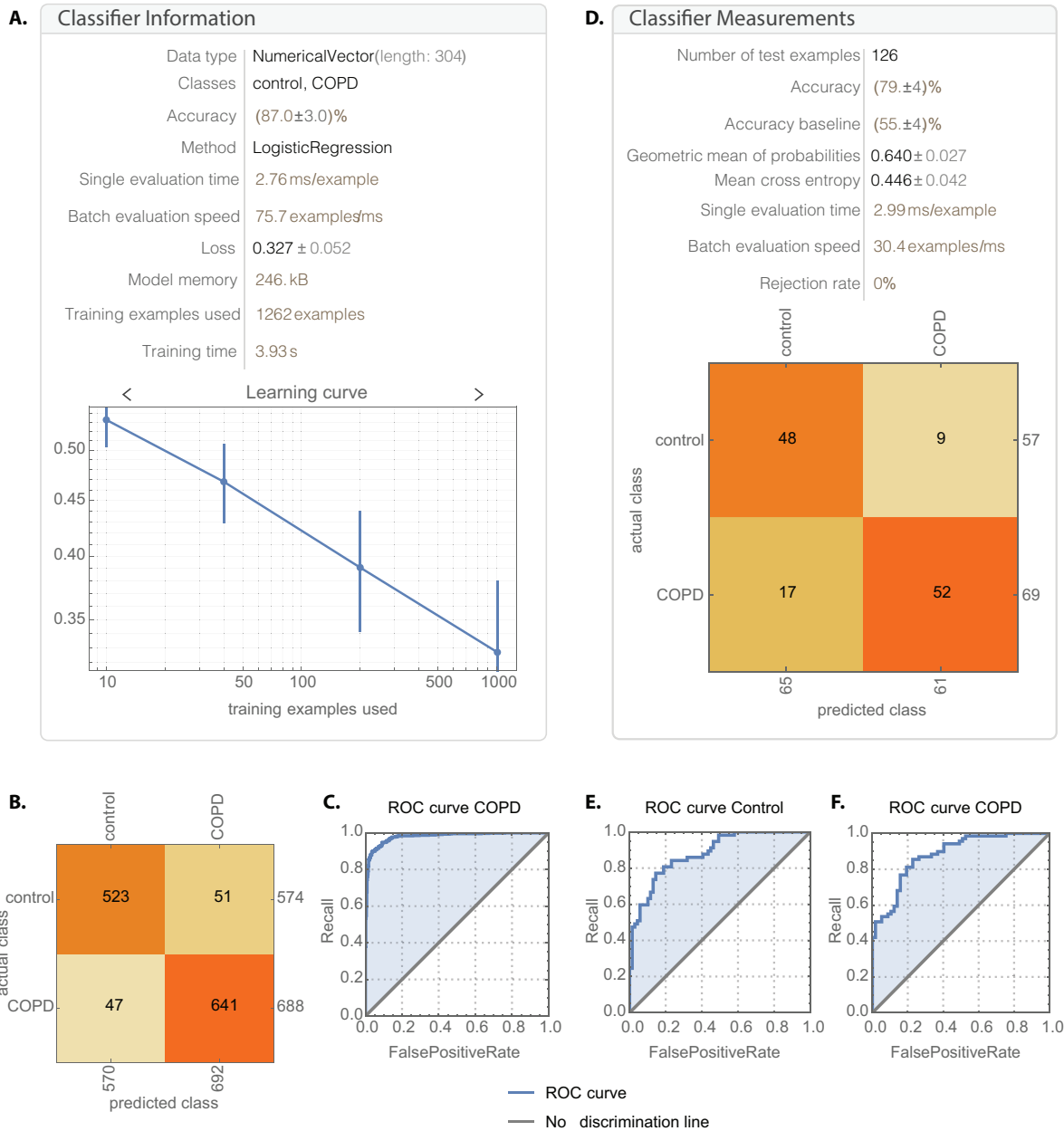


Figure 4.10: Trained logistic regression model can classify COPD and healthy profiles. (A) The logistic regression model trained on all the data achieves 87.0±3.0% accuracy, with the (B) confusion matrix and (C) ROC curves indicating good performance overall, with AUC 0.979. Training with 10-fold cross validation gives an average accuracy of 84.2%, with the worst testing model shown in (D) and its ROC for (E) Controls and (F) COPD shown respectively, with an AUC of 0.882.

4.5 Discussion

Chronic obstructive pulmonary disease causes damage to the lungs because of exposure to toxic irritants or genetic factors, and is a rising global health problem. With an increase in the elderly population's life expectancy and the number of smokers, the prevalence of COPD and its morbidity rates are expected to rise. Researchers are working to identify strategies that can help to clearly understand COPD, its pathology, and to find biomarkers in easily accessible body fluids to promote earlier detection of COPD and improve accuracy of diagnosis [5, 524, 525]. Our research objective was to identify age, sex and smoking status effects on gene expression between COPD and controls in blood. We curated and downloaded 7 microarray expression datasets for our meta-analysis on COPD. Using the raw expression data, we removed the background, annotated and summarized the probes, and merged the 7 datasets together by common gene names. This was followed by data normalization using BoxCox power transformation and downstream analyses to identify differentially expressed genes and genes that were biologically significant. This is the largest COPD meta-analysis and explores expression variability in 1,262 samples by modeling linear and binary effects of disease status, age, sex and smoking status.

Our ANOVA highlighted 1,513 statistically significant (BH-adjusted p -value <0.05 ; disease status factor) disease genes (see DF4 of online supplementary data files (Appendix B)). One of our genes, FAM13A, has previously been associated with COPD susceptibility [517, 539]. Other genes such as GPR15, CLEC4D and MPO have also been associated with COPD and inflammation within the lungs. Our GO and pathway enrichment results highlight some immune pathways (Table 4.3) and GO terms such as innate immune response, adaptive immune response and inflammation (DF7 of online supplementary data files (Appendix B)) that have previously been associated with COPD. For example, primary immunodeficiency (weakened immune system due to deficiencies in immune cell production) is linked to recurrent infections in subjects with COPD [540]. This recurrence in infections due to a weakened immune system also causes chronic inflammation and airway remodelling and obstruction [540]. Humoral deficiencies and inadequate antibody production and responses to infections also reduce the effectiveness of vaccinations such as the influenza and

pneumococcal vaccines [540, 541]. Studies have suggested that antibody replacement therapy can help to reduce the recurrence of bacterial and viral infections in COPD subjects [540, 542]. In our results, we highlighted the primary immunodeficiency KEGG pathway to determine how our genes are regulated in the pathway (Figure 4.4). Of the 8 genes highlighted in the T cell maturation portion of Figure 4.4, 6 of them are down-regulated in COPD subjects compared to our controls. IL2RG (alias gamma chain(γ C)) that regulates T cell development and differentiation was down-regulated in COPD subjects (Figure 4.4). IL2RG is also associated with severe combined immunodeficiency [92]. Previous findings suggest that down-regulation of the soluble common gamma chain is a mechanism to reduce inflammation by T cells in response to cigarette smoke in a COPD mouse model[543]. Up-regulated γ C promotes interferon- γ production and inflammation in the respiratory tract[543]. IL7-Ra and JAK3 are also linked to severe combined immunodeficiency. DCLRE1C (Artemis) and CD3E are both involved in Pro-T to Pre-T cell differentiation and were up- and down-regulated respectively in COPD subjects (Figure 4.4). Genes such as LCK ZAP70 and RFXAP are involved in T cell differentiation into CD8+ and CD4+ cells and were found to be down-regulated in COPD (Figure 4.4). In B-cell differentiation, our gene hits, BTK (B-cell development) and IKBKG (alias IKK γ) were up-regulated in COPD while Ig α was down-regulated (Figure 4.4). Reduced Ig α or deficiencies in Ig α promote reoccurring infections and disease exacerbation in COPD subjects [540, 544].

In the highlighted Cytokine-cytokine receptor interaction KEGG pathway there are different classes of cytokines such as chemokines, class I cytokines and the Tumor necrosis factor and Transforming growth factor beta families with varying expression (Figure 4.5). Cytokines play a major role in the inflammatory response observed in COPD subjects. For instance, CCR8 (chemokine) was up-regulated in COPD subjects (Figure 4.5). Increased levels of CCR8 has been previously observed in allergic asthmatics [545] and has a functional role in macrophage processes and release of cytokines in the lungs [546].

We also visualized our up- and down- regulated gene hits in the other enriched KEGG pathways (Table 4.3 and Figure B.2 - Figure B.6 of Appendix B). We highlighted our 45 gene hits in the

Pathways in Cancer KEGG pathway (Figure B.2 of Appendix B). COPD is a known risk factor for lung cancer and it leads to 1% of cancer cases each year [547]. Furthermore, there is a five-fold increase to developing lung cancer in patients with COPD compared to individuals with normal pulmonary function [547]. Some of our highlighted genes are involved in apoptosis (Fas and CASP9), DNA damage (MDM2), Extra-cellular matrix (ECM) receptor interaction (ECM) and proliferation (CyclinD1) (Figure B.2 of Appendix B). As for the KEGG Lysosome pathway (Figure B.3 of Appendix B), lysosome function and distribution in the cells of COPD subjects and smokers have been previously examined. The lysosomes in smokers have been previously shown to cluster around the nucleus of the cell and with reduced concentrations of lysosomes throughout the cell compared to subjects who did not smoke. Additionally, dysregulation of the lysosomal pathway has also been previously described in COPD patients [548].

We observed some down-regulated genes in the adherens junction pathway for COPD subjects (Figure B.4 of Appendix B). This may be connected to the increase in lung epithelial permeability due to smoking. Also, one study highlighted that apical junctional complex (AJC) genes were down-regulated in COPD smokers, and that the cigarette smoke promotes a cancer-like molecular phenotype by causing reprogramming of transcription of the AJC [549]. The hematopoietic cell lineage pathway highlights genes involved in the differentiation of immune cells from hematopoietic stem cells (Figure B.5 of Appendix B). As for the enriched measles pathway, research suggests that heavy smokers who had childhood measles has an increased risk for developing COPD [550]. The Reactome pathway analysis also resulted in immune related pathways such as Neutrophil degranulation, Signaling by Interleukins, Diseases of the Immune System and Signaling by the B Cell Receptor which all highlight components of the pathology of COPD (DF6 of online supplementary data files (Appendix B)).

Focusing on the 304 differentially expressed disease genes (filtered for biological effect), some of the top up-regulated genes are GPR15 (found on lymphocytes and involved in trafficking of lymphocytes), HK3 (glucose metabolism), CLEC4D (role in inflammation and immunity) and F5 (blood coagulation factor) [92] (Table 4.4. As for our top down-regulated genes CD3E (role

in T-cell development), DUSP7 (involved in MAPK signaling), TCF7 (role in natural killer cell development), RRAS2 (involved in cell proliferation). We also wanted to compare our gene list to a previously published meta-analysis. Reinhold et al., had a total of 6,243 genes which they grouped into 15 modules for each cohort [5]. Out of our 304 genes, 97 of them overlapped with their findings while 207 of our genes were unique. We used BINGO in Cytoscape v.3.7.1 for GO analysis on our 207 unique genes (Figure B.8 of Appendix B) [38, 39]. Our BINGO results (BH-adjusted p-value < 0.05) include GO terms such as defense response, response to bacterium, response to stress, response to wounding, immune response, cell adhesion, and inflammatory response (Figure B.8 of Appendix B).

In addition to exploring enriched GO terms associated with our 304 disease genes, Figure 4.6 - 4.7 highlight the genes that were up-regulated in COPD and were enriched in the Reactome pathways. Neutrophil degranulation (Figure 4.6) (genes up-regulated in COPD), and Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell pathway (genes up-regulated in COPD). Neutrophil degranulation (release of granules by exocytosis) has been associated with pulmonary disorders including asthma and COPD. In COPD patients' neutrophils are the highest number of inflammatory cells present in the bronchial walls [551]. Increase neutrophil degranulation induces tissue damage and this is due to high inflammatory state and constant priming of neutrophils by cytokines and chemokines [551]. Our up-regulated genes in the neutrophil granulation pathway include CEACAM6 (cell adhesion), MMP8 (tissue remodeling and breakdown of extracellular matrix), CLEC4D (cell-adhesion, cell signaling and inflammation), LTF (granules in neutrophils), MS4A3 (signal transduction), and DEFA4 (defense antimicrobial peptides). Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell pathway down-regulated genes include KLRB1 and KLRG1 (role in the regulation of natural killer cell function), CD3E (involved in adaptive immune response), ICAM2 (leukocyte adhesion and recirculation), SLAMF6 (natural killer cell activation) and CD81 and CD96 (role in adaptive immunity) [92].

To assess the effect of smoking status on gene expression, we focused on the biologically significant genes with a significant interaction between disease status and smoking status. We identified

39 disease genes that significantly interacted with smoking status (Figure 4.8). The baseline in Figure 4.8 was non-smoking controls. For the two control groups: current and former smokers, they both have elevated gene expression levels compared to non-smoking controls. This indicates changes due solely to smoking with moderate differences between former and current smokers. As for the COPD smokers and non-smokers, the majority of these genes are elevated compared to non-smoking controls with GGT6, PTGDS, TMEM136, IL24, MYOM2 and POU2AF1 being down-regulated in COPD compared to healthy non-smokers. Some of these genes have been associated with lung function and disorders such as GGT6 which plays a role in glutathione homeostasis and lung airspace epithelial barrier [552], IL-24 can induce apoptosis and helps control cancer cells [553] and POU2AF1 is a regulator of host defenses but cigarette smoke suppresses its gene expression [554] (Figure 4.8). In our analysis there was only 1 COPD non-smoker which was excluded from this analysis.

As for sex specific effects on gene expression, we identified 44 of the 304 disease genes to have a sex effect. The enriched pathways from using the genes that were up-regulated in males are highlighted in Figure B.7 of Appendix B. These genes are involved in Reactome pathways such as Neutrophil degranulation, Extracellular matrix organization, Collagen degradation, Degradation of the extracellular matrix, and antimicrobial peptides (Figure B.7 of Appendix B). Neutrophil degranulation was discussed above as being up-regulated by disease status in COPD subjects compared to controls. In COPD, the extracellular matrix of the airway and parenchyma of the lungs are restructured [555, 556]. Previous findings observed altered expression of elastin and collagen in COPD compared to controls, and the stage/severity of COPD affected extracellular matrix remodeling [555, 556]. Studies on COPD and sex, previously suggested higher prevalence in males due to them having higher smoking rates [557, 558]. However, currently with larger numbers of women smoking the prevalence of COPD in women is on the rise. Studies have shown that women are 50% more susceptible to COPD than males and why this is the case is still an on going debate [557, 558]. Some reasons include, smaller airways so larger concentrations of tobacco smoke in the lungs and hormonal effects [557, 558]. Of the 44 genes with a sex effect, we did not

find any genes with a significant interaction between disease status and sex.

Aging trends were visualized on the biologically significant disease genes. 95 genes showed significant aging trends compared to our baseline (<50) (DF13 of online supplementary data files (Appendix B)). Symptoms for COPD can be detected between ages 40 and 50 [559], and because of this we used our subjects grouped as <50 as our baseline. The data clustered into two distinct groups with similar gene expression patterns (Figure 4.9). Group 1 genes were significantly up-regulated for all age groups compared to the baseline (Figure 4.9). Pathway enrichment analysis indicated that the genes in group 1 are involved in the Neutrophil degranulation pathway (p-value 1.06e-04 and FDR 0.029) which has previously been described above as being up-regulated in COPD subjects. Genes within group 2 displayed an opposite trend with most genes being down-regulated with increasing age (Figure 4.9). These genes did not result in any statistically significant enrichment. However, genes in group 2 include C11orf74 (involved in transcription regulation) [92], CD163 (previously found to be over expressed in lungs of individuals with severe COPD) [560], TCF7 (natural killer and lymphoid cell development) [92], CYP1B1 (previously shown to be up-regulated in COPD and smokers) [561] and SASH1 (involved in TLR4 signaling and can promote cytokine production) [92]. In addition to this, we did not find any significant interacting genes between disease status and age.

To test the possibility of using blood expression data from micro-arrays to predict disease status, we performed machine learning with a logistic regression model using the 304 disease genes. This resulted in an average accuracy of 84.2% (Figure 4.10). These results are promising despite using aggregate expression versus cell-type specific expression. Previous studies explored using computed tomography (CT) images COPD patients and controls for disease classification [562]. Some studies also used patient reported data (such as heart rate, respiratory rate) to predict disease exacerbation and resulted in an ROC of 0.87 [563] and another with 70% sensitivity and 71% specificity [564].

Conducting a meta-analysis with microarray expression data limits our findings to annotated genes, and hinders us from discovering novel genes and looking at the entire transcriptome. Ad-

ditionally, using publicly available data limits us to specific factors we can explore in our analysis due to subject characteristics not being reported uniformly across datasets (see B.1 of Appendix B). For example, all studies did not report ethnicity and therefore we could not investigate the effect of ethnicity on gene expression in COPD. This would be a good factor to explore due to over 90% of COPD cases occurring in low-middle class communities [515, 519]. We also did not have consistently reported disease severity information to factor into our analysis and findings. Our selection criteria for the publicly available data limits our sample size (Figure 4.2). In addition to this, the limitations of available data resulted in unbalance in sample constitution: 1,262 samples with 574 controls and 688 COPD, of which 792 are males and 470 females, and have smoking status as 183 non-smokers, 418 smokers, and 661 former smokers. As for our machine learning algorithm, despite having a good predictive power and accuracy, we could not explore cell-type specific data. Furthermore, the observed confounding between studies suggests that samples would need to be analysed together with the current sample sets in new investigations, prior to prediction of status.

Our study highlights new gene candidates by factor (disease status, age, sex and smoking status) and genes that statistically interact between disease status and smoking status that can be studied further to understand their role in COPD. Future work to expand on our findings must include the use of cell-type specific expression data and RNA-sequencing data. Due to COPD being characterized by inflammation, increased macrophages and neutrophils and their release of cytokines, looking at cell-type specific data can give more insight on pathology of COPD. Using cell-type specific data for predicting disease states will also expand on our findings. RNA-sequencing data can introduce novel gene candidates and biomarkers for COPD. Furthermore, implementing proteomics and metabolomics can help characterize disease pathology and may lead to discovery of additional signatures for early detection of COPD using a systems biology approach.

CHAPTER 5

MICROARRAY GENE EXPRESSION DATASET RE-ANALYSIS REVEALS VARIABILITY IN INFLUENZA INFECTION AND VACCINATION.

Work presented in this chapter has been submitted to Frontiers in Immunology and a pre-print is available on bioRxiv: Rogers LRK, de los Campos G, Mias GI. Microarray Gene Expression Dataset Re-Analysis Reveals Variability in Influenza Infection and Vaccination. bioRxiv 702068; doi:<https://doi.org/10.1101/702068>

5.1 Abstract

Influenza, a communicable disease, affects thousands of people worldwide. Young children, elderly, immunocompromised individuals and pregnant women are at higher risk for being infected by the influenza virus. Our study aims to highlight differentially expressed genes in influenza disease compared to influenza vaccination, including variability due to age and sex. To accomplish our goals, we conducted a meta-analysis using publicly available microarray expression data. Our inclusion criteria included subjects with influenza, subjects who received the influenza vaccine and healthy controls. We curated 18 microarray datasets for a total of 3,481 samples (1,277 controls, 297 influenza infection, 1,907 influenza vaccination). We pre-processed the raw microarray expression data in R using packages available to pre-process Affymetrix and Illumina microarray platforms. We used a Box-Cox power transformation of the data prior to our down-stream analysis to identify differentially expressed genes. Statistical analyses were based on linear mixed effects model with all study factors and successive likelihood ratio tests (LRT) to identify differentially-expressed genes. We filtered LRT results by disease (Bonferroni adjusted p-value < 0.05) and used a two-tailed 10% quantile cutoff to identify biologically significant genes. Furthermore, we assessed age and sex effects on the disease genes by filtering for genes with a statistically significant (Bonferroni adjusted p-value < 0.05) interaction between disease and age, and disease and sex. We identified 4,889 statistically significant genes when we filtered the LRT results by disease factor, and gene enrichment analysis (gene ontology and pathways) included innate immune response, viral process, defense response to virus, Hematopoietic cell lineage and NF-kappa B signaling pathway. Our quantile filtered gene lists comprised of 978 genes each associated with influenza infection and vaccination. We also identified 907 and 48 genes with statistically significant (Bonferroni adjusted p-value < 0.05) disease-age and disease-sex interactions respectively. Our meta-analysis approach highlights key gene signatures and their associated pathways for both influenza infection and vaccination. We also were able to identify genes with an age and sex effect. This gives potential for improving current vaccines and exploring genes that are expressed equally across ages when considering universal vaccinations for influenza.

5.2 Introduction

The influenza virus, a respiratory pathogen, is responsible for seasonal influenza (also known as the flu), influenza pandemics and high rates of morbidity and mortality worldwide [565]. The influenza virus infects the upper respiratory tract by invading the epithelial cells, releasing viral RNA, replicating and spreading throughout the respiratory tract while also causing inflammation [566]. Influenza is a highly contagious disease and spreads easily via contact with an infected person's nasal discharges and cough droplets [9]. The main virulence factors are haemagglutinin (HA) and neuraminidase (NA) [566]. These surface glycoproteins are also important for determining the sub-type of the influenza virus. The influenza virus can also reduce host gene expression through their viral proteins [567, 568]. The viral proteins affect transcription and translation in the host which reduces the production of host proteins and promotes immune system evasion for the virus [567, 568]. The virus interferes with host gene expression to promote viral gene expression, and this affects the immune system of the host by reducing the expression of immune components such as the major histocompatibility (MHC) molecules antigen presentation, and interferon and cytokine signaling pathways [567, 569].

Influenza is a global health burden, and as a preventative method vaccinations are offered annually. Vaccines are modified annually because the influenza virus strains change and mutate every season [570]. The influenza vaccinations target the viral strains and sub-types that researchers predict would be most prevalent each flu season [9, 571]. Furthermore, there are groups in the population who are considered at a higher risk for influenza infection, and they include young children, elderly, individuals who are immunocompromised, and females who are pregnant [9]. The Centers for Disease Control and Prevention (CDC) has estimated, for the 2017-2018 season for influenza, 959,000 hospitalizations and over 79,000 deaths [9]. 90% of the deaths during the 2017-2018 flu season were within the elderly population, while about 48,000 of the hospitalizations were in children [9]. These estimates highlight that young children and especially the elderly are at higher risks for influenza and severe infections that can lead to hospitalization or death. Additionally, the CDC has recommended varying dosages for each vaccine for different age groups due to age-

dependent immune responses [9, 572]. Due to a decrease in efficacy of the influenza vaccines in the 65 and older population, they receive different dosages compared to younger age groups, in order to elicit a beneficial immune response [9, 572]. Contrasting between changes in gene expression due to immunosenescence in healthy subjects and the age-dependent immune responses to diseases such as influenza can help our understanding of how responses to different diseases vary with age. Due to the influenza virus constantly changing and the efficacy of the vaccine being dependent on one's age, researchers have started efforts to develop a universal vaccine [573–575]. The goal is for such a universal vaccine to provide protection to all influenza strains [576]. One approach, is to implement the use of highly conserved influenza peptides in vaccine formulations [575, 576].

Previous studies have investigated global blood gene expression to compare influenza disease to other respiratory diseases to assess severity and pathogenesis [577]. For example, influenza has been shown to induce a stronger immune response than respiratory syncytial virus by producing more respiratory cytokines [577, 578]. Studies also explored responses to vaccinations to highlight gene signatures. In our meta-analysis, our aim was to combine publicly available influenza microarray data to identify the effects of disease state (control, influenza infection and vaccination), age and sex on gene expression. We explored gene expression variation in blood for 3,481 samples (1,277 controls, 297 influenza infected, 1,907 influenza vaccinated) to identify genes and their pathways in influenza (Figure 5.1-5.2). This is to the best of our knowledge, the largest meta-analysis (18 datasets) to explore blood expression changes in influenza infection and vaccination. Our results provide gene signatures and pathways that can be targeted to improve influenza treatment and vaccinations. We also highlight disease associated genes that have interactions with age and sex, that can be used to further explore improving vaccinations, and aid efforts in identifying potential gene targets towards developing universal vaccinations to help reduce the burden of influenza.

5.3 Methods

We curated 18 influenza-related microarray datasets from public database repositories (Table 5.1) to investigate changes in gene expression due to disease status, sex and age. The 18 datasets were

from Affymetrix and Illumina microarray platforms (Table 5.1). We modified and implemented the data-analysis pipeline outlined by Brooks et al.[526]). To achieve our goal, after curating the datasets, we used the R programming language [51] to pre-process the raw gene expression data and to fit linear mixed effects models to determine statistically significant differentially expressed genes by factor (Figure 5.1). In addition, we identified genes that varied in expression due to disease status, sex, and age, and we also determined which gene ontology (GO) terms and pathways enrichment based on these gene sets (Figure 5.1).

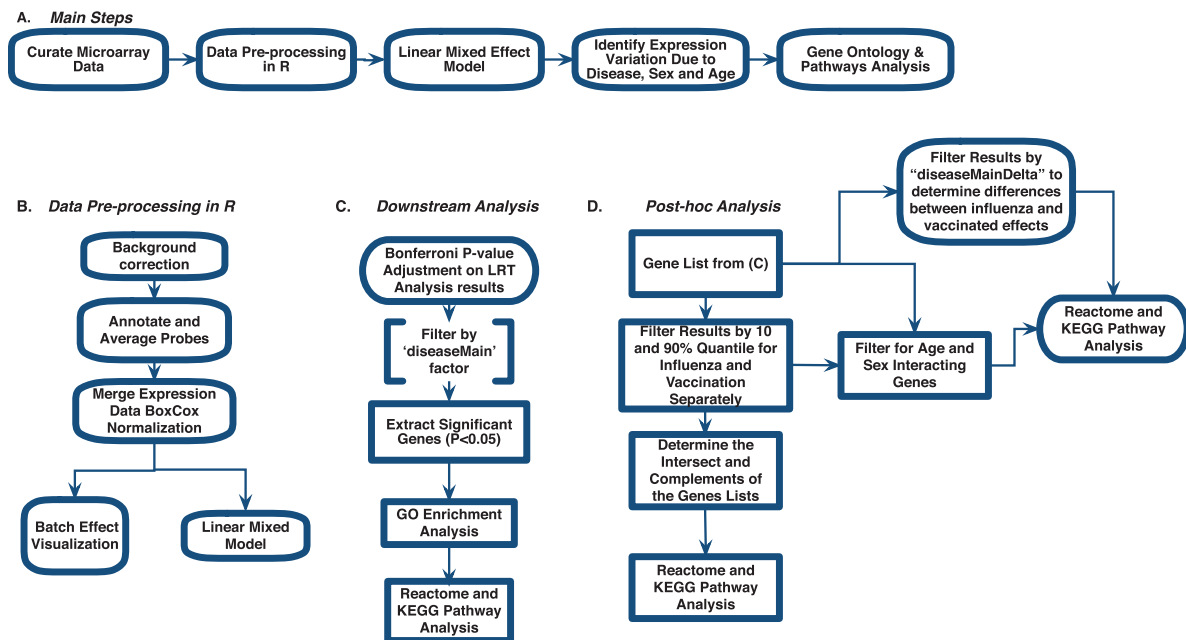


Figure 5.1: Meta-analysis Workflow to Assess Gene Expression Variation in Influenza Disease and Vaccination

Accession Number	Controls	Influenza Disease	Influenza Vaccine	Sex (M/F)	Age Range	Platform	Ref
GSE38900	31	16	0	20/27	0.025 - 1.57	Illumina HumanHT-12 V4.0 expression beadchip	[577]
GSE107990	171	0	500	238/433	23 - 89	Illumina HumanHT-12 V4.0 expression beadchip	[579]
GSE111368	130	229	0	177/182	18 - 71	Illumina HumanHT-12 V4.0 expression beadchip	[580]
GSE27131	7	14	0	16/5	25 - 59	Affymetrix Human Gene 1.0 ST Array	[581]
GSE29614	9	0	18	12//15	22 - 46	Affymetrix Human Genome U133 Plus 2.0 Array	[582]
GSE29615	28	0	55	38/45	21 - 46	Affymetrix HT HG-U133+ PM Array Plate	[582]
GSE47353	117	0	175	122/170	21 - 62	Affymetrix Human Gene 1.0 ST Array	[583]
GSE48762	274	0	150	202/222	22 - 49	Illumina HumanHT-12 V3.0 expression beadchip	[584]
GSE50628	0	10	0	2//8	4 - 9	Affymetrix Human Genome U133 Plus 2.0 Array	[585]
GSE52005	34	0	102	62//74	0.68 - 14.68	Illumina HumanHT-12 V4.0 expression beadchip	[586]
GSE74816	72	0	105	59/118	21 - 80	Affymetrix HT HG-U133+ PM Array Plate	[587]
GSE97485	10	0	0	6//4	27 - 72	Affymetrix Human Gene 1.0 ST Array	[588]
GSE34205	18	28	0	24/22	0.0416 - 11	Affymetrix Human Genome U133 Plus 2.0 Array	[578]
GSE41080	91	0	0	37/54	20 - 93	Illumina HumanHT-12 V3.0 expression beadchip	[589]
GSE74811	28	0	55	23/60	21 - 47	Affymetrix HT HG-U133+ PM Array Plate	[587]
GSE59654	39	0	117	68/88	22 - 90	Illumina HumanHT-12 V4.0 expression beadchip	[590]
GSE48018	111	0	320	431/0	18.2 - 32.1	Illumina HumanHT-12 V3.0 expression beadchip	[591]
GSE48023	107	0	310	0/417	18.5-40.2	Illumina HumanHT-12 V4.0 expression beadchip	[591]

Table 5.1: Demographics of curated influenza microarray datasets.

5.3.1 Data Curation: Gene Expression Omnibus

For our meta-analysis, we focused on influenza infection and vaccination. We searched public database repositories such as Gene Expression Omnibus (GEO) [47], Array Express (AE) [48] and Immune Space (IS) [592, 593] (Figure 5.2). To begin our data search, we found datasets with the keyword "influenza" and filtered for /textitHomo sapiens (Figure 5.2). Following this filter, we then removed duplicate records. For example, there were 15 duplicate records on GEO and 16 datasets on IS overlapped with our GEO records (Figure 5.2). We further filtered the results for datasets that were published, had non-ambiguous annotation, reported the age and sex of all subjects, and used blood or peripheral blood mononuclear cells (PBMCs) as the tissue type (Figure 5.2). Based on our inclusion criteria, we identified 18 datasets on GEO to use for our meta-analysis (Table 5.1 and SDF1 of online supplementary data files (Appendix C)). For datasets such as GSE29614 (SDY64 on IS), GSE29615 (SDY269 on IS), GSE74811 (SDY270 on IS), GSE59654 (SDY404 on IS), GSE74816 (SDY1119 on IS), GSE48023 (SDY1276 on IS), 48018 (SDY1276 on IS) that did not have the ages of the subjects reported on GEO, we used the annotation from IS to gather age and sex characteristics of the samples. Additionally, we excluded 4 duplicates in GSE34205: GSM844139, GSM844141, GSM844143 and GSM844196 (which are duplicates of GSM844138, GSM844140, GSM844142, and GSM844195 datasets respectively).

After filtering through and selecting the datasets to use in our meta-analysis, we downloaded the

raw gene expression data for each dataset, and created a file per study with sample characteristics (Table 5.1 and SDF1 of online supplementary data files (Appendix C)). Our selected datasets were further filtered to remove samples that did not fit our criteria. For instance, GSE38900 and GSE34205 have samples with respiratory syncytial virus (RSV), GSE48762 contains samples who received the pneumococcal vaccine, GSE50628 has samples with rota-virus infection and patients who experience seizures, and GSE97485 has samples with acute myeloid leukemia who received the influenza vaccine. Due to this, we excluded all subjects that had a pre-existing health condition, infections other than influenza and received vaccinations other than the influenza vaccine (SDF1 of online supplementary data files (Appendix C)).

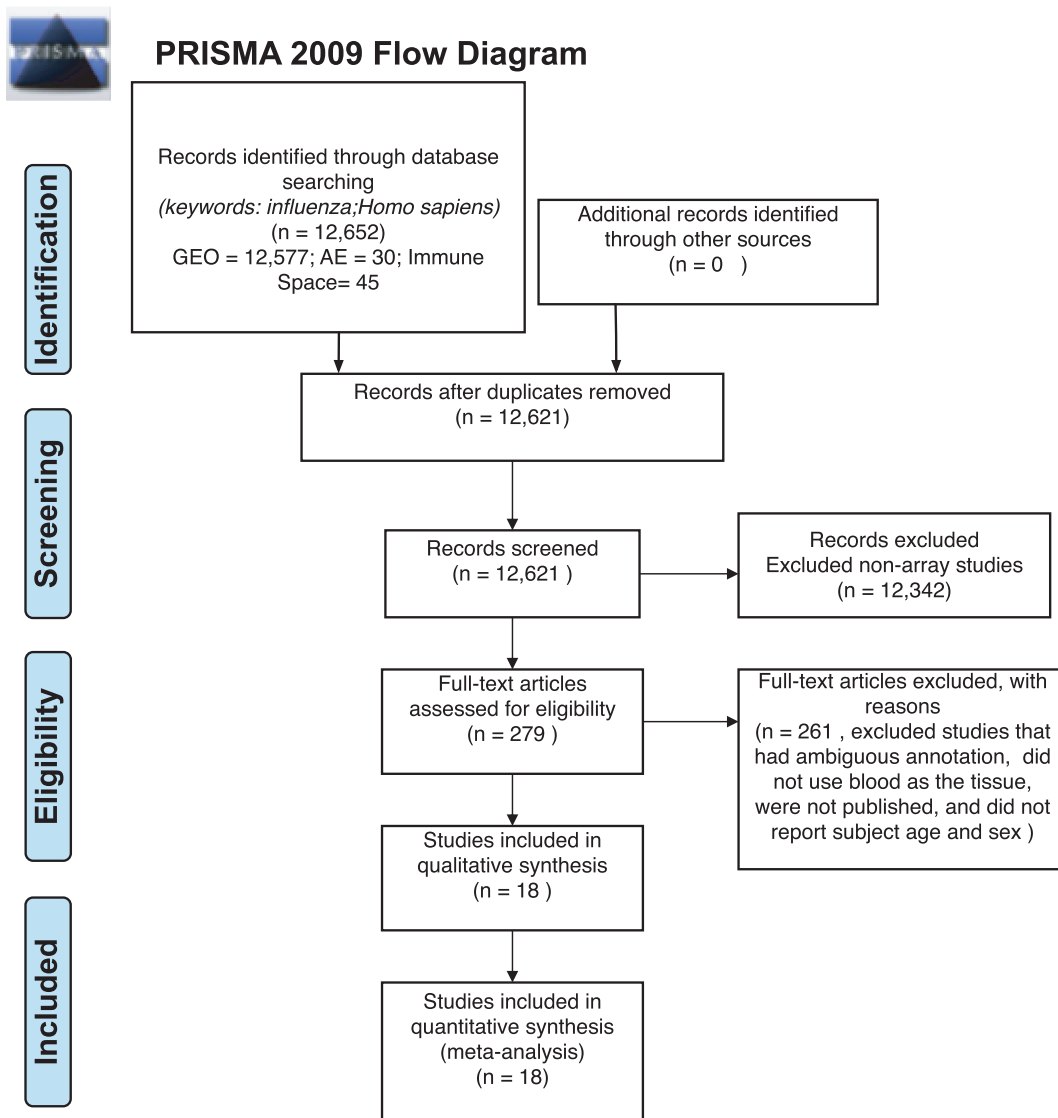


Figure 5.2: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist.

5.3.2 Data Pre-Processing in R and Mathematica

All raw expression files were downloaded directly from the GEO website and pre-processed in R using appropriate packages based on the type of microarray platform (Table 5.1). We carried out background correction and annotated and summarized all probes (Fig 5.1B). We used the `affy` package [52] to pre-process all of the data files for the expression data from Affymetrix Human Genome Plus 2.0 and the Affymetrix HT Human Genome U133 Plus PM. Specifically, we used the `expresso` function to pre-process the files using robust multi-array analysis (RMA) for background correction, conduct perfect-match probe correction, and to calculate expression values using ‘`avdiff`’ [52]. To summarize and remove replicate probes we used the `avereps` function from `limma` [53]. For the Affymetrix HT Human Genome U133 Plus PM, we created our own annotation package in R using the annotation obtained from GEO [51]. For the raw expression data from the Affymetrix Human Gene 1.1 ST microarray platform, we pre-processed the data using the `oligo` [534] and `affycoretools` [535] packages. To background correct the Affymetrix Human Gene 1.1 ST microarray data files we also used RMA and summarized and removed replicate probes using `avereps` function from `limma`. Our Illumina data files were pre-processed with the `limma` package. We used the NormExp Background Correction (`nec`) function from the `limma` package to remove the background of data files that reported the detection p-values. The (`nec`) function using the detection p-values when background correcting. Probes were annotated and summarized using the `aggregate` function from the `stats` package in base R [51, 53].

Following pre-processing, we merged expression data for the 18 datasets (Table 5.1 and SDF1 of online supplementary data files (Appendix C)) by matching gene symbols that were common across all datasets. We conducted a Box-Cox power transformation [54] and standardized the expression values using the functions `ApplyBoxCoxTransformExtended` and `StandardizeExtended` from the MathIOmica (version 1.2.0) package in Mathematica [49, 55] (Fig 5.1B and SDF2 of online supplementary data files (Appendix C)).

5.3.3 Linear Mixed Effects Modeling

We fitted a sequence of mixed-effects models to identify genes whose expression levels were affected by disease status (3 levels: control, influenza, vaccine) and those for which the effect of disease was modulated by either age or sex. Models were fitted using the `lmer` function of the `lme4` R-package [594]. Separate models were fitted to each of the genes. Our baseline model (M0) included the (fixed) effects of sex (M/F), age (a factor with 4 levels, (-1,3],(3,19], (19,65] and (65,100]), ethnicity (a factor with 7 levels, African-American, Caucasian, Asian, Hispanic, Middle Eastern, Other, Unclassified) and tissue (2 levels, blood and PBMCs) plus the random effects of study (18 levels, see Table 5.1 for accession numbers) and of the subject (we included the subject effect because some studies had repeated measures). We first expanded this model by adding the (fixed) main effect of disease status (a factor with three levels, M1). Our next model expanded M1 by adding interactions between disease status and age (M2-DxA) and disease status by sex (M2-DxS). P-values for the main effects of diseases as for disease-by-sex and disease-by-age were obtained using likelihood ratio tests (LRT) between the models described above (SDF3 of online supplementary data files (Appendix C)). LRTs were implemented using the `anova` function from base R to pairs of models. We used a sequential testing approach whereas: (i) we first identified genes with significant main effect of disease (this was based on a LRT between M1 and M0), (ii) among genes with significant main effect of diseases we tested the significance of DxA and DxS using a likelihood ratio test that had M1 as null hypothesis and the interaction models as alternative hypotheses. P-values were adjusted using Bonferroni, where for the first test (i) the number of tests was equal to the number of genes, and for the second one (ii) the number of tests was equal to the number of genes that passed the first test.

The filtering of genes based on Bonferroni-adjusted p-values for the main effect of disease (comparison of M1 to M0) allowed us to identify differentially expressed genes with respect to disease states (Figure 5.1). Using this gene list, we then conducted GO enrichment analysis (`GOAnalysis` function in `MathIOmica` package) and pathway enrichment analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG, using the `KEGGAnalysis` functions in `MathIOmica`),

and Reactome pathway enrichment analysis (`enrichPathway` function from the `ReactomePA` package in R [36]).

5.3.4 Determining Gene Expression Variability between Influenza Infection and Vaccination

We took a sequential testing approach to further analyze the identified statistically significant disease genes (SDF5 of online supplementary data files (Appendix C)). Using this gene list, we further filtered for biological effect by using calculated estimates (which compared influenza and vaccine expression to controls) (SDF4 of online supplementary data files (Appendix C)) and performed a two-tailed 10% quantile filter (i.e. 0.1 and 0.9 quantiles) to determine genes that were biologically significant in subjects who were vaccinated with influenza vaccinated and subjects infected with influenza disease. The biologically significant gene lists for the vaccinated and influenza subjects were further examined to identify genes in common, and genes only in the influenza list, and only in the vaccinated list (Figure 5.1). We performed GO and pathway enrichment analysis on these genes. Lastly, we filtered the disease (see SDF1 of online supplementary data files (Appendix C)) statistically significant gene list for interacting genes between disease and age (age groups: (-1,3],(3,19], (19,65], (65,100])) and disease and sex.

5.4 Results

Our data curation criteria resulted in 3,481 samples (1,277 controls, 297 influenza infection, 1,907 influenza vaccinated, 1,537 males and 1,944 females) (see SDF1 of online supplementary data files (Appendix C)). Our 3,481 samples are from 1,147 individuals. Some studies include repeated measures (in the curated studies individuals were followed for several days after vaccination or infection and varying timepoints were reported as a different samples for the same subject). We included all repeated measures in our downstream analysis and accounted for them in our model. The main results are summarized below, and further discussed in the Discussion Section 5.5 (Figure 5.3)..

5.4.1 Differentially Expressed Genes in Influenza Disease and Vaccination

Filtering our LRT analysis results by disease factor (see SDF3 of online supplementary data files (Appendix C)) for Bonferroni adjusted p-values (< 0.05), we identified 4,889 statistically significant disease genes (see SDF5 of online supplementary data files (Appendix C)). We performed GO enrichment analysis using BINGO in Cytoscape (version 3.7) [38, 39] and pathway enrichment analysis on the 4,889 genes (Figures C.1-C.5 of Appendix C and see SDF6-SDF8 of online supplementary data files (Appendix C)). We identified enriched GO terms such as: cell cycle checkpoint (51 genes), response to stimulus (987 genes), immune response (243 genes), transcription (122 genes), regulation of T-cell activation (62 genes), regulation of defense response to virus by host (8 genes) and immune system process (379 genes) (see SDF8 of online supplementary data files (Appendix C) for full table). We found 75 enriched KEGG pathways (SDF6 of online supplementary data files (Appendix C)). The enriched KEGG pathways include: Cell cycle (68 gene hits), Hematopoietic cell lineage (45 genes), NF-kappa B signaling pathway (46 genes), Metabolic pathways (341 genes), Primary immunodeficiency (23 genes), T cell receptor signaling pathway (44 genes), B cell receptor signaling pathway (29 genes) and also Influenza A (52 genes). We also highlighted the NF-kappa B signaling pathway and the Influenza A KEGG pathways that are relevant to disease with our calculated estimates which compared influenza infection and vaccination expression to that of healthy controls (Figures 5.4 -5.7).

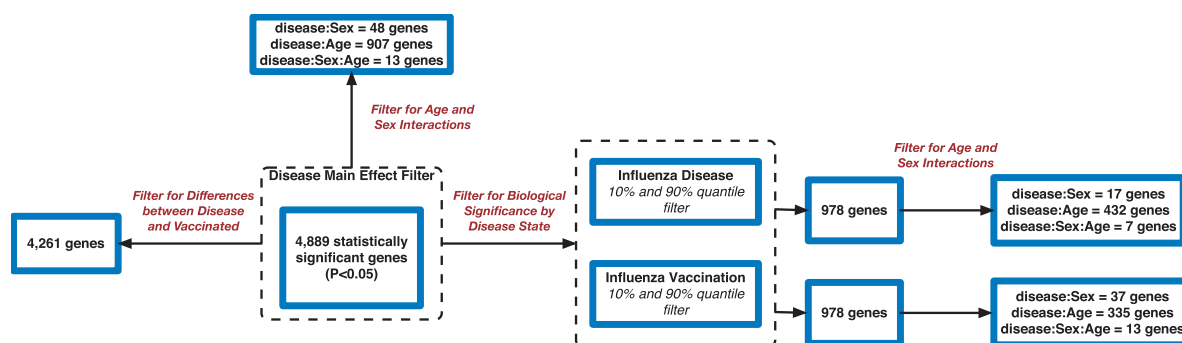


Figure 5.3: Flowchart of Gene Filtering Steps for Influenza Meta-analysis.

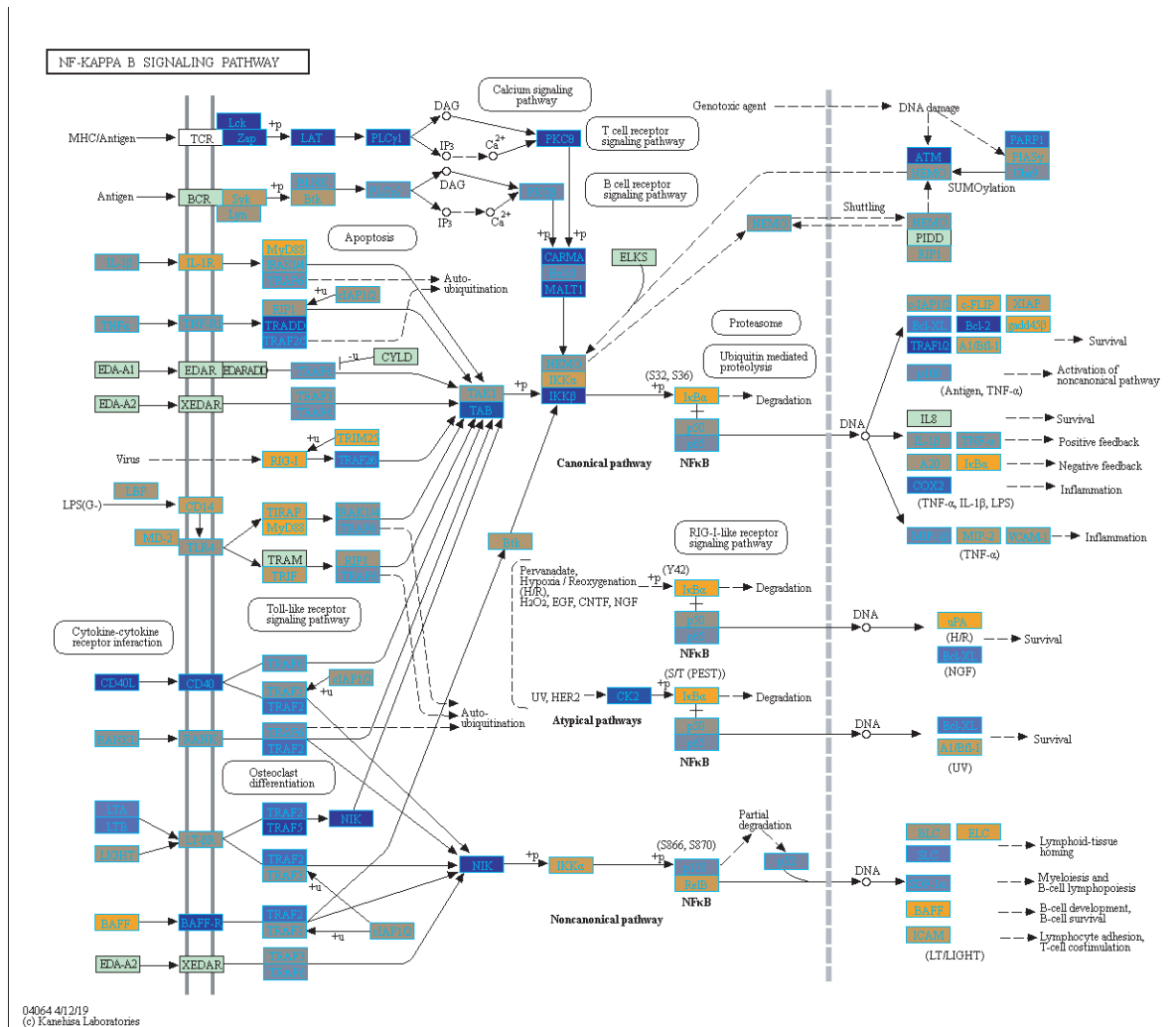


Figure 5.4: Highlighted NF-Kappa B Signaling KEGG Pathway (hsa04040) with Enriched Genes from the LRT Analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Infected Subjects [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Infected Subjects.

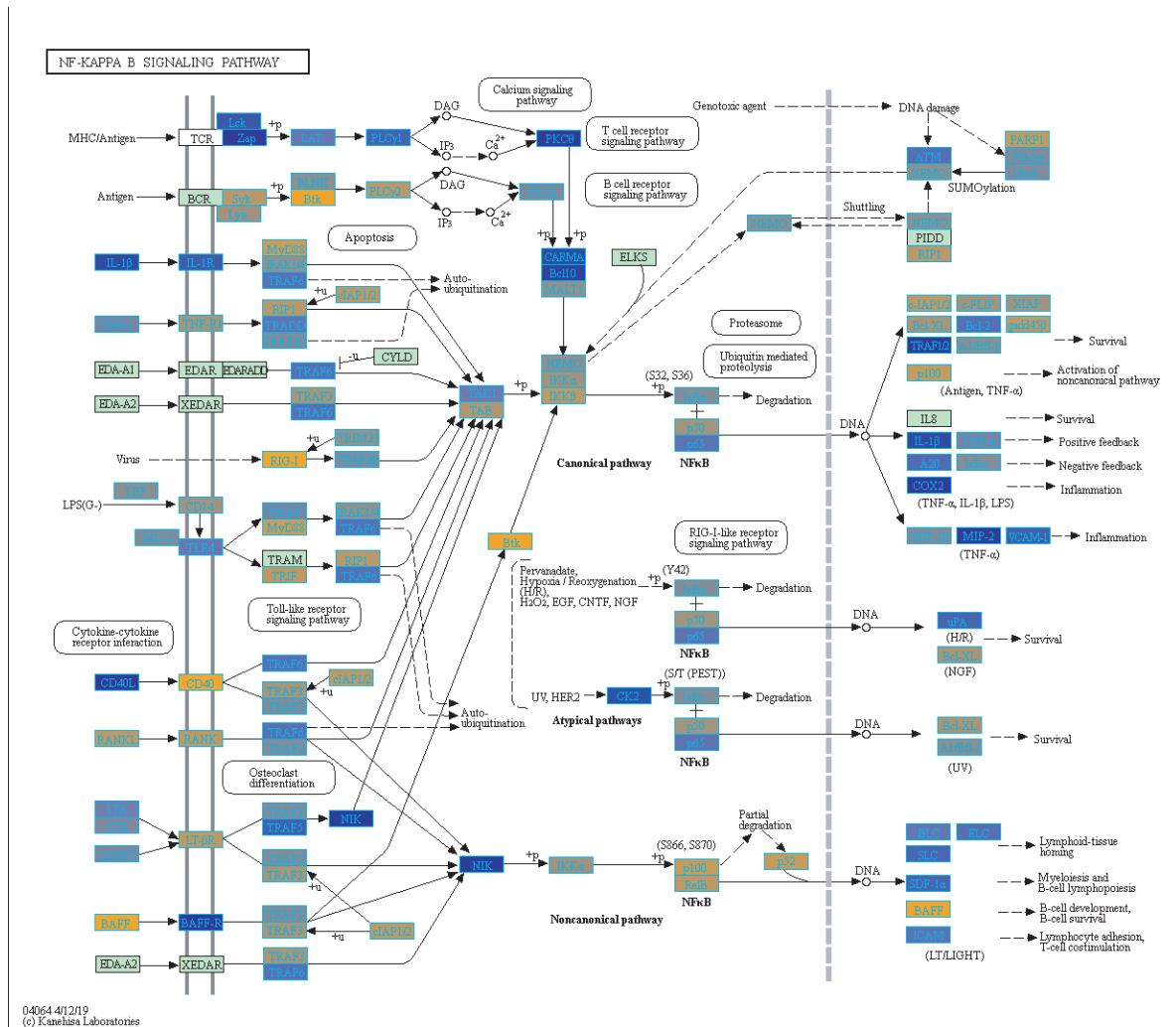


Figure 5.5: Highlighted NF-Kappa B Signaling KEGG Pathway (hsa04040) with Enriched Genes from the LRT Analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Vaccinated Subjects [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Vaccinated Subjects.

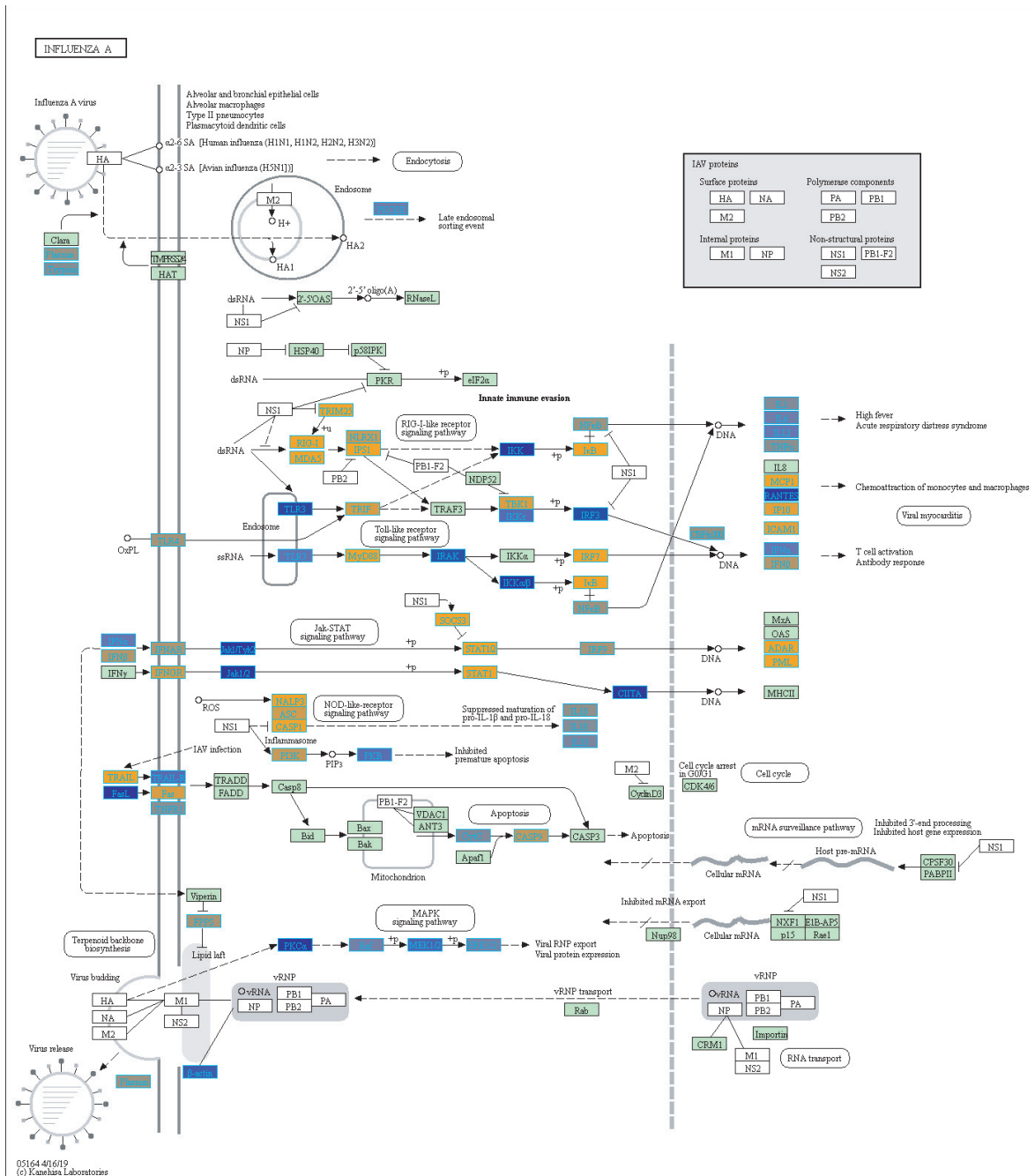


Figure 5.6: Highlighted Influenza A KEGG Pathway (hsa05164) with Enriched Genes from the LRT analysis (Bonferroni-adjusted p-value < 0.05) for Influenza Infected Subjects [1–3]. Yellow-colored genes are up-regulated and blue-colored genes are down-regulated in Influenza Infected Subjects.

90% quantile filter on the 4,889 genes to: (i) analyze the influenza disease estimates (compared expression to control) list to identify genes that are biologically significant and statistically significant (Bonferroni-adjusted p-value <0.05) in influenza infection (ii) analyze the influenza vaccination estimates with the same filtering approach to also identify significant genes for influenza vaccination. For influenza infection our 10% and 90% quantile cut-offs for biological significance were ≤ -0.6724464 and ≥ 0.5949655 respectively. For influenza vaccination, the 10% and 90% quantile cut-offs were ≤ -0.07157763 and ≥ 0.06719048 respectively. For influenza infection we identified 978 genes of the 4,889 to be biologically significant (Table 5.3 and SDF9 of online supplementary data files (Appendix C)), and we also identified 978 genes to be biologically significant for influenza vaccination (Table 5.3 and SDF10 of online supplementary data files (Appendix C)). We then compared the two gene lists to identify the intersection (genes in common), genes only in the influenza disease list, and genes only in the influenza vaccination list (Figure 5.1D and SDF11-SDF13 of online supplementary data files (Appendix C)). There were 334 genes in common across both lists (influenza disease and vaccination) (SDF17 of online supplementary data files (Appendix C)) that resulted in enriched Reactome pathways such as Interferon alpha/beta signaling (14 genes), Interferon gamma signaling (12 genes), Antiviral mechanism by IFN-stimulated genes (9 genes), and Cell Cycle Checkpoints (17 genes) (SDF20 of online supplementary data files (Appendix C)). There were 644 genes that were only in influenza infection list (SDF18 of online supplementary data files (Appendix C)) that resulted in enriched Reactome pathways including: Neutrophil degranulation (45 genes), Cell Cycle Checkpoints (27 genes), Amplification of signal from the kinetochores (13 genes), Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal (13 genes) and Mitotic Spindle Checkpoint (14 genes) (SDF21 of online supplementary data files (Appendix C)). Also, we identified another 644 genes that were only in the biologically significant list for the vaccinated subjects (SDF19 of online supplementary data files (Appendix C)). Enriched Reactome pathway analysis on these genes resulted in pathways such as Interferon Signaling (24 genes), Antigen processing-Cross presentation (14 genes), ER-Phagosome pathway (12 genes), Binding and Uptake of Ligands by Scavenger Receptors (8 genes) and Class I MHC mediated

antigen processing & presentation (30 gene) (SDF22 of online supplementary data files (Appendix C)).

We also explored the 4,889 genes to identify how many genes were different in gene expression when looking at influenza infected subjects compared to influenza vaccinated subjects. Of the 4,889 genes, 4,261 genes showed statistically significant differences between vaccination and infection with influenza (Figure 5.3 and SDF25 - SDF27 of online supplementary data files (Appendix C)).

KEGG ID	KEGG Pathway	Gene Count	<i>p-value</i>	<i>adjusted p-value</i>
path:hsa04060	Cytokine-cytokine receptor interaction	34	5.5E-09	1.4E-06
path:hsa04660	T cell receptor signaling pathway	19	6.9E-08	8.7E-06
path:hsa04650	Natural killer cell mediated cytotoxicity	19	3.8E-06	3.2E-04
path:hsa04672	Intestinal immune network for IgA production	11	5.9E-06	3.8E-04
path:hsa04640	Hematopoietic cell lineage	14	1.8E-05	9.1E-04
path:hsa05340	Primary immunodeficiency	8	1.3E-04	4.8E-03
path:hsa04064	NF-kappa B signaling pathway	13	1.4E-04	4.8E-03
path:hsa04622	RIG-I-like receptor signaling pathway	11	1.6E-04	4.8E-03
path:hsa04068	FoxO signaling pathway	16	1.7E-04	4.8E-03
path:hsa05166	HTLV-I infection	23	6.4E-04	1.5E-02
path:hsa05162	Measles	15	6.4E-04	1.5E-02
path:hsa04062	Chemokine signaling pathway	18	9.8E-04	2.1E-02
path:hsa05330	Allograft rejection	7	1.1E-03	2.2E-02
path:hsa04380	Osteoclast differentiation	14	1.4E-03	2.6E-02
path:hsa05320	Autoimmune thyroid disease	8	1.9E-03	3.2E-02
path:hsa04110	Cell cycle	13	2.3E-03	3.6E-02
path:hsa04010	MAPK signaling pathway	21	2.8E-03	4.1E-02
path:hsa04630	Jak-STAT signaling pathway	15	2.9E-03	4.1E-02
path:hsa05164	Influenza A	16	3.3E-03	4.4E-02

Table 5.2: Enriched KEGG Pathways from Statistically Significant Genes with an Interaction Between Disease Status and Age.

Influenza Infection			
Down-Regulated		Up-Regulated	
Gene	Difference of Means	Gene	Difference of Means
NELL2	-1.687	UGCG	2.005
UBASH3A	-1.583	CD177	1.875
ABCB1	-1.513	OTOF	1.844
PID1	-1.457	HP	1.625
CACNA2D3	-1.428	SSH1	1.491
PTGDR	-1.423	DTL	1.431
CD40LG	-1.392	GPR84	1.428
PTGDR2	-1.390	HJURP	1.420
TLE2	-1.379	CDC45	1.395
NCR3	-1.357	SLC1A3	1.390
Influenza Vaccination			
Down-Regulated		Up-Regulated	
Gene	Difference of Means	Gene	Difference of Means
TOP1MT	-0.179	GBP1	0.354
ARNTL	-0.176	MYOF	0.347
DIDO1	-0.172	STAT1	0.284
PDE4D	-0.169	PSTPIP2	0.281
TMX4	-0.168	SAMD9L	0.276
ZNF589	-0.166	OAS3	0.269
SLC37A3	-0.165	WARS	0.263
GNB5	-0.165	BATF2	0.263
ENO2	-0.162	ANKRD22	0.256
AP3M2	-0.159	C1QB	0.255

Table 5.3: Top 10 Up- and Down- Regulated Differentially Expressed Genes from the Influenza Infected and Influenza Vaccination Biologically Significant Gene Lists (based on estimates).

5.4.2 Age and Sex Effect on Gene Expression in Influenza

Using the 4,889 genes disease significant genes from above, we Bonferroni-adjusted the p-values for both the age and sex factors. We then further filtered the 4,889 list by the age factor p-values (Bonferroni-adjusted p-value < 0.05) to identify statistically significant interacting genes between disease state and age (DxA). We also repeated this approach for the sex factor interaction with disease (DxS). Of the 4,889 statistically significant (Bonferroni-adjusted p-value < 0.05) disease genes, 907 of them had a statistically significant interaction with disease and age (SDF28 of online supplementary data files (Appendix C)). KEGG enrichment, our results include: Cytokine-cytokine receptor interaction (34 genes), T cell receptor signaling pathway (19 genes), Natural killer cell mediated cytotoxicity (19 genes), Intestinal immune network for IgA production (11 gene hits), Hematopoietic cell lineage (14 genes), Primary immunodeficiency (8 genes), NF-kappa B signaling pathway (13 genes), and Influenza A (16 genes) (SDF30 of online supplementary data files (Appendix C), Table 5.2). We also looked at the biologically significant gene lists for influenza infection and vaccination (based on effect as discussed above) to determine which of these genes also had a significant interaction with disease and age. Of the 978 in the influenza infection biologically significant list, 432 had a statistically significant ((Bonferroni-adjusted p-value < 0.05 for disease and age factor) interaction with disease and age (Figure 5.3 and SDF32 of online supplementary data files (Appendix C)). In the biologically significant gene list for influenza vaccinated subjects 335 genes also had a statistically significant (Bonferroni-adjusted p-value < 0.05 for disease and age factor) interaction with disease and age (Figure 5.3 and SDF35 of online supplementary data files (Appendix C)).

Furthermore, we explored differences in gene expression (based on mean differences across groups) in subjects with influenza infection, influenza vaccination and controls across the 4 age groups: (-1,3],(3,19], (19,65], (65,100] using the gene lists of identified disease:age interacting genes. First we calculated the mean expression for control subjects younger than 3 (age group: (-1,3]). This served as our baseline for all comparisons to influenza infection and vaccination. We calculated the difference in means for the subjects within the other age groups only focusing on the

healthy subjects and used the younger than 3 as our baseline to find the difference in means (Figure 5.8). We also calculated the difference in mean expression for all influenza infected subjects and used the influenza infected subjects younger than 3 as the baseline for comparisons of relative expression (Figure 5.9A). In addition to this, we calculated difference in means by comparing influenza infected samples to the control baseline (younger than 3) (Figure 5.9B). We repeated the above steps with our vaccinated subjects to explore how expression changes with age and disease state (Figure 5.10). We also plotted the difference in means comparing influenza vaccinated subjects to influenza infected subjects to highlight temporal patterns of the 907 interacting (disease:age) genes (Figure 5.11).

We also filtered our gene lists (statistically significant disease genes and the biologically significant for influenza disease and vaccination gene lists) for genes with a statistically significant disease interaction with sex (Figure 5.3). We identified 48 of the 4,889 disease genes (Bonferroni-adjusted p -value < 0.05 for disease and sex factor) that interacted with disease and sex ((Figure 5.3) and SDF29 of online supplementary data files (Appendix C)). In the influenza infected biologically significant gene list there were 17 genes that interacted with disease and sex (Bonferroni-adjusted p -value < 0.05 for disease and sex factor), and 7 genes had an interaction with disease, sex and age (Bonferroni-adjusted p -value < 0.05 for disease, sex and age factor) (Figure 5.3 and see also SDF33 and SDF34 of online supplementary data files (Appendix C)). We did not find any statistically significant enrichment in pathways for these genes. As for the biologically significant influenza vaccination genes, 37 of them were associated with disease and sex interactions (Bonferroni-adjusted p -value < 0.05 for disease and sex factor), and 13 genes had associated interactions with disease, sex and age (Bonferroni-adjusted p -value < 0.05 for disease, sex and age factor) (Figure 5.3 and see also SDF36 and SDF37 of online supplementary data files (Appendix C)). We also did not find any enriched pathways for these genes.

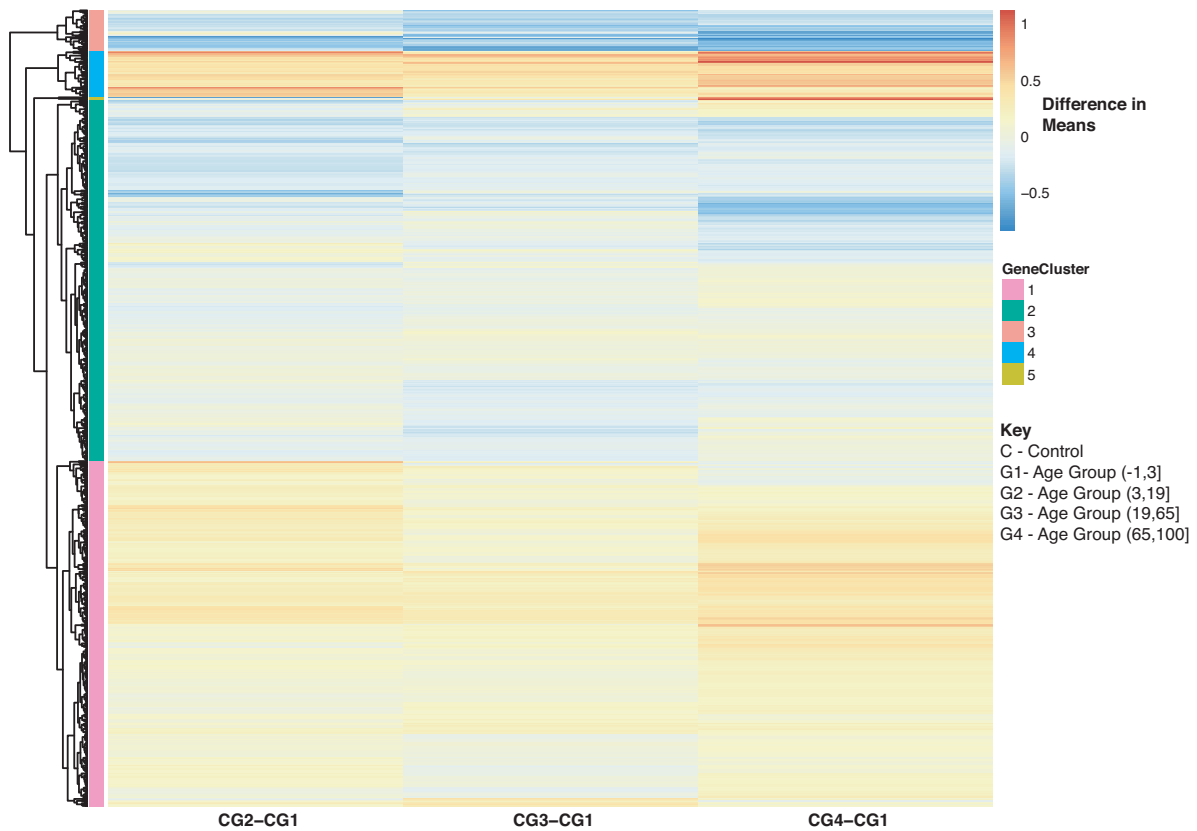


Figure 5.8: Heatmap of Statistically Significant (Bonferroni-adjusted p-value < 0.05) Genes with an Interaction Between Disease State and Age for Healthy Controls. Difference in means calculated by comparing control subjects in age groups 2-4 to control subjects in age group 1 (baseline).

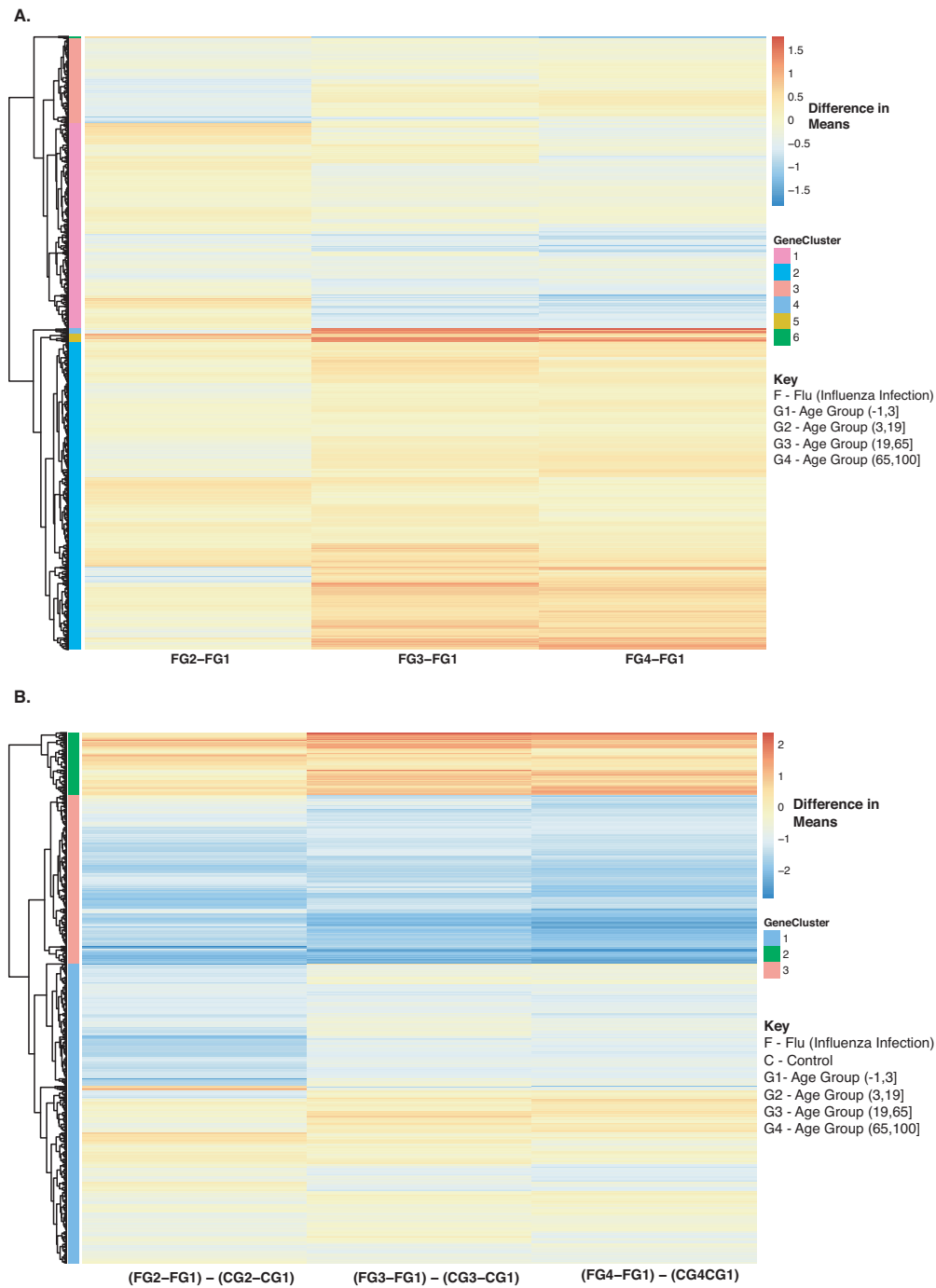


Figure 5.9: Heatmap of Statistically Significant (Bonferroni-adjusted p -value <0.05) Genes with an Interaction Between Disease State and Age for Influenza Infected Subjects. (A) Difference in means calculated by comparing influenza infected subjects in age groups 2-4 to influenza infected subjects in age group 1 (baseline). (B) Comparison of influenza infected subjects to control subjects in the different age groups by calculating the difference between the baseline-adjusted means for influenza infected subjects (A) and control subjects (Figure 5.8).

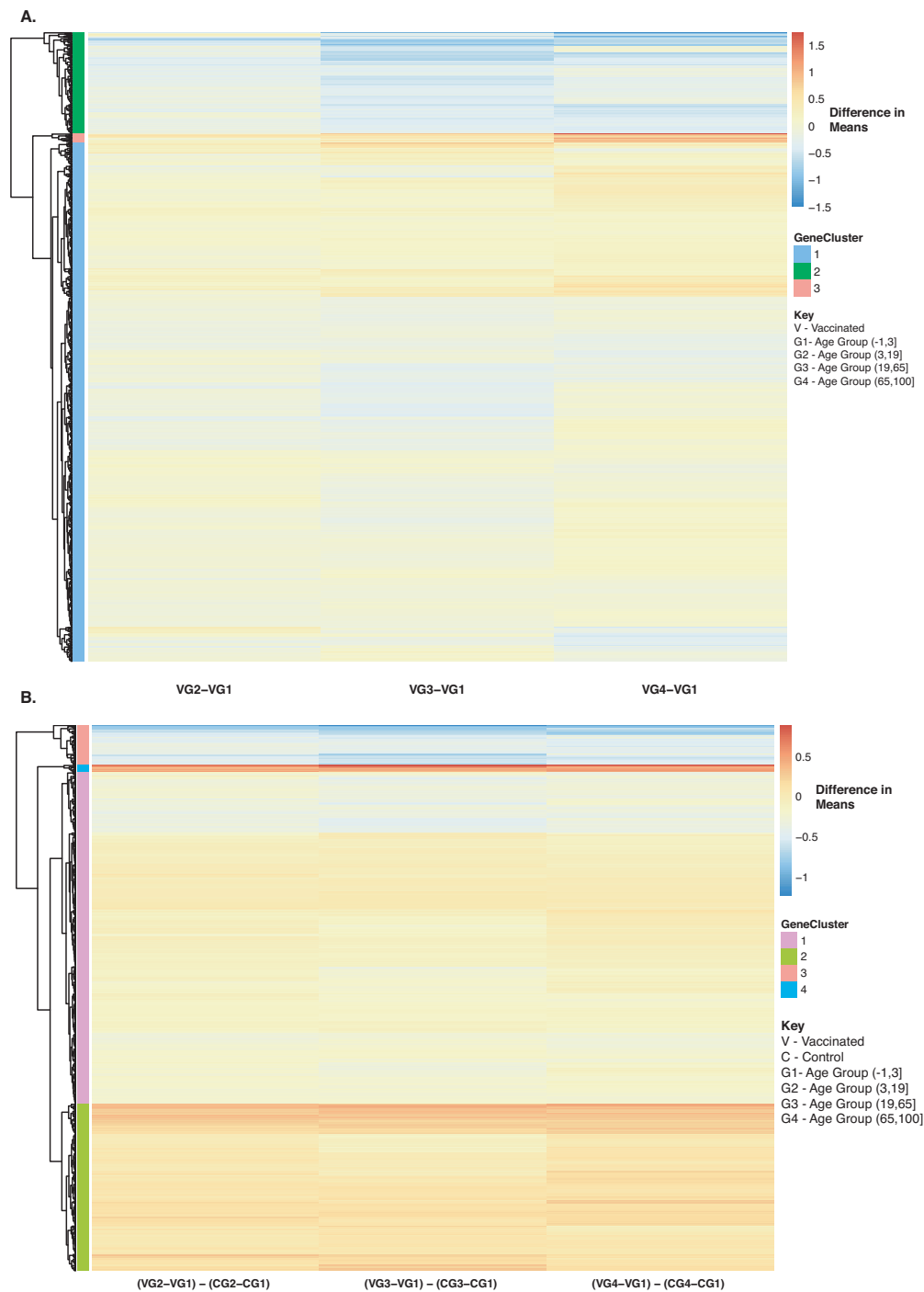


Figure 5.10: Heatmap of Statistically Significant (Bonferroni-adjusted p -value <0.05) Genes with an Interaction Between Disease State and Age for Influenza Vaccinated Subjects. (A) Difference in means calculated by comparing influenza vaccinated subjects in age groups 2-4 to influenza vaccinated subjects in age group 1 (baseline). (B) Comparison of influenza vaccinated subjects to control subjects in the different age groups by calculating the difference between the baseline-adjusted means for influenza vaccinated subjects (A) and control subjects (Figure 5.8).

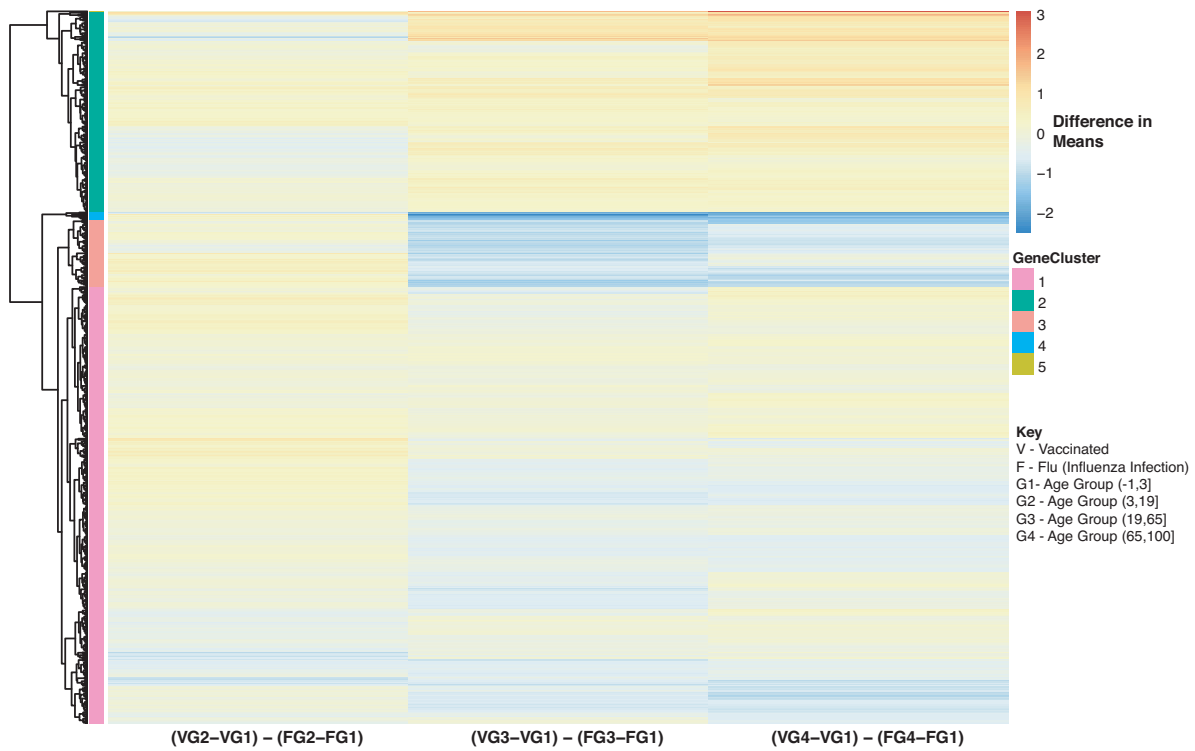


Figure 5.11: Heatmap of Statistically Significant (Bonferroni-adjusted p-value < 0.05) Genes with an Interaction Between Disease State and Age for Influenza Vaccinated Subjects Compared to Influenza Infected Subjects. Comparison of baseline-adjusted means for influenza vaccinated subjects (Figure 5.10A) and influenza infected subjects (Figure 5.9A)

5.5 Discussion

Every year there is a new vaccine available to reduce the amount of influenza cases worldwide. The influenza virus is constantly changing and researchers have to predict the most common strains that will affect the population each season. During the flu season, the majority of hospitalizations and deaths from influenza are within the elderly population [9]. Young children are also at high risk for severe infections of influenza due to their underdeveloped immune system [10]. Current vaccine development methods, though effective are also flawed. In some cases, the influenza strains can mutate after the strains for the vaccine have been selected for the upcoming flu season, which then reduces the effectiveness of the vaccine [595]. Exploring how gene expression varies in influenza infection, vaccination, and comparison of the differences may highlight prospective biomarkers/-gene signatures for improving vaccinations. In addition, because of the observed age-dependency

in influenza infection, investigating gene expression temporal patterns across various ages can also provide insight on how genes change due to underdevelopment and immunosenescence.

We identified 18 microarray expression datasets that passed our inclusion criteria for a meta-analysis on influenza (Table 5.1). We collected the raw expression microarray data for all datasets, pre-processed them and combined by common gene names. With 3,481 samples (including repeated measures) we modeled the pre-processed expression data with a mixed effects model and carried out LRT analysis. Our LRT analyses resulted in 4,889 statistically significant (Bonferroni-adjusted p-value <0.05) disease genes (see SDF5 of online supplementary data files (Appendix C)). These results include CD177 which plays a role in innate immune response by regulating chemotaxis of neutrophils [92, 596], BCL11B which regulates T-cell differentiation [92, 597], HMGB1 protein has been shown to promote viral replication [598] and plays a role in inflammation [92], TPP2 plays a role in major histocompatibility complex (MHC) presentation and TANK is involved in NF-kappa B signalling.

We highlighted the KEGG NF-kappa B signaling pathway using the estimates from influenza infection and vaccination (Figure 5.4) and Figure 5.5). The NF-kappa B pathway is activated during influenza infection which up-regulates antiviral genes [599] and can regulate viral synthesis [600]. Previous studies have also reported that the influenza virus is capable of regulating antiviral activity by NF-kappa B and promote replication in hosts [600]. In the NF-kappa B pathway, we observed similar expression patterns between disease and vaccinated subjects, including down regulation of genes involved in MCH/Antigen presentation for both physiological states. There are also some differences in gene expression observed such as CD40 and PARP1 up-regulated in vaccinated samples. CD40 has previously been shown to regulate immune response and promotes protection against the virus [601, 602] while PARP1 has been highlighted as a host factor that can regulate the polymerase activity of influenza [603]. In Figure 5.5, the genes in our vaccine list in the RIG-I-like receptor signaling pathway are down-regulated, compared to influenza infected subjects (Figure 5.4). The RIG-I-like receptors have been previously shown to be involved in sensing viral RNA and regulating an antiviral immune response [604]. Other genes such as ICAM which is

involved in lymphocyte adhesion and T-cell costimulation, and BLC and ELC involved in lymphoid tissue honing are all down-regulated in vaccinated subjects compared to infected subjects (Figure 5.4) and Figure 5.5). We also highlighted expression of genes in the Influenza A KEGG pathway for influenza infected and influenza vaccinated (Figure 5.6) and Figure 5.7). Although there are similarities in both figures, some key differences in expression are observed in genes connected with high fever (IL-1 and IL-6). Studies have shown elevated levels of IL-1 β and IL-6 following infection with influenza A [580, 605].

Additionally, we compared our biologically significant gene list for influenza infection (978 genes) to Dunning et al., who identified whole blood RNA signatures in hospitalized adults with influenza. Their findings indicated that genes involved in interferon-related pathways were activated at the start of the infection and by day 4 had started to decrease with a shift in inflammatory and neutrophil related pathways [580]. Our findings also indicate enrichment for neutrophil related pathways in the case of influenza infection. Dunning et al. list a top 25 gene set (controls versus influenza subjects), from which 22 genes overlap with our findings (978 genes, see SDF9 of online supplementary data files (Appendix C)). 5 of our top 10 up-regulated gene list overlap with the Dunning et al. 25-gene set (Table 5.3), namely UGCG, CD177, OTOF, HP and SSH1.

Furthermore, our identified biologically significant gene lists for influenza infection and vaccination (using a 2-tailed 10% quantile filter on expression estimates of effect size compared to healthy control) have 334 genes in common, with 644 genes being unique to influenza infection and 644 being unique to influenza vaccination (SDF17-SDF19 of online supplementary data files (Appendix C)). Following pathway enrichment, we observed that the genes that are unique to each disease state (influenza infected and vaccinated) are involved in different processes. For example, the biologically significant genes only in influenza infected samples were enriched in pathways such as neutrophil degranulation and cell cycle checkpoints (SDF21 of online supplementary data files (Appendix C)). Neutrophil degranulation is a defensive process neutrophils undergo to protect the host against invading pathogens. On the other hand, pathways involved in interferon signaling and antigen processing were enriched for the genes only in the vaccinated gene list (SDF22 of

online supplementary data files (Appendix C)). This indicates that with the actual infection the body undergoes different processes to that induced by vaccination (Figures C.3-C.4 of Appendix C and see SDF23 and SDF24 of online supplementary data files (Appendix C)).

The 48 genes for which we identified a statistically significant interaction between disease and sex are highlighted in SDF29 of the online supplementary data files (Appendix C). Sex-specific gene expression has been previously observed in influenza. Studies have observed females exhibited a stronger immune response to influenza vaccine compared to males within the first day [606]. Another study suggested that males have a stronger immune response to influenza infection [607]. These findings indicate sex and influenza is still to be explored and our gene list may offer new candidates to be investigated for their role in influenza.

With regards to aging and influenza, the statistics of the disease burden indicates specific age groups are at higher risk for infection [9]. This is in part due to immune system development and deterioration. For example, B and T cell function diminishes with age [10, 608]. In our analysis, we identified 907 disease-associated genes with a statistically significant interaction with age that were also enriched in immune related KEGG pathways (SDF30 of online supplementary data files (Appendix C)). Figure 5.8 compares the mean differences of healthy subjects to the baseline (healthy children younger than 3). There are 4 major groups (Figure 5.8 and see SDF47 of online supplementary data files (Appendix C)): with reference to Figure 5.8, genes in Cluster 1 were up-regulated compared to the baseline for all age comparisons, Cluster 2 and 3 genes were generally down-regulated compared to the baseline, and Cluster 4 genes are up-regulated and increase with age. Genes in Cluster 1 and 2 are involved in Reactome pathways such as cytokine signaling, interferon signaling and the immune system. Cluster 3 genes are involved in Reactome pathways such as interferon signaling and cell cycle while Cluster 4 genes are involved in cellular senescence, signaling by interleukins and immune system.

In Figure 5.9 we further explored changes in gene expression across age groups due to influenza infection of our 907 disease:age interacting genes. Figure 5.9A is compares influenza infected subjects in age groups 2,3 and 4 to the baseline (infection subjects under 3). In Figure 5.9A

there are three major groups (cluster numbering with respect to Figure 5.9A): Cluster 1 (gradual decrease with age), 2 (genes up-regulated with increase in age), and 3 (gradual down-regulation with age). Genes in Cluster 1 are in Reactome pathways such as cytokine signaling, interferon signaling, antiviral mechanism by IFN-stimulated genes and chemokine receptors. Cluster 2 genes are involved in regulatory T lymphocytes, transcription, protein repair, and interleukin-2 signaling while Cluster 3 genes are involved in gene transcription (see SDF48 of online supplementary data files (Appendix C)). Figure 5.9B instead compares influenza infected subjects to controls by looking at difference in means. There are three groups of expression patterns (cluster numbering with respect to Figure 5.9B): Cluster 1 shows a gradual increase with age, in Cluster 2 expression intensifies with age and in Cluster 3 genes are down-regulated compared to the control subjects younger than 3 (see SDF49 of online supplementary data files (Appendix C)). Genes in Clusters 1 and 2 were not in any enriched Reactome pathways but are associated with transcription and signaling pathways. Genes in Cluster 3 were in Reactome pathways that include cytokine signaling, interferon signaling, antiviral mechanism by IFN-stimulated genes and chemokine receptors.

As for the vaccinated subjects with respect to Figure 5.10A, we observe a gradual decrease in gene expression for gene Cluster 2 and a gradual increase in expression for genes in Cluster 1 and 3 compared to the baseline (young vaccinated subjects under age 3) (SDF50 of online supplementary data files (Appendix C)). Genes in Cluster 1 were not enriched in pathways while genes in Cluster 2 were enriched in Reactome pathways that include interferon and cytokine signaling, antiviral mechanism and response. Genes in Cluster 3 were enriched in Reactome pathways that include interferon and cytokine signaling, cellular senescence and immune system. When we compared vaccinated subjects to control subjects across ages we observed 3 main trends, with respect to Figure 5.10B: Cluster 1 (pathways include antiviral mechanisms, interferon and cytokine signaling) 2 (pathways such as immune response and cell migration, immunological synapse and chemokine receptors) and 4 (mitochondrial translation) genes are all up-regulated in vaccinated subjects with Cluster 3 (pathways include interferon and cytokine signaling and immune system) genes being down-regulated (Figure 5.10B and SDF51 of online supplementary data files (Appendix C)).

We also compared influenza vaccinated subjects to influenza infected subjects to explore changes in expression with age. There are 3 major groups (cluster numbering with respect to Figure: Cluster 1 and 3 genes show a gradual decrease in expression with age in vaccinated subjects compared to influenza infected subjects. Genes in Cluster 2 show a gradual increase in expression with age in influenza vaccinated subjects. Figures C.6-C.8 of Appendix C also explore temporal patterns with age.

Our heatmaps display temporal patterns with age in response to influenza infection and vaccination. These genes that are associated with disease and age interactions, are all involved in immune-related pathways. Exploring how gene expression changes with age in immune related genes can help further characterize the disease and improve treatments. For example, age, pre-existing health conditions and influenza history (previous infection or vaccination) are all factors that can affect the efficacy of the vaccine [609]. There is an on-going effort to improve efficacy of vaccination in the elderly population. Studies have suggested that antibody titers decline drastically in older adults from seroconversion to day 180 after vaccination [609]. The decay of antibody titers also highlight the importance of determining the right time and how many times one should be vaccinated. Vaccines for the elderly population have been modified to increase the dosage and use adjuvants to increase immunogenicity [610]. and Ramsay et al., also showed that vaccination during the current influenza season provides stronger protection than vaccinations from previous seasons [611].

The vaccine type also plays a role in immunogenicity within hosts. For example, Nakaya et al., were able to detect larger antibody titers and plasmablasts generated in the trivalent inactivated vaccine (TIV) compared to the live attenuated vaccine (LAIV), and differentially expressed genes mostly related to interferon signaling [582]. LAIV responses in young children are higher than in adults. For instance, LAIV when compared to inactivated vaccines induced smaller concentrations of antibodies in response to HA in adults [612]. Previous findings have shown the benefit of taking a systems biology approach to assess gene expression responses to vaccinations [587, 609, 613]. Our findings not only identify genes that are different between controls compared to infected and

vaccinated subjects, but with our methodology we were also able to assess differences between the influenza infected and vaccinated subjects while still investigating disease genes that interact with age and sex. Our temporal patterns with age for each disease state helps to clarify how age might be playing a role.

As we have previously observed [526], meta-analyses using microarray expression data have multiple limitations: Our findings are limited to only genes that have been annotated and are existing probes on the arrays, and also have to be consistently utilized across array platforms. Hence, we are unable to probe global gene expression, and are limited to mRNA profiling. These can be expanded in future studies using RNA-sequencing data, and the newer single-cell sequencing approaches that would allow cell-specific information to be discerned, which is important in evaluating immune responses and the interplay between various cell types. Taking a similar approach to our microarray dataset analysis using RNA-sequencing data will promote the discovery of novel genes by being able to explore the entire transcriptome. Additionally, we are limited by the varying annotations of the available public datasets, and can only explore characteristics that are uniformly reported. For example, we did not have virus strain information for all samples or vaccine details so we were unable to include such info in our analysis. In addition to this, our study is unbalanced (particularly with respect to disease state, where a limited number of influenza infection samples were available: 3,481 samples (1,277 controls, 297 influenza infection, 1,907 influenza vaccinated, 1,537 males and 1,944 female). We additionally used repeated measures, which we accounted for in our mixed effects model.

Despite the limitations introduced by using micro-array data, our study identified gene candidates by factor (disease status, age and sex) that can be examined further to understand their role in influenza infection and vaccination. We also highlighted 907 genes that have an age-effect on gene expression. These genes can be further explored to determine their role in influenza infection and how they can be further analyzed for their role in implementing effective universal vaccines regardless of age. All these consideration are of paramount importance in designing the next generation of vaccines, as we move forward towards a universal influenza vaccine.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

6.1 Conclusion

In this dissertation we examined neurodegenerative and respiratory diseases to explore age-dependent changes in gene expression. We were also especially interested in identifying the effects of other factors such as sex and tissue on gene expression. We accomplished this by using publicly available microarray gene expression data from public database repositories to explore gene expression variation in diseases which have previously been shown to have an age-dependency.

Chapter 2 explored the effects of age, tissue and sex on gene expression in Alzheimer's Disease. We identified temporal patterns for control and AD samples (2,088 samples). We found 3,735 statistically significant genes of which 352 of them were biologically significant. Of the 352 disease genes we found genes that were directly interacting with sex and age. We were also able to highlight aging trends across genes and explore differences in gene expression due to different brain regions. We identified novel genes and associated pathways that can be further studied for their role in AD.

In Chapter 3, we conducted a literature review to explore how *Streptococcus pneumoniae* causes disease and how the host protects itself against the pathogen. We discuss how the state of one's immune system plays a big role in susceptibility to the disease. Understanding how the bacteria causes the diseases and how weakened host defenses promote infection and spreading throughout the body sheds light on how important a strong immune system is. Pneumococcal infections are more prevalent in young children with under-developed immune systems and the elderly and immunocompromised who have a weakened immune system. We explore the treatments available and analyze their efficacy in different age groups. Overall, pneumococcal disease burden has been reduced due to the available vaccines however efficacy varies with age.

Chapter 4 studied how age, sex and smoking status affected gene expression profiles in COPD compared to healthy controls (1,262 samples). Our linear model and ANOVA resulted in 1,513 statistically significant differentially expressed disease genes. Filtering for biological significance we identified 304 genes. These genes were associated with immune and cell processes. We also highlighted genes that significantly interacted with smoking status. Our results compared

smokers to non-smokers with and without COPD and the temporal patterns highlighted the effects of smoking and how it can promote the pathology of the disease. We also successfully trained a model via machine learning to predict disease status (COPD or control).

In the final chapter, Chapter 5, we assessed expression profiles in controls, influenza infected samples and influenza vaccinated samples (3,481 samples). In addition to comparing these disease states, we also highlighted important genes with interactions between disease and age group and disease and sex. This meta-analysis resulted in 4,889 statistically significant genes. Of these genes we looked for biological significance in influenza infected samples (978 genes) and influenza vaccinated samples (978 genes). We were able to identify pathways and genes that both disease states (influenza infected and vaccinated) had in common and genes that were unique to each disease state. Our results also highlight temporal patterns that clearly show aging trends in response to infection and vaccination compared to controls.

6.2 Limitations and Future Directions

The use of microarray expression data for our meta-analyses introduced limitations to our findings. We were limited to previously annotated genes which prevented us from identifying novel gene signatures due to not being able to span the entire transcriptome. Using publicly available data also caused our study design to be dependent on the annotation of researchers. We were only able to explore factors that were uniformly reported. For example, COPD is more prevalent in low-middle class communities, but because socioeconomic background information was not reported for samples, we could not explore this factor in our analyses. Additionally, in the case of AD, there is a higher prevalence of AD in African-Americans, but we were not able to explore the effect of race or ethnicity on gene expression for our meta-analysis because this information was not reported for all samples.

Despite the limitations introduced by using pre-existing microarray data, we were successful in identifying statistically significant gene lists by factor of differentially expressed genes for Chapters 2,4 and 5. We yielded gene candidates that can be further investigated for their role in the

diseases. We were also successful in identifying disease genes that were affected by age, tissue, sex and smoking status (COPD). To build on the findings in my dissertation, curating datasets that enable meta-analyses with balanced design in terms of disease states, sex and ethnicity can improve our interpretation of our findings and allow us to investigate variation in susceptibility and effectiveness of treatments. As meta-analyses continue to be a popular approach to explore diseases and have stronger power due to larger sample sizes, promoting uniformed data annotation/ sample characteristics reporting will allow for more study factors to be explored which will promote a thorough understanding of how gene expression profiles can be affected by one's characteristics or background. Furthermore, implementing RNA-sequencing data in meta-analyses of diseases rather than microarray expression data will promote the identification of novel genes and biomarkers for AD, COPD and Influenza. In addition to using RNA-sequencing expression data, other technologies such as proteomics and personalized 'omics can help with characterizing disease pathology and highlight additional biomarkers while also improving individual outcomes .

APPENDICES

APPENDIX A

DATA-DRIVEN ANALYSIS OF AGE, SEX, AND TISSUE EFFECTS ON GENE EXPRESSION VARIABILITY IN ALZHEIMER'S DISEASE SUPPLEMENTARY DATA

A.1 Online Supplementary Data Files

Our datasets, data files and results generated in our Alzheimer's Disease meta-analysis have been deposited to FigShare. The supplementary file names begin with the prefix “ST” and are referred to throughout the chapter. To access the FigShare online repository: <https://doi.org/10.6084/m9.figshare.7435469>

A.2 Supplementary Figures

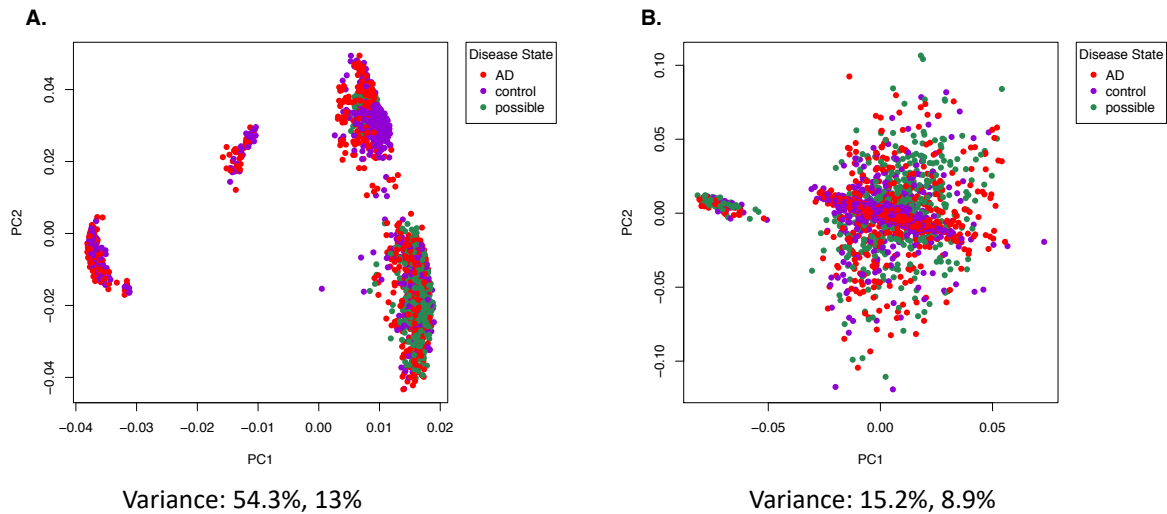


Figure A.1: Principal component analysis of the disease factor before (A) and after (B) batch correction with ComBat.

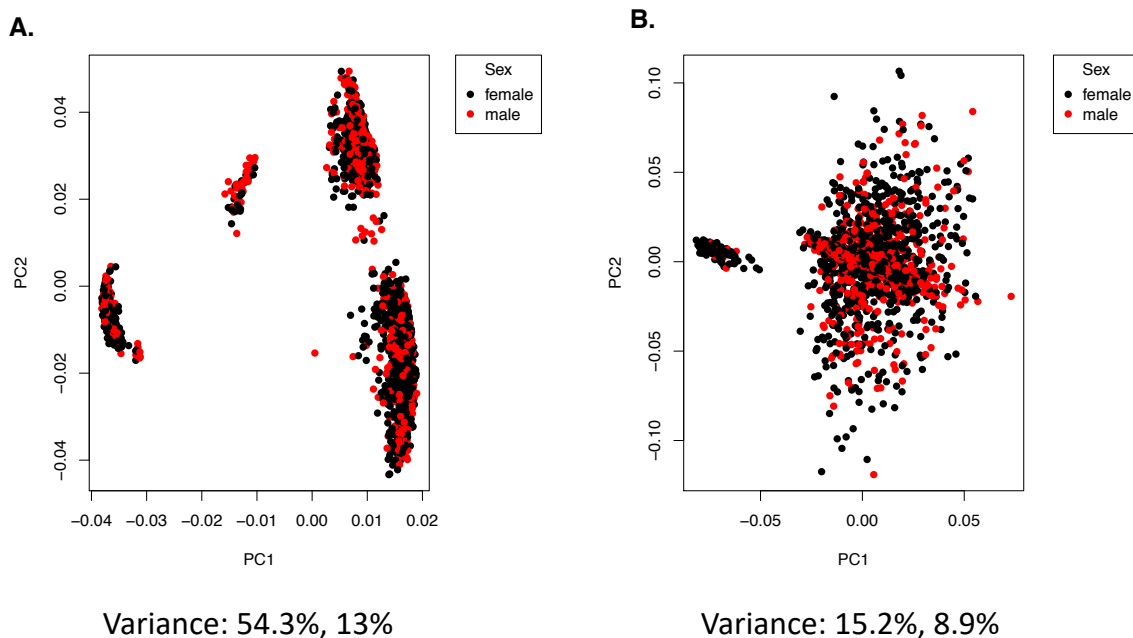


Figure A.2: Principal component analysis of the sex factor before (A) and after (B) batch effect correction with ComBat.

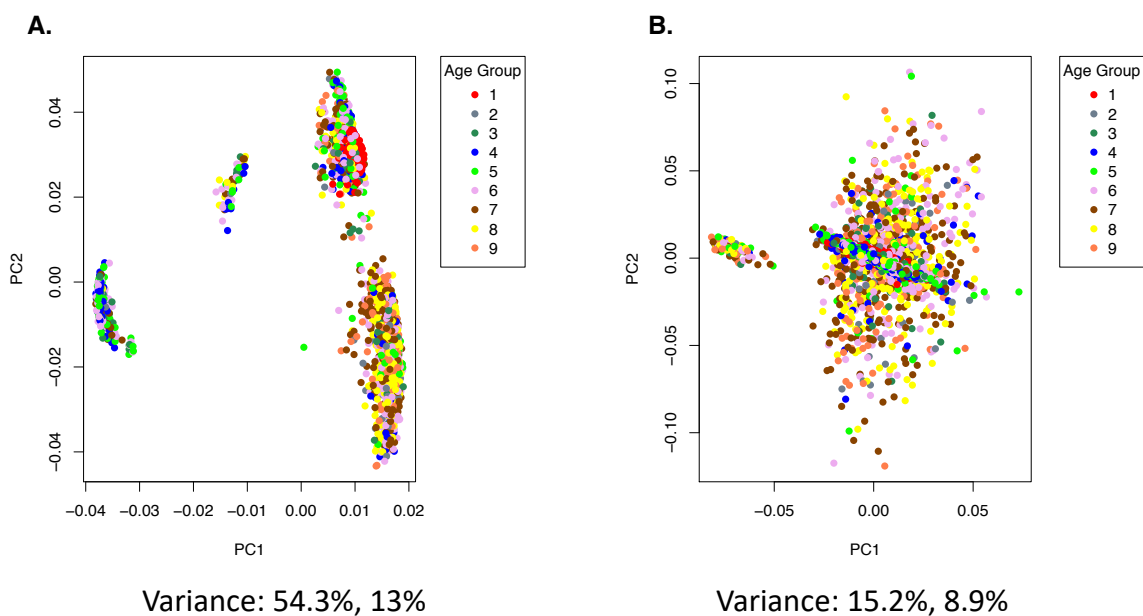


Figure A.3: Principal component analysis of the age group factor before (A) and after (B) batch effect correction with ComBat.

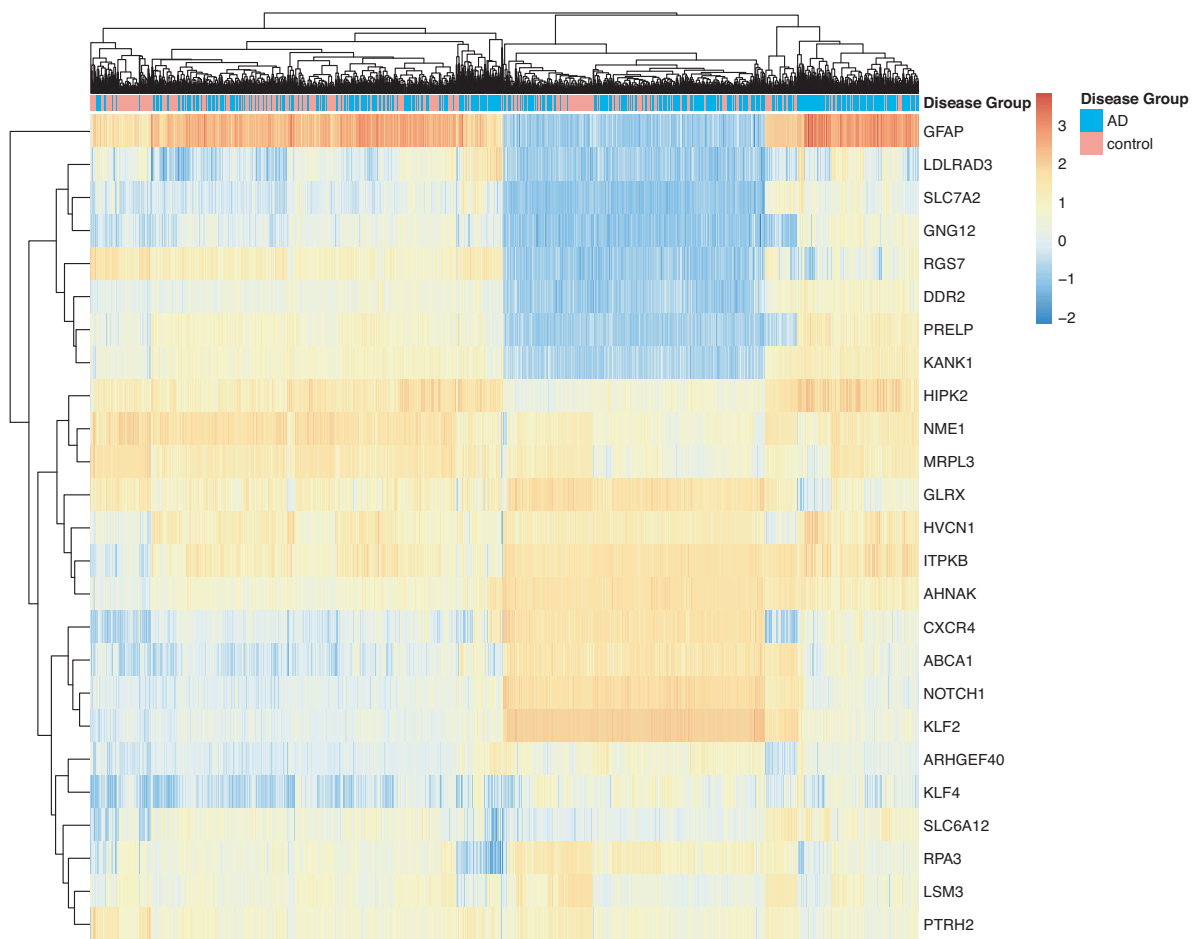


Figure A.4: Heatmap with gene clustering of the top 25 differentially expressed disease (control-AD) gene list.

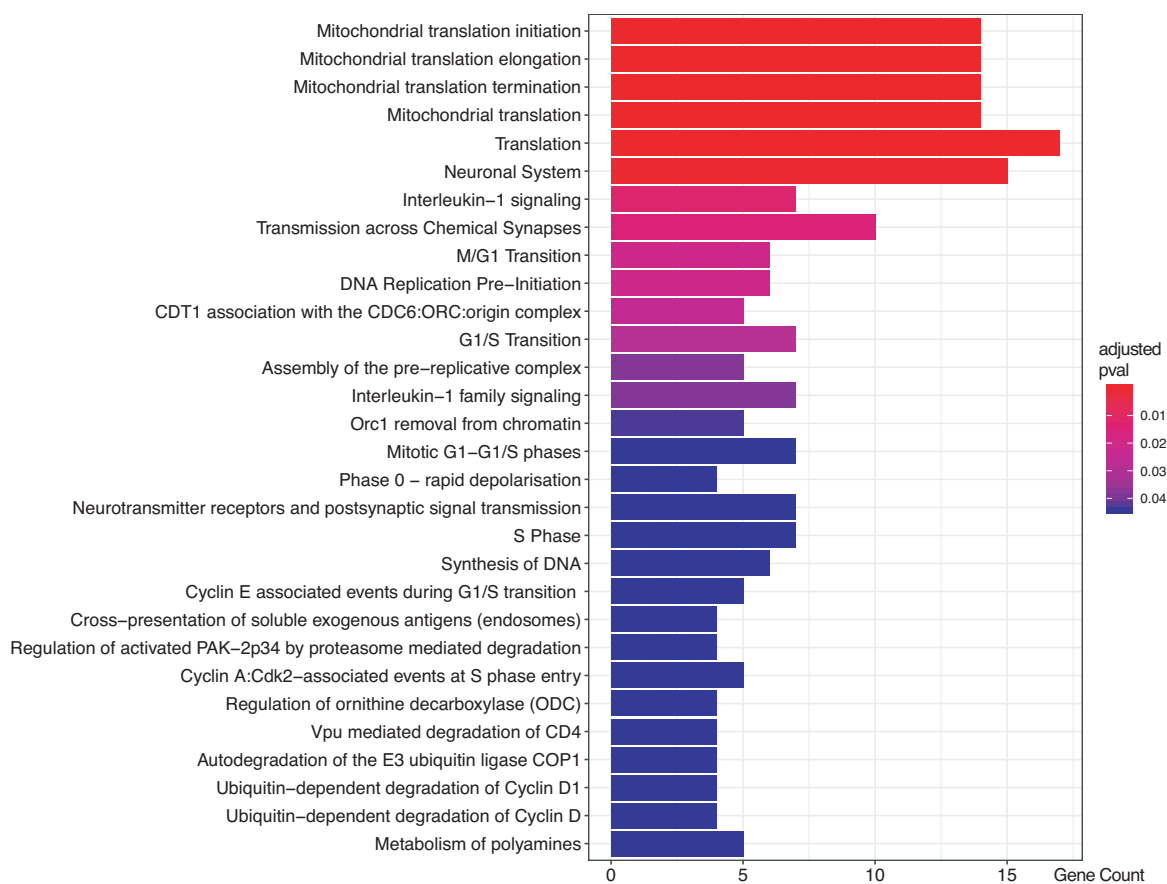


Figure A.5: Reactome pathway analysis bar plot of enriched pathways and number of gene hits. Gene list: Genes that were down-regulated in Alzheimer's disease but up-regulated in healthy controls.

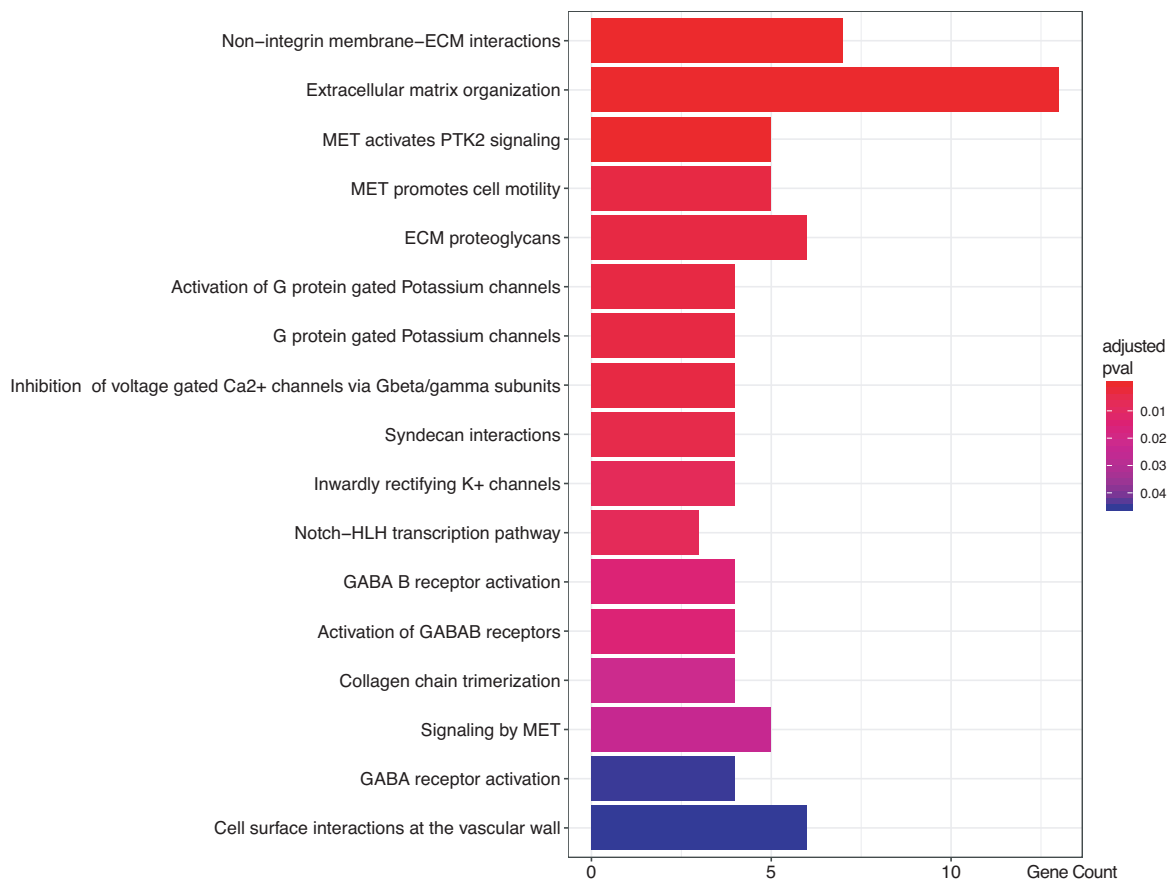
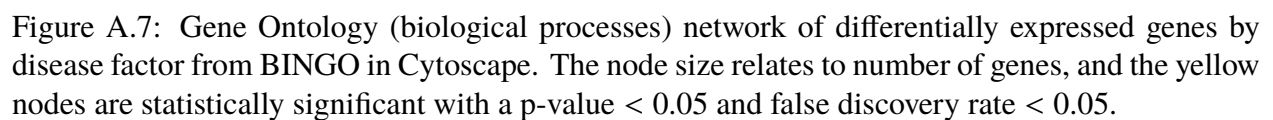


Figure A.6: Reactome pathway analysis bar plot of enriched pathways and number of gene hits. Gene list: Genes that were up-regulated in Alzheimer's disease but down-regulated in healthy controls.



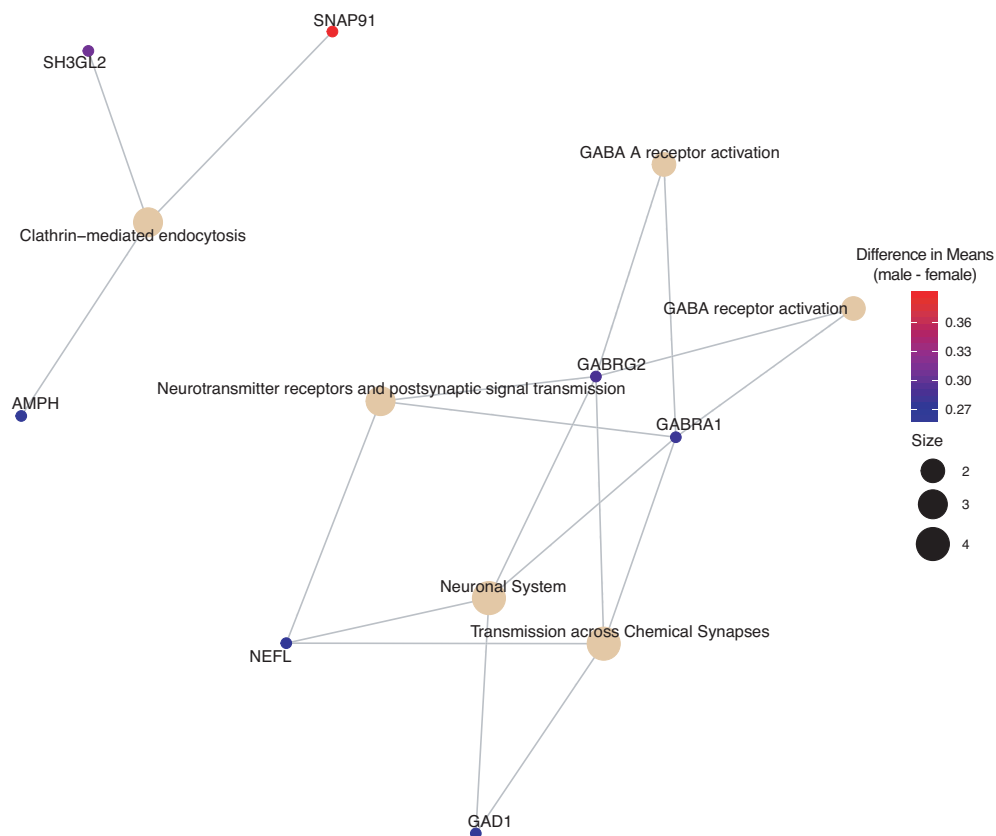


Figure A.8: Pathway-gene network of enriched Reactome pathways using the differentially expressed disease genes with a sex effect (prior to selecting for interacting genes) that were up-regulated in males



Figure A.9: Heatmap with gene clustering to visualize gene expression of differentially expressed disease (control-AD) gene list with a sex effect (prior to selecting for interacting genes).

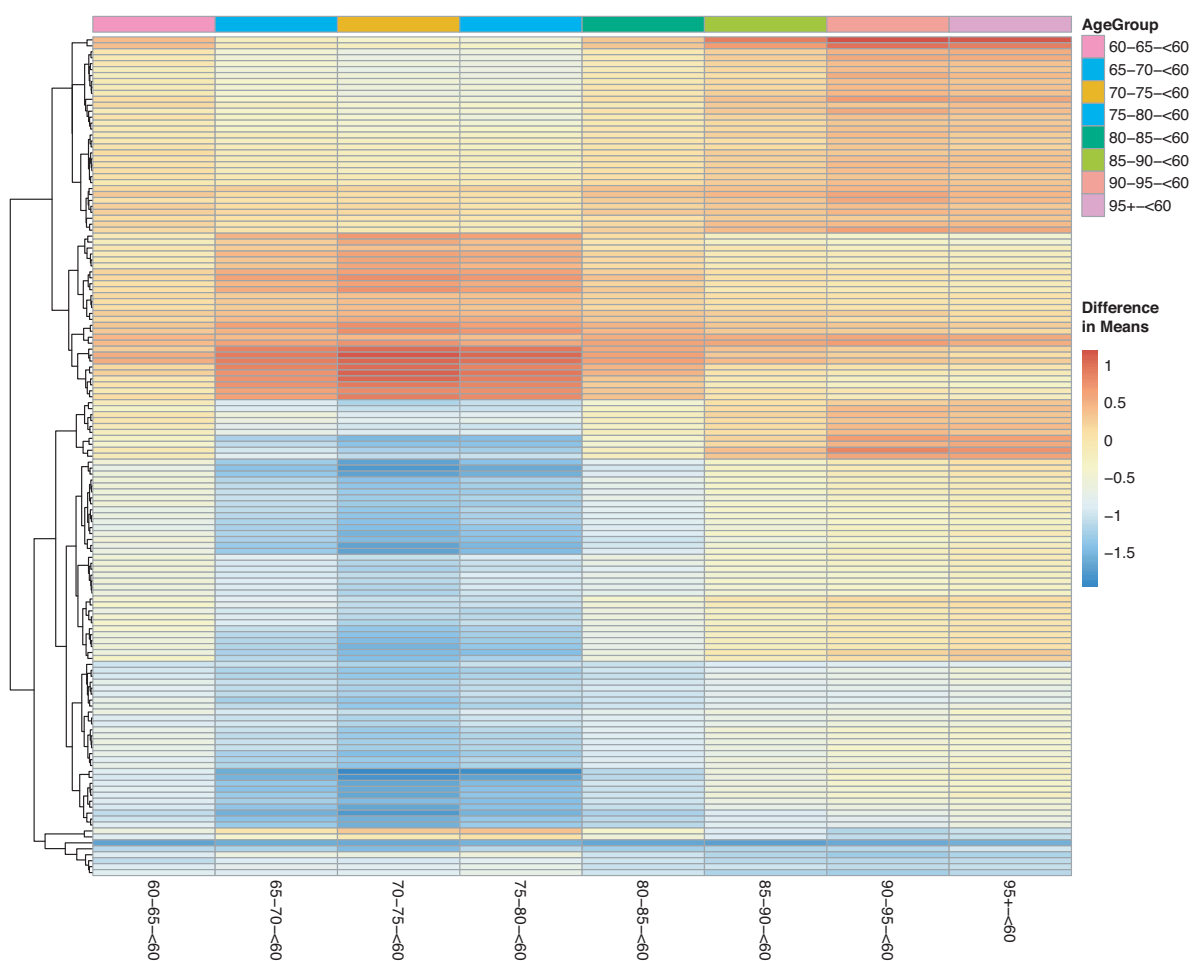


Figure A.10: Heatmap with gene clustering to visualize age group effect (prior to selecting for interacting genes) using difference in means on the differentially expressed disease (control-AD) gene list.

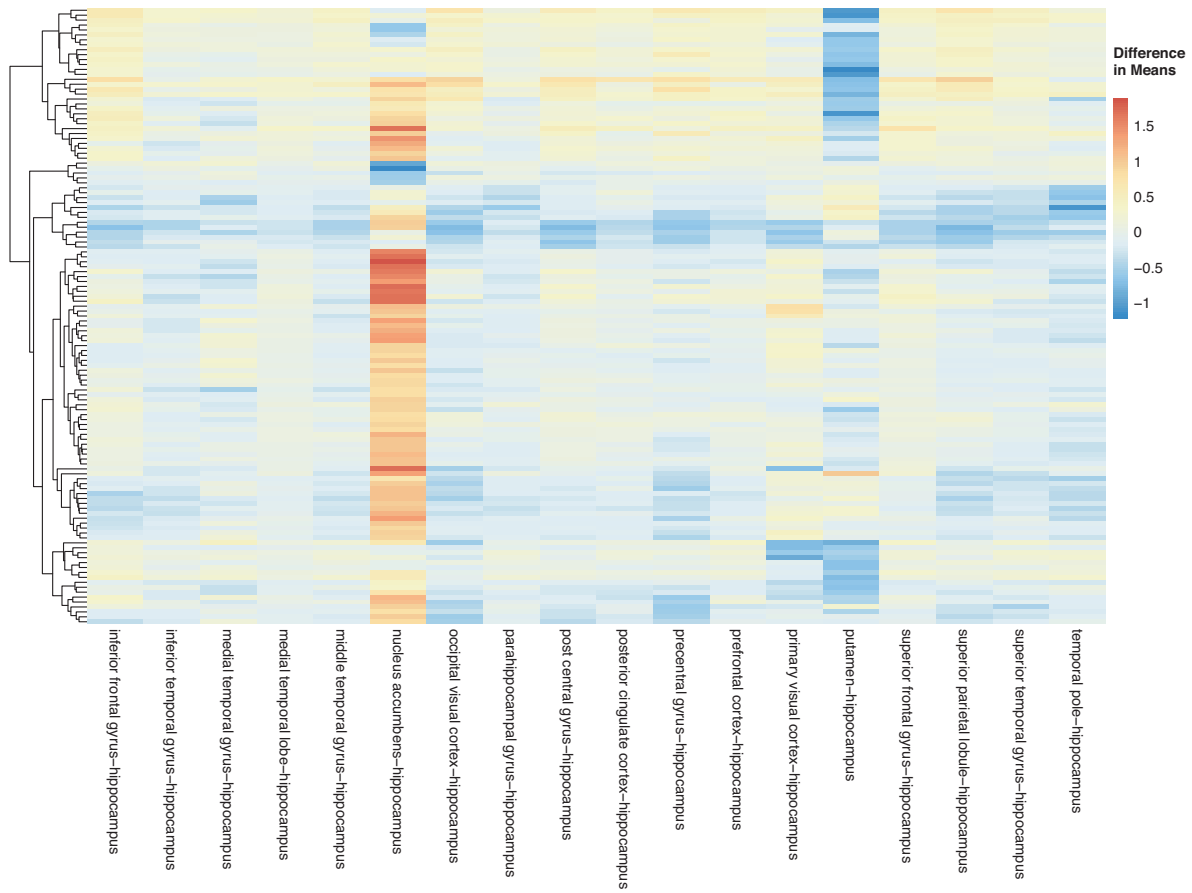


Figure A.11: Heatmap with gene clustering to visualize tissue (hippocampus as baseline) effect using the difference in means (prior to selecting for interacting genes) on the differentially expressed disease (control-AD) gene list.

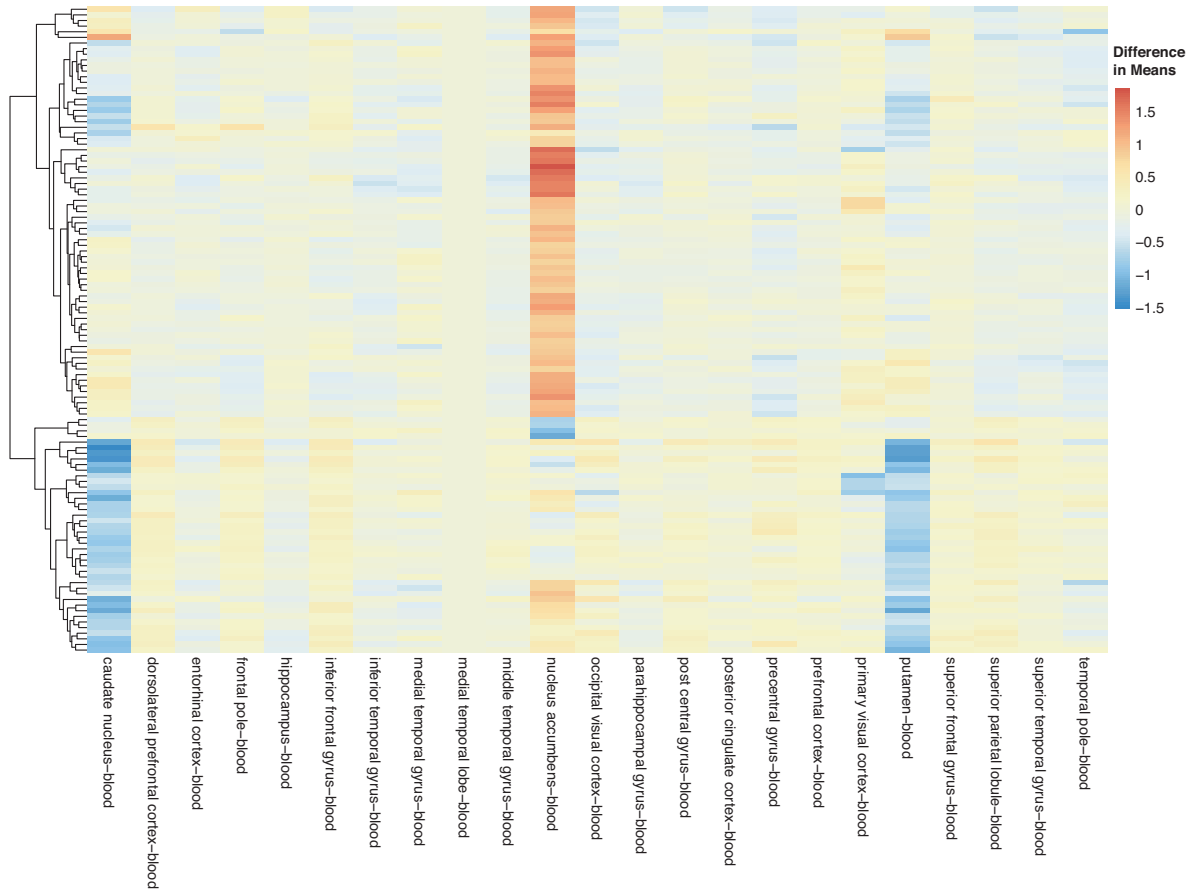


Figure A.12: Heatmap with gene clustering to visualize tissue (blood as baseline) effect using the differences in means between binary comparisons (prior to selecting for interacting genes) on the differentially expressed disease (control-AD) gene list .

A.3 Supplementary Tables

Dataset	Cognitive Data Reported	Additional Notes	Brain Bank
GSE84422	<ul style="list-style-type: none"> Braak stage Neuropathological category Clinical dementia rating CERAD scores Sum of neurofibrillary tangles density Average plaque density 	<ul style="list-style-type: none"> AD: Probable/possible/definite Post-mortem APOE genotype not reported Full spectrum of clinical and neuropathological disease severity Excluded subjects with non-AD neuropathology Mount Sinai and JJ Peters Institutional Review Boards approved protocols 	Mount Sinai Medical Center Brain Bank
GSE28146	<ul style="list-style-type: none"> Mini-mental state examination Braak stage (averaged) Neurofibrillary tangle density (averaged) 	<ul style="list-style-type: none"> AD: Incipient/moderate/severe Post-mortem APOE genotype not reported Alzheimer's Disease and Related Disorders Association criteria 	Brain Bank of Alzheimer's Disease Research Center at the University of Kentucky
GSE48350	<ul style="list-style-type: none"> Braak stage Mini-mental state examination 	<ul style="list-style-type: none"> Post-mortem APOE genotype reported Excluded subjects with evidence of alcoholism, co-existing major psychiatric illness or major depression, pre-existing brain damage, brain metastases and cerebral vascular disease Excluded subjects with non-AD neuropathology 	National Institute on Aging Alzheimer's Disease brain banks
GSE5281	<ul style="list-style-type: none"> Braak stage (range provided, not reported per sample) CERAD scores (range provided, not reported per sample) 	<ul style="list-style-type: none"> AD: Late-onset AD Post-mortem APOE genotype not reported Clinically and neuropathologically classified late-onset AD-afflicted individuals Braak stage of V or VI 	Sun Health Research Institute and Alzheimer's Disease Center at Washington University and Duke University
GSE63060-1	<ul style="list-style-type: none"> Clinical dementia rating and sum of boxes score (averaged) Mini-mental state examination 	<ul style="list-style-type: none"> Living volunteers APOE genotype not reported Ethical approval received from Institutional Research Ethics Committee 	Not applicable

Table A.1: Additional information reported from datasets on samples used for the meta-analysis.

Table A.1 (cont'd)

Dataset	Cognitive Data Reported	Additional Notes	Brain Bank
GSE29378	<ul style="list-style-type: none"> Braak Stage reported for some samples Plaque disease burden Disease Duration 	<ul style="list-style-type: none"> AD:Late-onset AD Post-mortem APOE genotype reported for some samples National Institute for Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorder Association diagnostic criteria for clinical AD Neuropathologic confirmation at autopsy 	Alzheimer's Disease Center, Oregon Health and Sciences University and Human Brain and Spinal Fluid Resource Center
E-MEXP-2280	<ul style="list-style-type: none"> Braak Stage reported for some samples MAPT haplotype 	<ul style="list-style-type: none"> AD: Braak stage VI Post-mortem APOE genotype reported All patients were screened for Microtubule Associated Protein Tau (MAPT) and Progranulin (GRN) mutations and MAPT haplotyping 	Netherlands Brain Bank

Quantile	Disease (control-AD)	Sex (male-female)	AgeGroup (i-<60)	Tissue (i-blood)	Tissue (i-hippocampus)
0.1%	-0.1980181	-0.1850368	-1.8342734	-1.4193659	-1.2176260
1%	-0.1409218	-0.1662240	-1.5978577	-1.1791001	-0.8806144
2.5%	-0.1242022	-0.1531487	-1.4093136	-0.9535290	-0.6994610
5%	-0.1109185	-0.1286464	-1.2380410	-0.7905619	-0.5988491
10%	-0.0944796	-0.0863796	-1.0477827	-0.6359497	-0.5187091
90%	0.1195751	0.2502144	0.3308682	0.7932871	0.8181017
95%	0.1398357	0.2678312	0.4815650	1.0342074	1.0113406
97.5%	0.1597702	0.2788782	0.6502154	1.2459840	1.2049823
99%	0.1851621	0.3036726	0.8852537	1.5229805	1.6578388
99.9%	0.2625072	0.3698125	1.1441852	1.7230551	1.7531744

Table A.2: Quantiles on differences of means between group comparisons from TukeyHSD analysis for each factor with the 10% and 90% highlighted.

Gene	diff	lwr	upr	tukey.p.adj
SNAP91	0.38789336	0.3162492	0.45953751	<5.91E-12
AMPH	0.261109	0.227946	0.29427199	5.91E-12
CABP1	0.25221566	0.2197738	0.28465749	5.91E-12
CCK	0.2736884	0.2290262	0.31835057	5.91E-12
CHGB	0.27223361	0.2330743	0.31139295	5.91E-12
CPQ	-0.15134075	-0.1739617	-0.12871979	5.91E-12
CXCR4	-0.18569224	-0.2153576	-0.15602687	5.91E-12
DIRAS2	0.26469562	0.2210822	0.30830901	5.91E-12
EEF1A2	0.29830496	0.2546981	0.34191185	5.91E-12
GABRG2	0.28727303	0.2421887	0.33235734	5.91E-12
GFAP	0.26148848	0.2372962	0.28568071	5.91E-12
GJA1	0.30536761	0.279178	0.33155722	5.91E-12
KLF2	-0.16010858	-0.1822269	-0.13799028	5.91E-12
MYT1L	0.25975404	0.2232677	0.29624039	5.91E-12
NEFL	0.27515335	0.2353672	0.31493946	5.91E-12
NRN1	0.26422817	0.2300824	0.29837391	5.91E-12
RGS4	0.27860758	0.2432385	0.31397667	5.91E-12
SERPINI1	0.26217204	0.2301731	0.29417102	5.91E-12
SH3GL2	0.30717515	0.2681965	0.34615382	5.91E-12
IL13RA1	-0.15586061	-0.1816785	-0.13004272	5.91E-12
ERC2	0.26822985	0.2230493	0.31341042	5.91E-12
GAD1	0.26177649	0.2178336	0.30571939	5.91E-12
SLC40A1	-0.18276614	-0.2136628	-0.15186946	5.91E-12

Table A.3: TukeyHSD results (male-female) table of statistically significant differentially expressed disease genes with sex effect.

Table A.3 (cont'd)

Gene	diff	lwr	upr	tukey.p.adj
ITIH5	-0.16639611	-0.1947932	-0.137999	5.91E-12
FAM19A1	0.269102	0.2230286	0.31517537	5.91E-12
FGF13	0.25310759	0.2088214	0.29739382	5.92E-12
AHNAK	-0.10311728	-0.1242383	-0.08199628	5.93E-12
RPA3	-0.13337106	-0.1607507	-0.10599143	5.93E-12
EZR	-0.1182311	-0.141882	-0.09458025	5.93E-12
ITPKB	-0.11649742	-0.1417882	-0.09120668	5.93E-12
GABRA1	0.27928408	0.2277225	0.33084567	5.93E-12
MAP3K1	-0.16567888	-0.1964814	-0.13487641	5.93E-12
NOTCH1	-0.10639043	-0.1270303	-0.0857506	5.93E-12
HVCN1	-0.10966873	-0.1333269	-0.0860106	5.93E-12
PCDH8	0.26623656	0.2038808	0.32859232	5.93E-12
LDLRAP1	-0.13026125	-0.1611422	-0.09938027	5.93E-12
GMPR	-0.14621751	-0.1812927	-0.11114232	5.94E-12
CYBRD1	-0.1288122	-0.1605469	-0.09707747	5.94E-12
PRKD2	-0.09889483	-0.1237886	-0.07400105	5.94E-12
PRKX	-0.12798343	-0.1609683	-0.09499854	5.97E-12
STMN2	0.25735011	0.1839117	0.33078852	8.42E-12
HIP1	-0.11198115	-0.1436469	-0.08031542	1.16E-11
FOS	-0.15132275	-0.1952695	-0.10737604	2.55E-11
FAM107B	-0.10351231	-0.1344653	-0.07255936	7.72E-11
RNF135	-0.0875083	-0.1183122	-0.05670445	2.92E-08
ID3	-0.10925012	-0.1502501	-0.06825012	1.94E-07

APPENDIX B

META-ANALYSIS OF GENE EXPRESSION MICROARRAY DATASETS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE SUPPLEMENTARY DATA

B.1 Online Supplementary Data Files

Our datasets, data files and results generated in our COPD meta-analysis have been deposited to FigShare. The supplemental file names begin with the prefix “DF” and are referred to throughout the chapter. To access the FigShare online repository: <https://doi.org/10.6084/m9.figshare.8233175>

B.2 Supplementary Figures

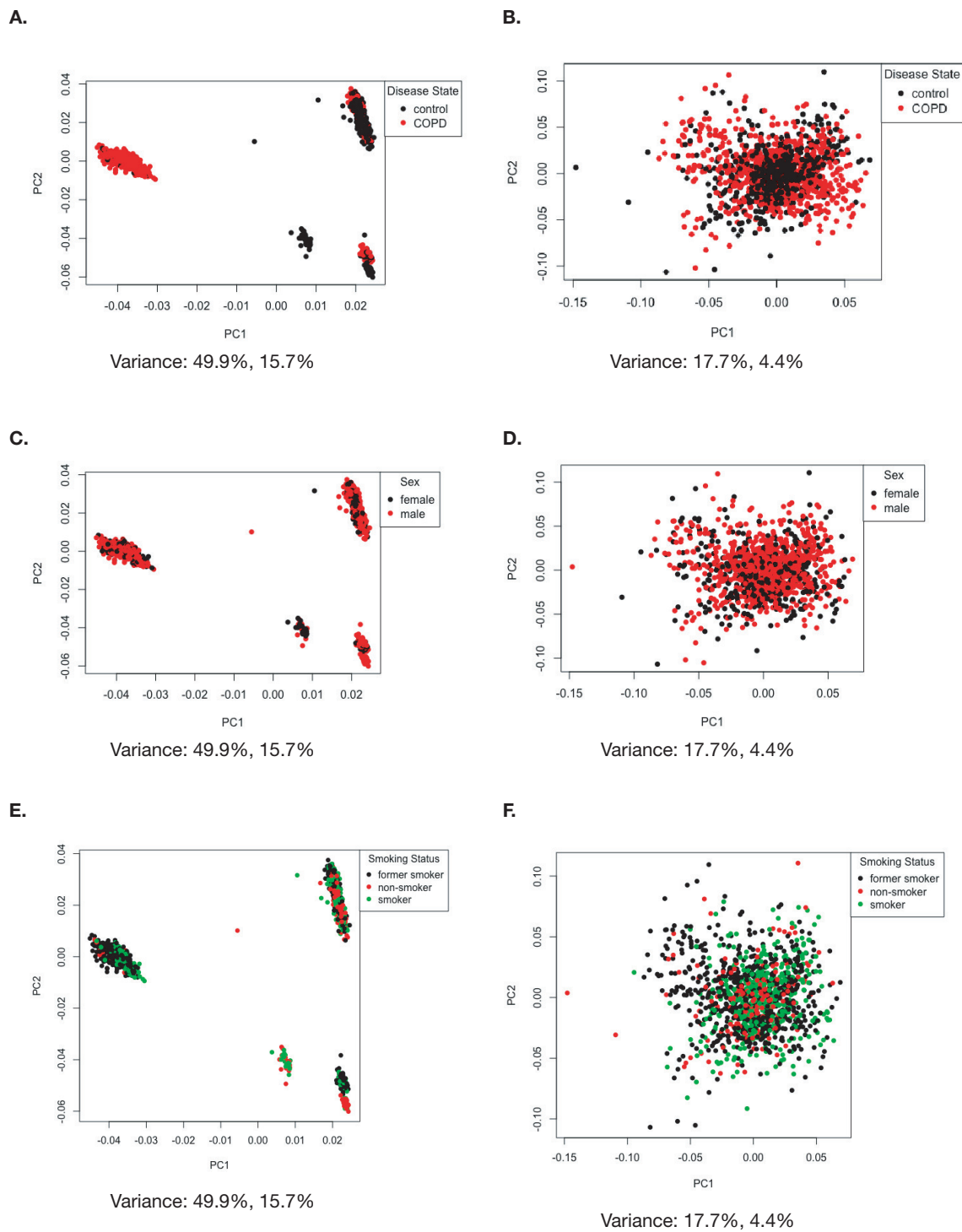


Figure B.1: Principal Component Analysis to visualize changes in variation in datasets before and after combat.

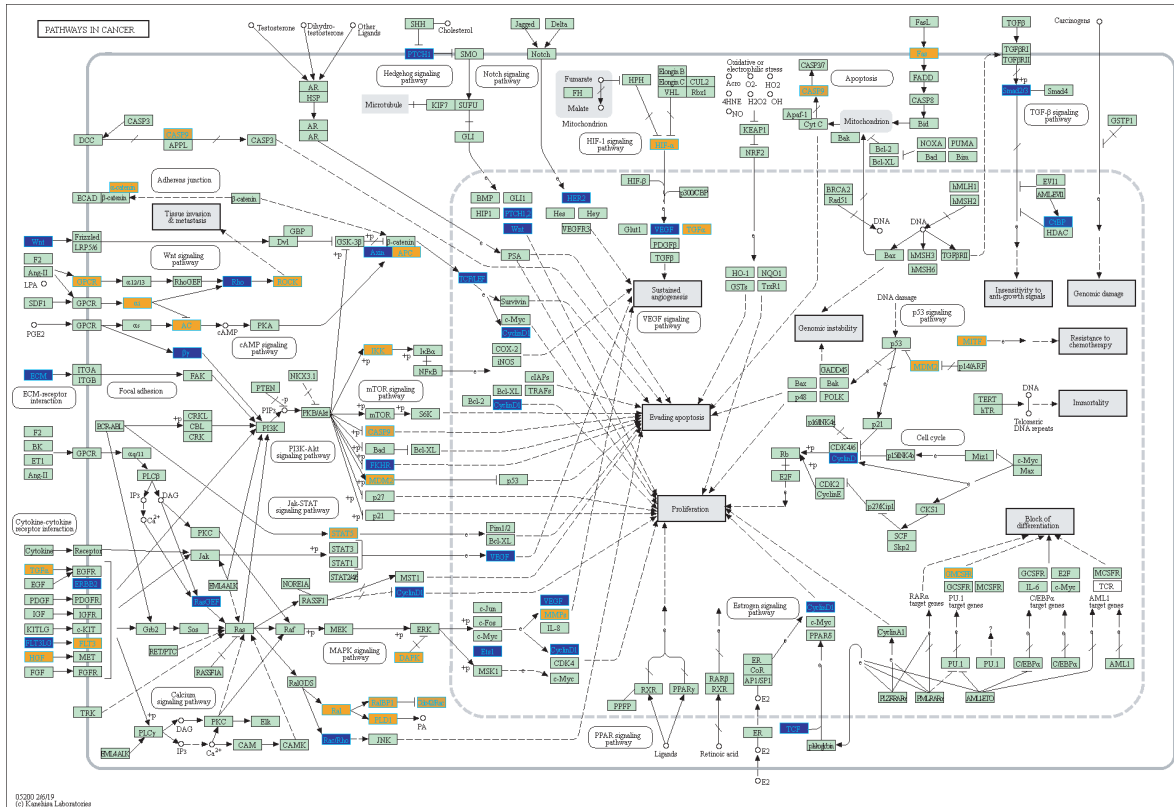


Figure B.2: Highlighted Pathways in Cancer KEGG Pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor) [1–3]

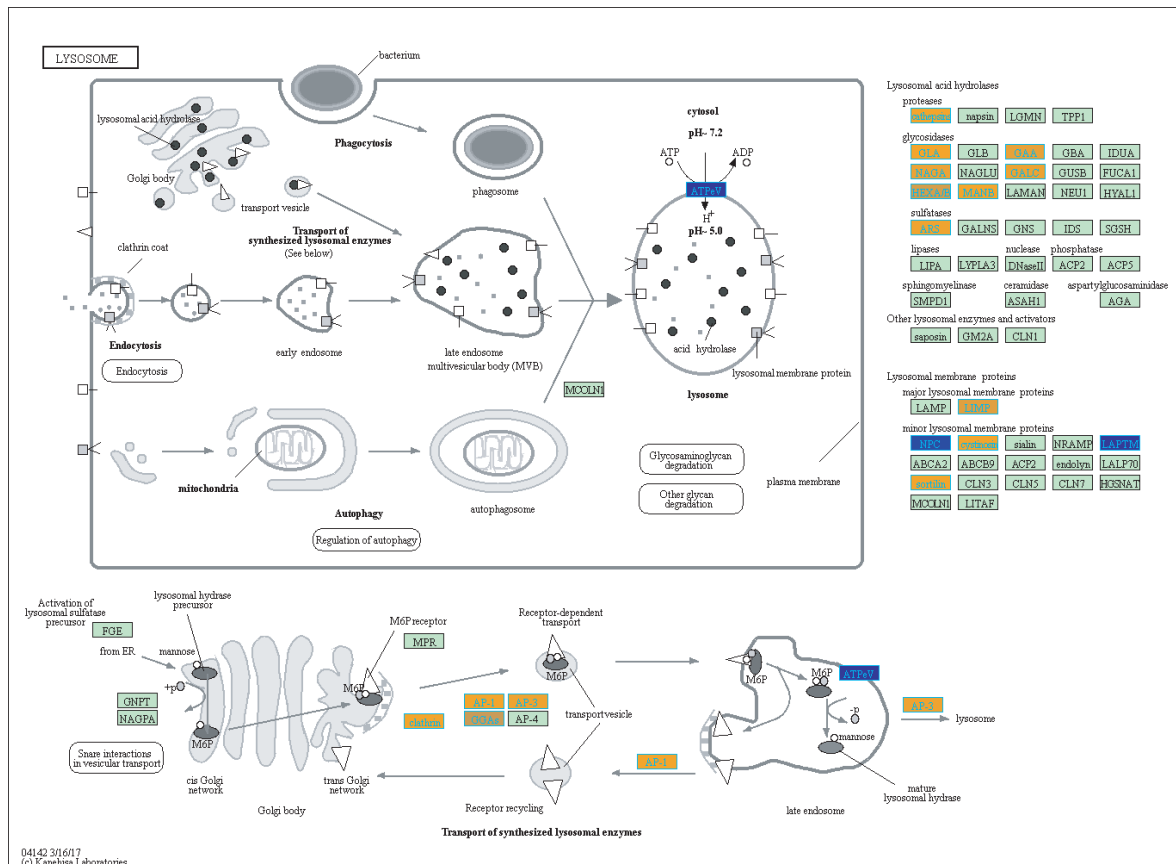


Figure B.3: Highlighted Lysosome KEGG Pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor). [1–3]

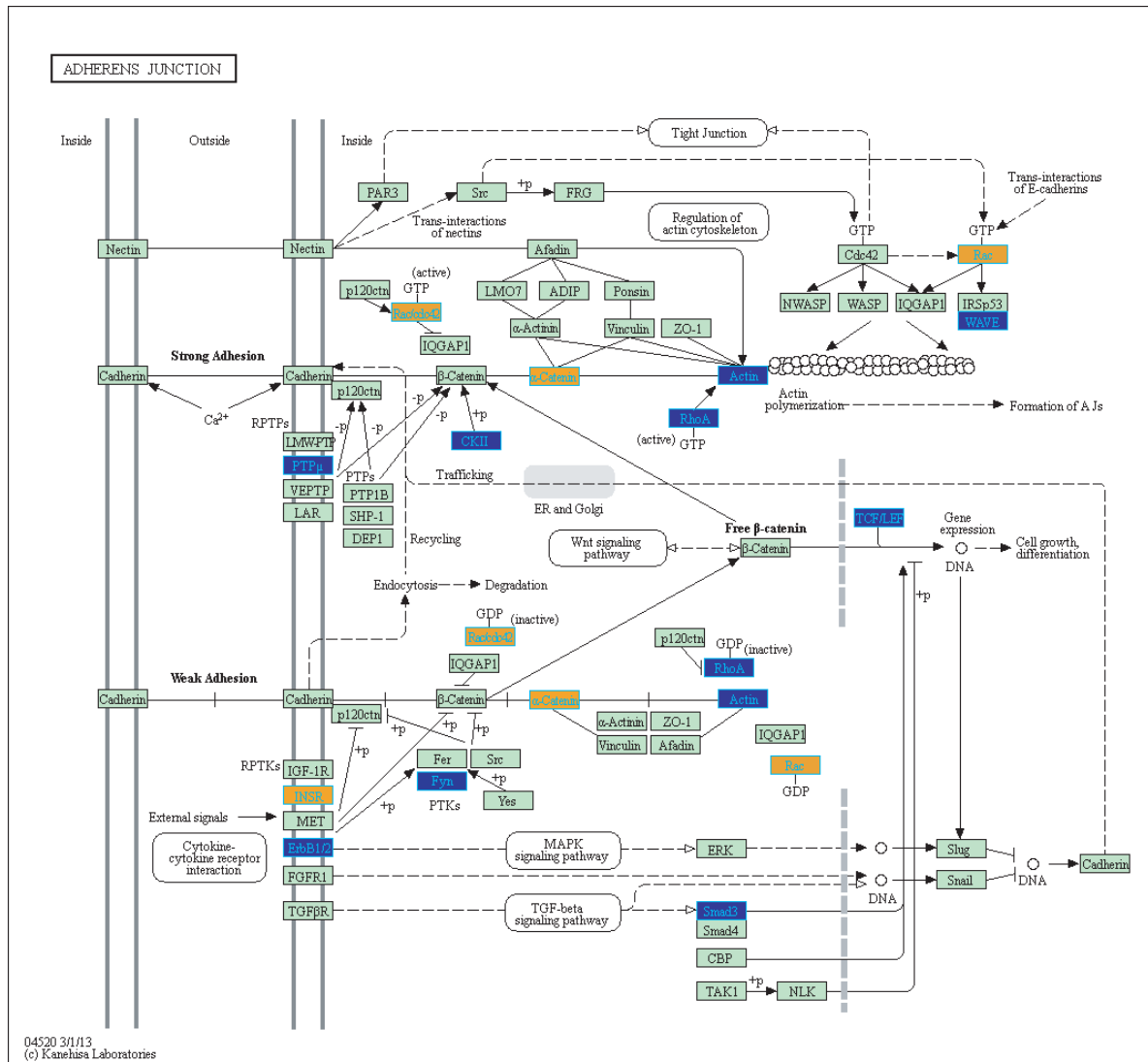


Figure B.4: Highlighted Adherens KEGG Pathway with enriched genes from the ANOVA (BH-adjusted p-value < 0.05; disease status factor)[1–3]

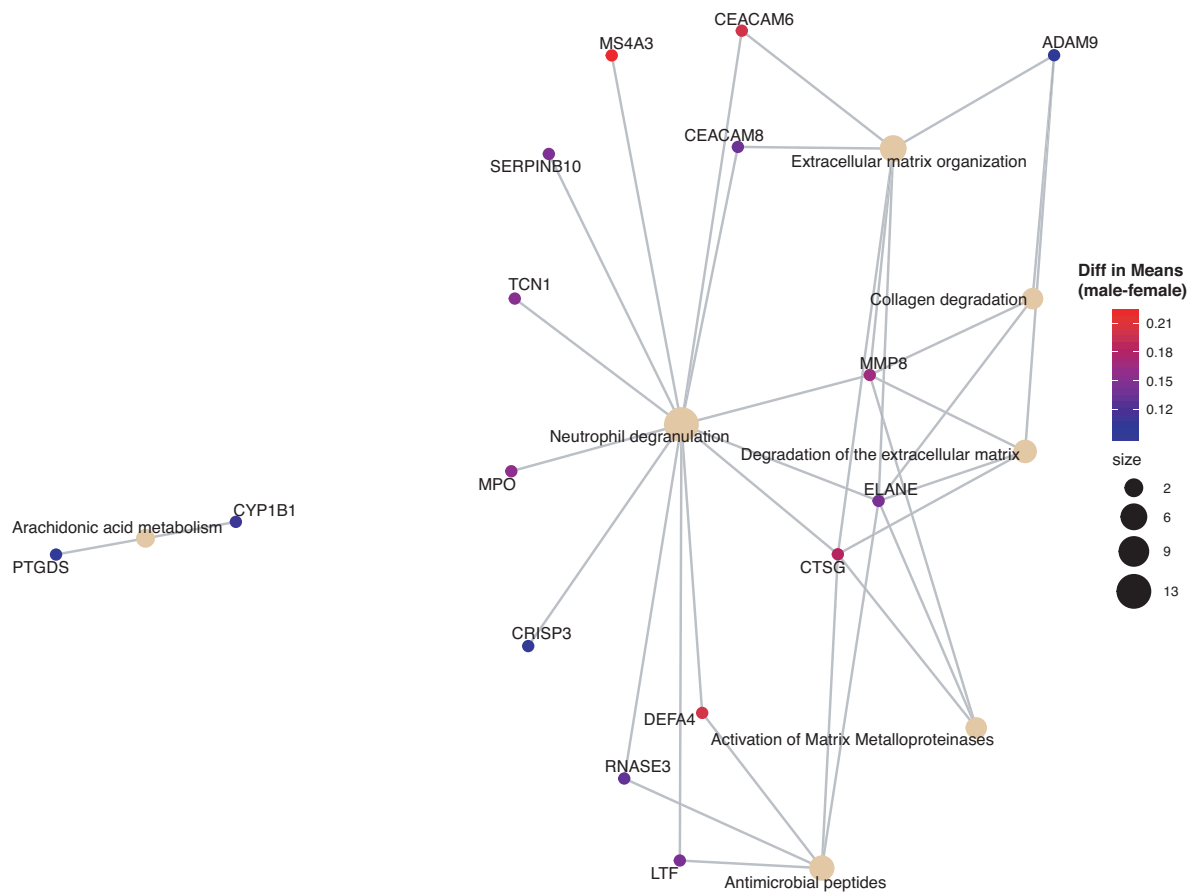


Figure B.7: Enriched Reactome pathway-gene network using the differentially expressed disease genes with a sex effect (no significant interaction between sex and disease) that were up-regulated in males).

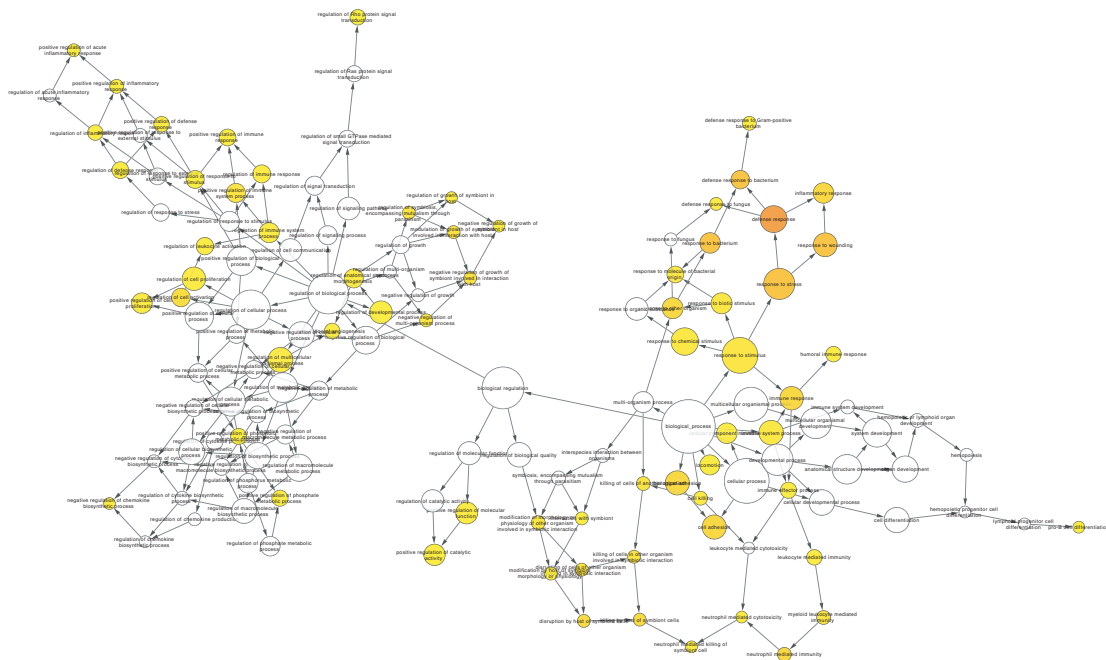


Figure B.8: Gene ontology results from BINGO using our 207 unique statistically significant disease genes filtered for biological effect. Our 304 biologically significant genes were compared to Reinhold et al., [5]

APPENDIX C

MICROARRAY GENE EXPRESSION DATASET RE-ANALYSIS REVEALS VARIABILITY IN INFLUENZA INFECTION AND VACCINATION SUPPLEMENTARY DATA

C.1 Online Supplementary Data Files

Our datasets, data files and results generated in our Influenza meta-analysis have been deposited to FigShare. The supplementary file names begin with the prefix “SDF” and are referred to throughout the chapter. To access the FigShare online repository: <https://doi.org/10.6084/m9.figshare.8636498>

C.2 Supplementary Figures

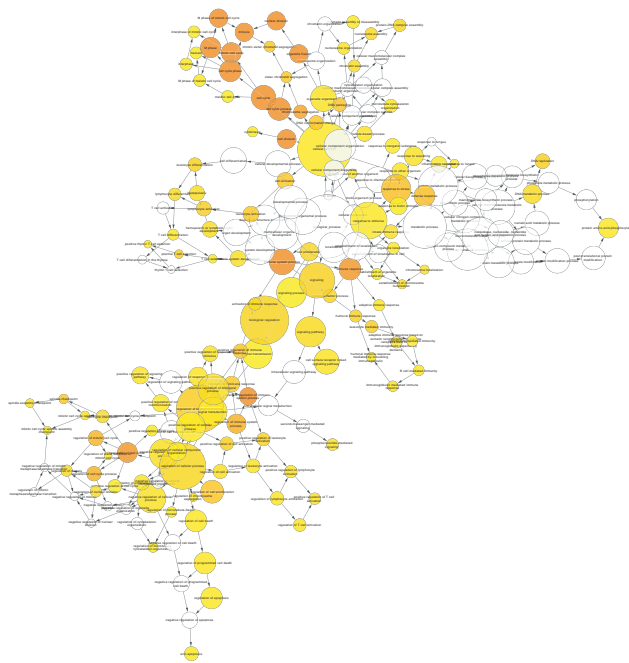


Figure C.1: Gene Ontology of Biologically Significant Genes for Influenza Infected Subjects using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05 .

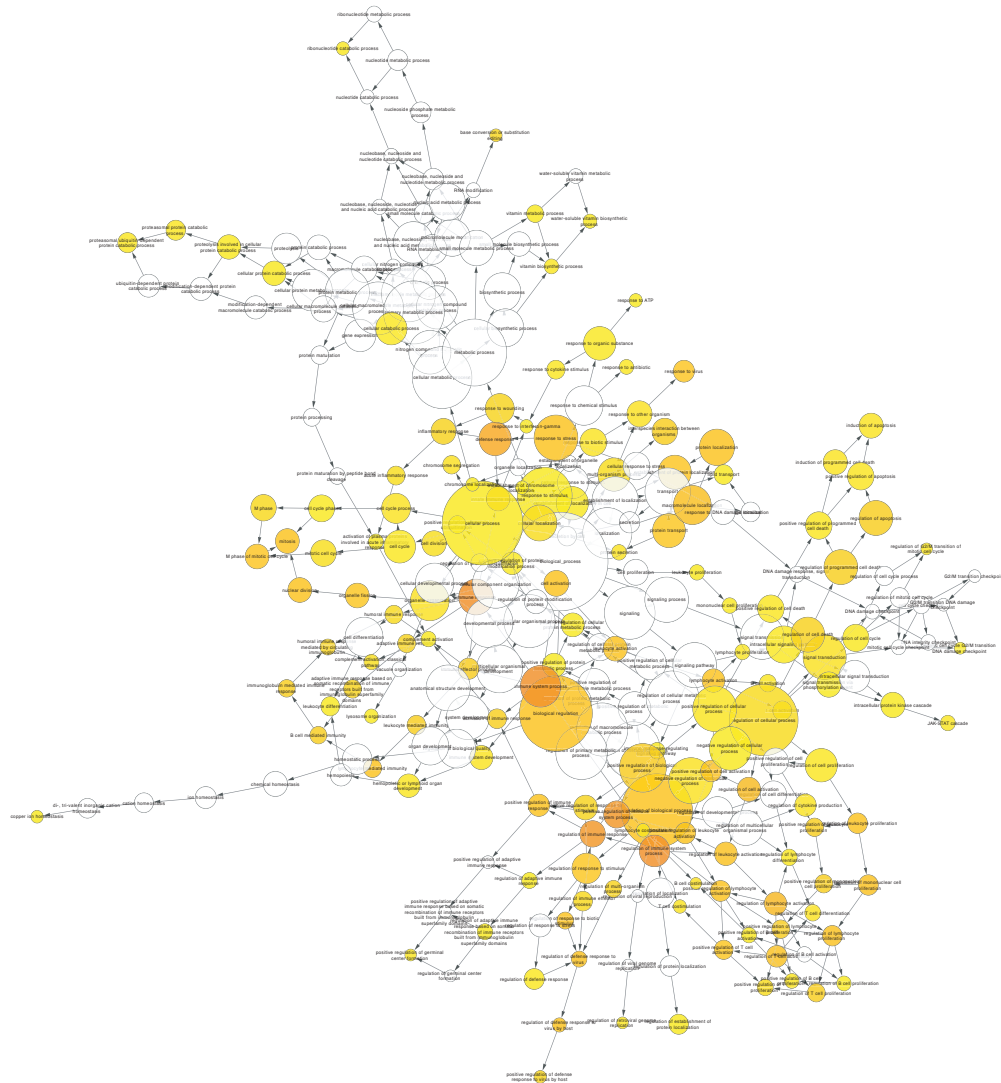


Figure C.2: Gene Ontology of Biologically Significant Genes for Influenza Vaccinated Subjects using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.



Figure C.5: Gene Ontology of Biologically Significant Common Genes for the Influenza Infected and Vaccinated Subjects using BINGO. The node size relates to number of genes, and the yellow nodes are statistically significant with a p-value < 0.05 and false discovery rate < 0.05.

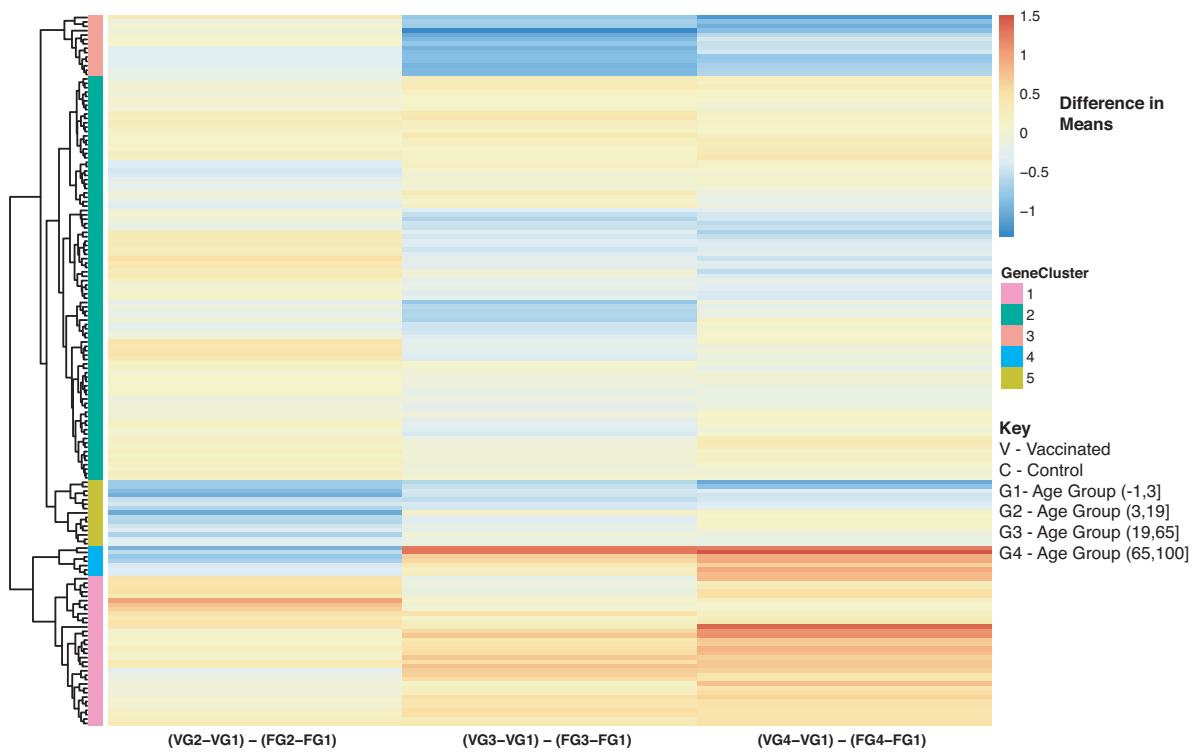


Figure C.6: Heatmap of Biologically Significant Common Genes for the Influenza Infected and Vaccinated Subjects with an Interaction Between Disease State and Age. Comparison of baseline-adjusted means for influenza vaccinated subjects and influenza infected subjects

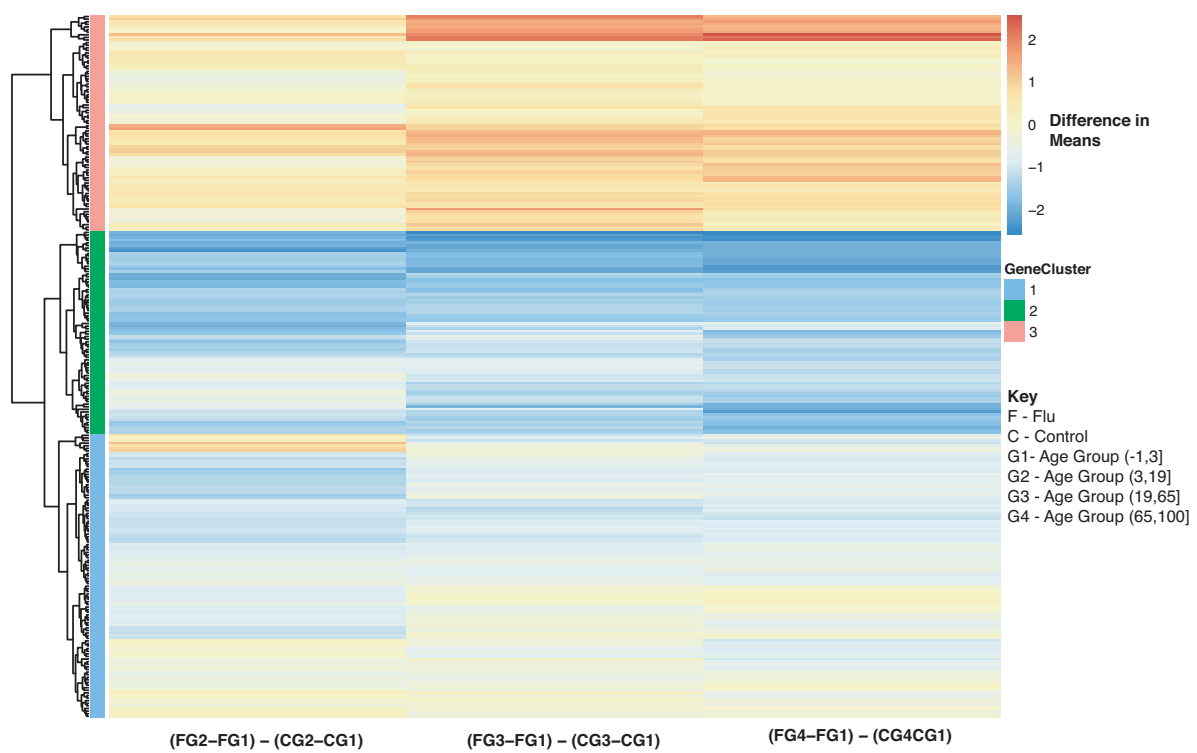


Figure C.7: Heatmap of Biologically Significant Genes Only in the Influenza Infected Gene List with an Interaction Between Disease State and Age. Comparison of baseline-adjusted means for influenza infected subjects and controls

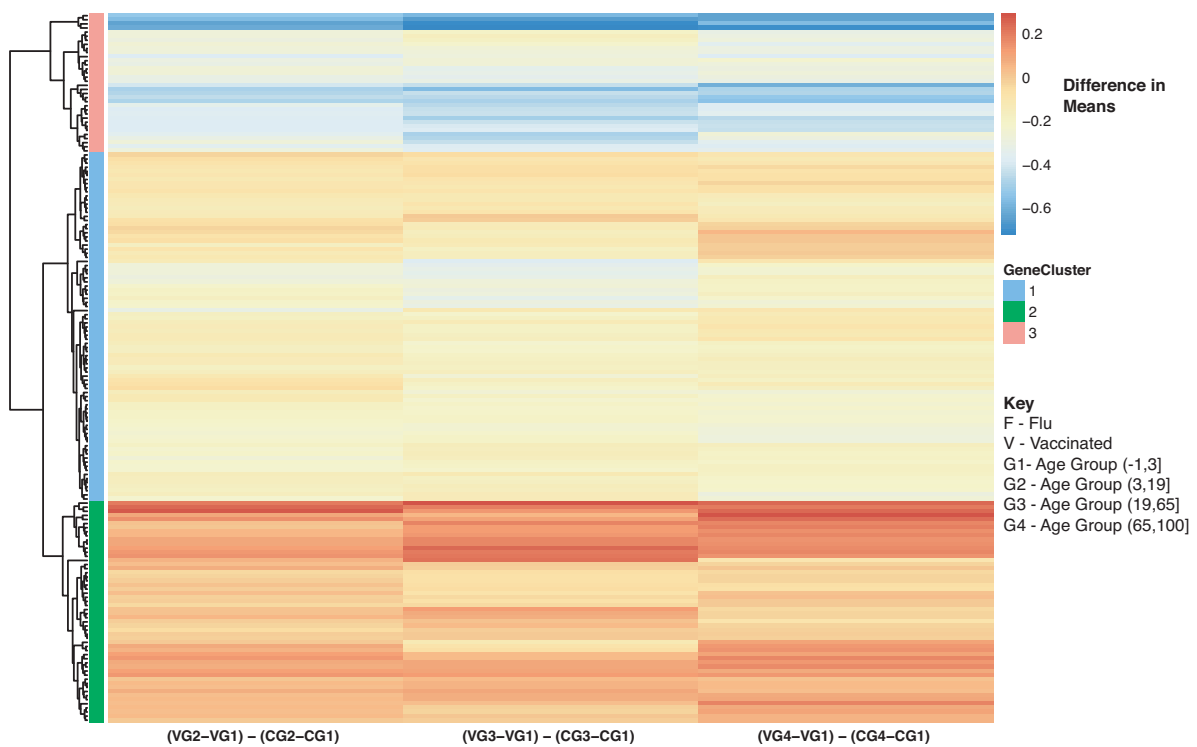


Figure C.8: Heatmap of Biologically Significant Genes Only in the Influenza Vaccinated Gene List with an Interaction Between Disease State and Age. Comparison of baseline-adjusted means for influenza vaccinated subjects and controls

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, 2000.
- [2] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. Kegg as a reference resource for gene and protein annotation. *Nucleic Acids Res*, 44(D1):D457–62, 2016.
- [3] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima. Kegg: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*, 45(D1):D353–D361, 2017.
- [4] Institute for Health Metrics and Evaluation (IHME). Gbd compare data visualization. (May 2018), 2017.
- [5] Dominik Reinhold, Jarrett D Morrow, Sean Jacobson, Junxiao Hu, Benjamin Ringel, Max A Seibold, Craig P Hersh, Katerina J Kechris, and Russell P Bowler. Meta-analysis of peripheral blood gene expression modules for copd phenotypes. *PloS one*, 12(10):e0185682, 2017.
- [6] Danielle Aw, Alberto B Silva, and Donald B Palmer. Immunosenescence: emerging challenges for an ageing population. *Immunology*, 120(4):435–446, 2007.
- [7] C. Lopez-Otin, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–217, 2013.
- [8] Ian Paul Johnson. Age-related neurodegenerative disease research needs aging models. *Frontiers in aging neuroscience*, 7:168, 2015.
- [9] Centers for Disease Control and Prevention. Seasonal influenza (flu). 2018. Available at <https://www.cdc.gov/flu/about/index.html>, Accessed: 2019-06-26.
- [10] Lavidia RK Brooks and George I Mias. Streptococcus pneumoniae’s virulence and host immunity: aging, diagnostics and prevention. *Frontiers in immunology*, 9:1366, 2018.
- [11] MA Julie Westerink, Harry W Schroeder Jr, and Moon H Nahm. Immune responses to pneumococcal vaccines in children and adults: rationale for age-specific vaccination. *Aging and disease*, 3(1):51, 2012.
- [12] Centers for Disease Control and Prevention. Recommended immunization schedules for persons aged 0 through 18 years. *Morbidity and Mortality Weekly Report*, 61(4), 2012.
- [13] Centers for Disease Control and Prevention. Intervals between pcv13 and ppsv23 vaccines: Recommendations of the advisory committee on immunization practices (acip). *Morbidity and Mortality Weekly Report*, 64(34):944–947, 2015.
- [14] Centers for Disease Control and Prevention. Chronic obstructive pulmonary disease (copd), 2019, (Accessed: 2019-06-02). Available at <https://www.cdc.gov/copd/basics-about.html>.

- [15] Jan M Van Deursen. The role of senescent cells in ageing. *Nature*, 509(7501):439, 2014.
- [16] Bennett G Childs, Matej Durik, Darren J Baker, and Jan M Van Deursen. Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nature medicine*, 21(12):1424, 2015.
- [17] J. W. Rowe, T. Fulmer, and L. Fried. Preparing for better health and health care for an aging population. *JAMA*, 316(16):1643–1644, 2016.
- [18] Steven Black, Ennio De Gregorio, and Rino Rappuoli. Developing vaccines for an aging population. *Science translational medicine*, 7(281):281ps8–281ps8, 2015.
- [19] E. Jaul and J. Barron. Age-related diseases and clinical and public health implications for the 85 years old and over population. *Front Public Health*, 5:335, 2017.
- [20] United Nations Department of Economic and Social Affairs. World population ageing 2015, 2015. Accessed: 2018-11-01.
- [21] Sören Dallmeyer, Pamela Wicker, and Christoph Breuer. How an aging society affects the economic costs of inactivity in germany: empirical evidence and projections. *European Review of Aging and Physical Activity*, 14(1):18, 2017.
- [22] S. Jevtic, A. S. Sengar, M. W. Salter, and J. McLaurin. The role of the immune system in alzheimer disease: Etiology and treatment. *Ageing Res Rev*, 40:84–94, 2017.
- [23] M. P. Mattson and T. V. Arumugam. Hallmarks of brain aging: Adaptive and pathological modification by metabolic states. *Cell Metab*, 27(6):1176–1199, 2018.
- [24] Burton P Drayer. Imaging of the aging brain. *Radiology*, 166:785–796, 1988.
- [25] Christopher A Taylor, Sujay F Greenlund, Lisa C McGuire, Hua Lu, and Janet B Croft. Deaths from alzheimer’s disease—united states, 1999–2014. *MMWR. Morbidity and mortality weekly report*, 66(20):521, 2017.
- [26] Kevin A. Matthews, Wei Xu, Anne H. Gaglioti, James B. Holt, Janet B. Croft, Dominic Mack, and Lisa C. McGuire. Racial and ethnic estimates of alzheimer’s disease and related dementias in the united states (2015–2060) in adults aged > 65 years. *Alzheimer’s & Dementia*, 2018.
- [27] R. Brookmeyer, N. Abdalla, C. H. Kawas, and M. M. Corrada. Forecasting the prevalence of preclinical and clinical alzheimer’s disease in the united states. *Alzheimers Dement*, 14(2):121–129, 2018.
- [28] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans. Alzheimer disease in the united states (2010–2050) estimated using the 2010 census. *Neurology*, 80(19):1778–83, 2013.
- [29] Colin L. Masters, Randall Bateman, Kaj Blennow, Christopher C. Rowe, Reisa A. Sperling, and Jeffrey L. Cummings. Alzheimer’s disease. *Nature Reviews Disease Primers*, 1:15056, 2015.

- [30] Phillip De Jager, Yiyi Ma, Cristin McCabe, Jishu Xu, Badri N Vardarajan, Daniel Felsky, Hans-Ulrich Klein, Charles C White, Mette A Peters, Ben Lodgson, et al. A multi-omic atlas of the human frontal cortex for aging and alzheimer’s disease research. *bioRxiv*, page 251967, 2018.
- [31] M. Toepper. Dissociating normal aging from alzheimer’s disease: A view from cognitive neuroscience. *J Alzheimers Dis*, 57(2):331–352, 2017.
- [32] J. M. Winkler and H. S. Fox. Transcriptome meta-analysis reveals a central role for sex steroids in the degeneration of hippocampal neurons in alzheimer’s disease. *BMC Syst Biol*, 7:51, 2013.
- [33] Q. Wang, W. X. Li, S. X. Dai, Y. C. Guo, F. F. Han, J. J. Zheng, G. H. Li, and J. F. Huang. Meta-analysis of parkinson’s disease and alzheimer’s disease revealed commonly impaired pathways and dysregulation of nrf2-dependent genes. *J Alzheimers Dis*, 56(4):1525–1539, 2017.
- [34] Shirin Moradifard, Moslem Hoseinbeyki, Shahla Mohammad Ganji, and Zarrin Minuchehr. Analysis of microrna and gene expression profiles in alzheimer’s disease: A meta-analysis approach. *Scientific reports*, 8(1):4767, 2018.
- [35] John W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2):99–114, 1949.
- [36] G. Yu and Q. Y. He. Reactomepa: an r/bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst*, 12(2):477–9, 2016.
- [37] G. Yu, L. G. Wang, Y. Han, and Q. Y. He. clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS*, 16(5):284–7, 2012.
- [38] S. Maere, K. Heymans, and M. Kuiper. Bingo: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–9, 2005.
- [39] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.
- [40] M. Wang, P. Roussos, A. McKenzie, X. Zhou, Y. Kajiwara, K. J. Brennand, G. C. De Luca, J. F. Crary, P. Casaccia, J. D. Buxbaum, M. Ehrlich, S. Gandy, A. Goate, P. Katsel, E. Schadt, V. Haroutunian, and B. Zhang. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to alzheimer’s disease. *Genome Med*, 8(1):104, 2016.
- [41] Eric M Blalock, Heather M Buechel, Jelena Popovic, James W Geddes, and Philip W Landfield. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient alzheimer’s disease. *Journal of chemical neuroanatomy*, 42(2):118–126, 2011.

- [42] Nicole C Berchtold, David H Cribbs, Paul D Coleman, Joseph Rogers, Elizabeth Head, Ronald Kim, Tom Beach, Carol Miller, Juan Troncoso, John Q Trojanowski, et al. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proceedings of the National Academy of Sciences*, 2008.
- [43] Winnie S Liang, Travis Dunckley, Thomas G Beach, Andrew Grover, Diego Mastroeni, Douglas G Walker, Richard J Caselli, Walter A Kukull, Daniel McKeel, John C Morris, et al. Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. *Physiological genomics*, 28(3):311–322, 2007.
- [44] Sanjana Sood, Iain J Gallagher, Katie Lunnon, Eric Rullman, Aoife Keohane, Hannah Crossland, Bethan E Phillips, Tommy Cederholm, Thomas Jensen, Luc JC van Loon, et al. A novel multi-tissue rna diagnostic of healthy ageing relates to cognitive health status. *Genome biology*, 16(1):185, 2015.
- [45] Jeremy A Miller, Randall L Woltjer, Jeff M Goodenbour, Steve Horvath, and Daniel H Geschwind. Genes and pathways underlying regional and cell type changes in alzheimer’s disease. *Genome medicine*, 5(5):48, 2013.
- [46] Iraad F Bronner, Zoltán Bochdanovits, Patrizia Rizzu, Wouter Kamphorst, Rivka Ravid, John C van Swieten, and Peter Heutink. Comprehensive mrna expression profiling distinguishes tauopathies and identifies shared molecular pathways. *PLoS One*, 4(8):e6826, 2009.
- [47] R. Edgar, M. Domrachev, and A. E. Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 30(1):207–10, 2002.
- [48] Alvis Brazma, Helen Parkinson, Ugis Sarkans, Mohammadreza Shojatalab, Jaak Vilo, Niran Abeygunawardena, Ele Holloway, Misha Kapushesky, Patrick Kemmeren, Gonzalo Garcia Lara, et al. Arrayexpress—a public repository for microarray gene expression data at the ebi. *Nucleic acids research*, 31(1):68–71, 2003.
- [49] George Mias. *Databases: E-Utilities and UCSC Genome Browser*, chapter 4, pages 133–170. Springer International Publishing, Cham, 2018.
- [50] Wolfram Research, Inc. *Mathematica*, version 11.2 edition, 2017.
- [51] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [52] Laurent Gautier, Leslie Cope, Benjamin M Bolstad, and Rafael A Irizarry. affy—analysis of affymetrix genechip data at the probe level. *Bioinformatics*, 20(3):307–315, 2004.
- [53] Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.
- [54] RM Sakia. The box-cox transformation technique: a review. *The statistician*, pages 169–178, 1992.

- [55] G. I. Mias, T. Yusufaly, R. Roushangar, L. R. Brooks, V. V. Singh, and C. Christou. Math-iomica: An integrative platform for dynamic omics. *Sci Rep*, 6:37237, 2016.
- [56] Vegard Nygaard, Einar Andreas Rødland, and Eivind Hovig. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, 17(1):29–39, 2016.
- [57] W Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1):118–127, 2007.
- [58] Rafael Irizarry and Michael Love. Ph525x series - biomedical data science, 2015. Accessed: 2018-01-18.
- [59] Paul Pavlidis. Using anova for gene selection from microarray studies of the nervous system. *Methods*, 31(4):282–289, 2003.
- [60] George Mias. *Analysis of Variance for Multiple Tests*, chapter 6.3, pages 133–170. Springer International Publishing, Cham, 2018.
- [61] J Martin Bland and Douglas G Altman. Multiple significance tests: the bonferroni method. *BMJ*, 310(6973):170, 1995.
- [62] David A. Elliott, Cyndi Shannon Weickert, and Brett Garner. Apolipoproteins in the brain: implications for neurological and psychiatric disorders. *Clinical lipidology*, 51(4):555–573, 2010.
- [63] Meredith N. Braskie, Neda Jahanshad, Jason L. Stein, Marina Barysheva, Katie L. McMahon, Greig I. de Zubicaray, Nicholas G. Martin, Margaret J. Wright, John M. Ringman, Arthur W. Toga, and Paul M. Thompson. Common alzheimer’s disease risk variant within the clu gene affects white matter microstructure in young adults. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(18):6764–6770, 2011.
- [64] R. H. Swerdlow. Mitochondria and mitochondrial cascades in alzheimer’s disease. *J Alzheimers Dis*, 62(3):1403–1416, 2018.
- [65] Paula I Moreira, Cristina Carvalho, Xiongwei Zhu, Mark A Smith, and George Perry. Mitochondrial dysfunction is a trigger of alzheimer’s disease pathophysiology. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1802(1):2–10, 2010.
- [66] I. G. Onyango, J. Dennis, and S. M. Khan. Mitochondrial dysfunction in alzheimer’s disease and the rationale for bioenergetics based therapies. *Aging Dis*, 7(2):201–14, 2016.
- [67] F. L. Heppner, R. M. Ransohoff, and B. Becher. Immune attack: the role of inflammation in alzheimer disease. *Nat Rev Neurosci*, 16(6):358–72, 2015.
- [68] Linda J Van Eldik, Maria C Carrillo, Patricia E Cole, Dominik Feuerbach, Barry D Greenberg, James A Hendrix, Matthew Kennedy, Nick Kozauer, Richard A Margolin, José L Molinuevo, et al. The roles of inflammation and immune mechanisms in alzheimer’s disease. *Alzheimer’s & Dementia: Translational Research & Clinical Interventions*, 2(2):99–109, 2016.

- [69] S. E. Marsh, E. M. Abud, A. Lakatos, A. Karimzadeh, S. T. Yeung, H. Davtyan, G. M. Fote, L. Lau, J. G. Weinger, T. E. Lane, M. A. Inlay, W. W. Poon, and M. Blurton-Jones. The adaptive immune system restrains alzheimer's disease pathogenesis by modulating microglial function. *Proc Natl Acad Sci U S A*, 113(9):E1316–25, 2016.
- [70] G. A. Prieto, B. H. Trieu, C. T. Dang, T. Bilousova, K. H. Gylys, N. C. Berchtold, G. Lynch, and C. W. Cotman. Pharmacological rescue of long-term potentiation in alzheimer diseased synapses. *J Neurosci*, 37(5):1197–1212, 2017.
- [71] Yuzhi Chen, Angela M Bodles, Donna L McPhie, Rachael L Neve, Robert E Mrak, and W Sue T Griffin. App-bp1 inhibits a β 42 levels by interacting with presenilin-1. *Molecular neurodegeneration*, 2(1):3, 2007.
- [72] Bin Zhang, Qiuwei Wang, Tingting Miao, Bin Yu, Pei Yuan, Jing Kong, and Beiyi Lu. Whether alzheimer's diseases related genes also differently express in the hippocampus of ts65dn mice? *International journal of clinical and experimental pathology*, 8(4):4120, 2015. Available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466988/>, Accessed: 2019-03-01.
- [73] Yuzhi Chen, Wenyun Liu, Donna L McPhie, Linda Hassinger, and Rachael L Neve. App-bp1 mediates app-induced apoptosis and dna synthesis and is increased in alzheimer's disease brain. *J Cell Biol*, 163(1):27–33, 2003.
- [74] Yuzhi Chen, Donna L McPhie, Joseph Hirschberg, and Rachael L Neve. The amyloid precursor protein-binding protein app-bp1 drives the cell cycle through the sm checkpoint and causes apoptosis in neurons. *Journal of Biological Chemistry*, 275(12):8929–8935, 2000.
- [75] Yuzhi Chen, Rachael L Neve, and Helena Liu. Neddylaton dysfunction in alzheimer's disease. *Journal of cellular and molecular medicine*, 16(11):2583–2591, 2012.
- [76] Daphna Laifenfeld, Lucas J Patzek, Donna L McPhie, Yuzhi Chen, Yona Levites, Anne M Cataldo, and Rachael L Neve. Rab5 mediates an amyloid precursor protein signaling pathway that leads to apoptosis. *Journal of Neuroscience*, 27(27):7141–7153, 2007.
- [77] Lars Feuk, Jonathan Prince, Gerome Breen, Tesfai Emahazion, Andrew Carothers, David St Clair, and Anthony Brookes. Apolipoprotein-e dependent role for the fas receptor in early onset alzheimer's disease: finding of a positive association for a polymorphism in the tnfrsf6 gene. *Human genetics*, 107(4):391–396, 2000.
- [78] Brati Das and Riqiang Yan. Role of bace1 in alzheimer's synaptic function. *Translational neurodegeneration*, 6(1):23, 2017.
- [79] Robert Vassar. Bace1: The β -secretase enzyme in alzheimer's disease. *Journal of Molecular Neuroscience*, 23(1-2):105–114, 2004.
- [80] Bart De Strooper and Wim Annaert. Novel research horizons for presenilins and γ -secretases in cell biology and disease. *Annual review of cell and developmental biology*, 26:235–260, 2010.

- [81] Nathalie Jurisch-Yaksi, Ragna Sannerud, and Wim Annaert. A fast growing spectrum of biological functions of γ -secretase in development and disease. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1828(12):2815–2827, 2013.
- [82] Lutgarde Serneels, Tim Dejaegere, Katleen Craessaerts, Katrien Horr , Ellen Jorissen, Thomas Tousseyn, S bastien H bert, Marcel Coolen, Gerard Martens, An Zwijsen, et al. Differential contribution of the three aph1 genes to γ -secretase activity in vivo. *Proceedings of the National Academy of Sciences*, 102(5):1719–1724, 2005.
- [83] Carlos A Saura, Se-Young Choi, Vassilios Beglopoulos, Seema Malkani, Dawei Zhang, BS Shankaranarayana Rao, Sumantra Chattarji, Raymond J Kelleher III, Eric R Kandel, Karen Duff, et al. Loss of presenilin function causes impairments of memory and synaptic plasticity followed by age-dependent neurodegeneration. *Neuron*, 42(1):23–36, 2004.
- [84] Paul O’Callaghan, Fredrik Noborn, Dag Sehlin, Jin-ping Li, Lars Lannfelt, Ulf Lindahl, and Xiao Zhang. Apolipoprotein e increases cell association of amyloid- β 40 through heparan sulfate and lrp1 dependent pathways. *Amyloid*, 21(2):76–87, 2014.
- [85] Krish Chandrasekaran, Kimmo Hatanp  , Stanley I Rapoport, and Daniel R Brady. Decreased expression of nuclear and mitochondrial dna-encoded genes of oxidative phosphorylation in association neocortex in alzheimer disease. *Molecular brain research*, 44(1):99–104, 1997.
- [86] Krish Chandrasekaran, Tony Giordano, Daniel R Brady, James Stoll, Lee J Martin, and Stanley I Rapoport. Impairment in mitochondrial cytochrome oxidase gene expression in alzheimer disease. *Molecular brain research*, 24(1-4):336–340, 1994.
- [87] W Davis Parker, Janice Parks, Christopher M Filley, and BK Kleinschmidt-DeMasters. Electron transport chain defects in alzheimer’s disease brain. *Neurology*, 44(6):1090–1090, 1994. Available at <https://www.ncbi.nlm.nih.gov/pubmed/8208407>.
- [88] William R Markesbery. Oxidative stress hypothesis in alzheimer’s disease. *Free Radical Biology and Medicine*, 23(1):134–147, 1997.
- [89] Rui Bi, Wen Zhang, Deng-Feng Zhang, Min Xu, Yu Fan, Qiu-Xiang Hu, Hong-Yan Jiang, Liwen Tan, Tao Li, Yiru Fang, et al. Genetic association of the cytochrome c oxidase-related genes with alzheimer’s disease in han chinese. *Neuropsychopharmacology*, 43(11):2264, 2018.
- [90] Michael J Berridge. The inositol trisphosphate/calcium signaling pathway in health and disease. *Physiological reviews*, 96(4):1261–1296, 2016.
- [91] Adriana Ferreira. Calpain dysregulation in alzheimer’s disease. *ISRN biochemistry*, 2012, 2012.
- [92] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, and D. Lancet. The genecards suite: From gene data

- mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*, 54:1 30 1–1 30 33, 2016.
- [93] Eduardo Bonilla, Kurenai Tanji, Michio Hirano, Tuan H Vu, Salvatore DiMauro, and Eric A Schon. Mitochondrial involvement in alzheimer’s disease. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1410(2):171–182, 1999.
 - [94] Brenna C Beckelman, Xueyan Zhou, C Dirk Keene, and Tao Ma. Impaired eukaryotic elongation factor 1a expression in alzheimer’s disease. *Neurodegenerative Diseases*, 16(1-2):39–43, 2016.
 - [95] Beatriz Calvo-Flores Guzmán, Chitra Vinnakota, Karan Govindpani, Henry J Waldvogel, Richard LM Faull, and Andrea Kwakowsky. The gabaergic system as a therapeutic target for alzheimer’s disease. *Journal of neurochemistry*, 146(6):649–669, 2018.
 - [96] Claire L Padgett and Paul A Slesinger. Gabab receptor coupling to g-proteins and ion channels. In *Advances in pharmacology*, volume 58, pages 123–147. Elsevier, 2010.
 - [97] Aaradhita Upadhyay, Seyyedmohsen Hosseinibarkooie, Svenja Schneider, Anna Kaczmarek, Laura Torres-Benito, Natalia Mendoza-Ferreira, Melina Overhoff, Roman Rombo, Vanessa Grysko, Min Jeong Kye, and et al. Neurocalcin delta knockout impairs adult neurogenesis whereas half reduction is not pathological. *Frontiers in Molecular Neuroscience*, 12, Feb 2019.
 - [98] Oliver Preische, Stephanie A Schultz, Anja Apel, Jens Kuhle, Stephan A Kaeser, Christian Barro, Susanne Gräber, Elke Kuder-Buletta, Christian LaFougere, Christoph Laske, et al. Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic alzheimer’s disease. *Nature medicine*, page 1, 2019.
 - [99] Keiji TANAKA. The proteasome: Overview of structure and functions. *Proceedings of the Japan Academy, Series B*, 85(1):12–36, 2009.
 - [100] Miriam Kolog Gulko, Gabriele Heinrich, Carina Gross, Blagovesta Popova, Oliver Valerius, Piotr Neumann, Ralf Ficner, and Gerhard H. Braus. Sem1 links proteasome stability and specificity to multicellular development. *PLOS Genetics*, 14(2):e1007141, Feb 2018.
 - [101] Yaw Lin Lin, Changyue Chen, Kylie F Keshav, Ellen Winchester, and Anindya Dutta. Dissection of functional domains of the human dna replication protein complex replication protein a. *The Journal of biological chemistry*, 271 29:17190–8, 1996.
 - [102] Yuri B Yurov, Svetlana G Vorsanova, and Ivan Y Iourov. The dna replication stress hypothesis of alzheimer’s disease. *The Scientific World Journal*, 11:2602–2612, 2011.
 - [103] Vicent Bonet-Costa, Laura Corrales-Diaz Pomatto, and Kelvin J. A. Davies. The proteasome and oxidative stress in alzheimer’s disease. *Antioxidants & Redox Signaling*, 25(16):886–901, Dec 2016.
 - [104] M Lopez Salon, L Pasquini, M Besio Moreno, J.M Pasquini, and E Soto. Relationship between β -amyloid degradation and the 26s proteasome in neural cells. *Experimental Neurology*, 180(2):131 – 143, 2003.

- [105] Sangsoo Oh, Hyun Seok Hong, Enmi Hwang, Hae Jin Sim, Woojin Lee, Su Jeon Shin, and Inhee Mook-Jung. Amyloid peptide attenuates the proteasome activity in neuronal cells. *Mechanisms of Ageing and Development*, 126(12):1292–1299, Dec 2005.
- [106] Frédéric Checler, Cristine Alves da Costa, Karine Ancolio, Nathalie Chevallier, Elvira Lopez-Perez, and Philippe Marambaud. Role of the proteasome in alzheimer’s disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1502(1):133 – 138, 2000.
- [107] Wei Kong, Xiaoyang Mou, Qingzhong Liu, Zhongxue Chen, Charles R Vanderburg, Jack T Rogers, and Xudong Huang. Independent component analysis of alzheimer’s dna microarray gene expression data. *Molecular neurodegeneration*, 4(1):5, 2009.
- [108] Stephanie Dauth, Thomas Grevesse, Harry Pantazopoulos, Patrick H Campbell, Ben M Maoz, Sabina Berretta, and Kevin Kit Parker. Extracellular matrix protein expression is brain region dependent. *Journal of Comparative Neurology*, 524(7):1309–1336, 2016.
- [109] Manveen K Sethi and Joseph Zaia. Extracellular matrix proteomics in schizophrenia and alzheimer’s disease. *Analytical and bioanalytical chemistry*, 409(2):379–394, 2017.
- [110] Hala Salim Sonbol. Extracellular matrix remodeling in human disease. *Journal of microscopy and ultrastructure*, 6(3):123, 2018.
- [111] Srikanth Rangaraju, Marla Gearing, Lee-Way Jin, and Allan Levey. Potassium channel kv1.3 is highly expressed by microglia in human alzheimer’s disease. *Journal of Alzheimer’s disease*, 44(3):797–808, 2015.
- [112] Laura Thei, Jennifer Imm, Eleni Kaisis, Mark L Dallas, and Talitha L Kerrigan. Microglia in alzheimer’s disease: a role for ion channels. *Frontiers in neuroscience*, 12, 2018.
- [113] Virginie Stygelbout, Karelle Leroy, Valerie Pouillon, Kunie Ando, Eva D’amico, Yonghui Jia, H Robert Luo, Charles Duyckaerts, Christophe Erneux, Stephane Schurmans, et al. Inositol trisphosphate 3-kinase b is increased in human alzheimer brain and exacerbates mouse alzheimer pathology. *Brain*, 137(2):537–552, 2014.
- [114] Hongyan Li and Rong Wang. A focus on cxcr4 in alzheimer’s disease. *Brain circulation*, 3(4):199, 2017.
- [115] Jessica L. Ables, Joshua J. Breunig, Amelia J. Eisch, and Pasko Rakic. Not(ch) just development: Notch signalling in the adult brain. *Nature Reviews Neuroscience*, 12(5):269–283, May 2011.
- [116] Hamed Owanj, Hai Jie Yang, and Zhi Wei Feng. Nucleoside diphosphate kinase nm23-m1 involves in oligodendroglial versus neuronal cell fate decision in vitro. *Differentiation*, 84(4):281–293, Nov 2012.
- [117] Jose Vina and Ana Lloret. Why women have more alzheimer’s disease than men: gender and mitochondrial toxicity of amyloid- β peptide. *Journal of Alzheimer’s disease*, 20(s2):S527–S533, 2010.

- [118] J. L. Podcasy and C. N. Epperson. Considering sex and gender in alzheimer disease and other dementias. *Dialogues Clin Neurosci*, 18(4):437–446, 2016.
- [119] S. Seshadri, P. A. Wolf, A. Beiser, R. Au, K. McNulty, R. White, and R. B. D’Agostino. Lifetime risk of dementia and alzheimer’s disease. the impact of mortality on risk estimates in the framingham study. *Neurology*, 49(6):1498–504, 1997.
- [120] Hongde Liu, Kun Luo, and Donghui Luo. Guanosine monophosphate reductase 1 is a potential therapeutic target for alzheimer’s disease. *Scientific Reports*, 8(1):2759, 2018.
- [121] Luke W Bonham, Celeste M Karch, Chun C Fan, Chin Tan, Ethan G Geier, Yunpeng Wang, Natalie Wen, Iris J Broce, Yi Li, Matthew J Barkovich, et al. Cxcr4 involvement in neurodegenerative diseases. *Translational psychiatry*, 8(1):73, 2018.
- [122] Paola Bezzi, Maria Domercq, Liliana Brambilla, Rossella Galli, Dominique Schols, Erik De Clercq, Angelo Vescovi, Giacinto Bagetta, George Kollias, Jacopo Meldolesi, and et al. Cxcr4-activated astrocyte glutamate release via tnfa: amplification by microglia triggers neurotoxicity. *Nature Neuroscience*, 4(7):702–710, Jul 2001.
- [123] B. Zhang, C. Gaiteri, L. G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezhnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin, E. Fluder, B. Clurman, S. Melquist, M. Narayanan, C. Suver, H. Shah, M. Mahajan, T. Gillis, J. Mysore, M. E. MacDonald, J. R. Lamb, D. A. Bennett, C. Molony, D. J. Stone, V. Gudnason, A. J. Myers, E. E. Schadt, H. Neumann, J. Zhu, and V. Emilsson. Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer’s disease. *Cell*, 153(3):707–20, 2013.
- [124] B. Antonsson, D. B. Kassel, G. Di Paolo, R. Lutjens, B. M. Riederer, and G. Grenningloh. Identification of in vitro phosphorylation sites in the growth cone protein scg10. effect of phosphorylation site mutants on microtubule-destabilizing activity. *J Biol Chem*, 273(14):8439–46, 1998.
- [125] C. Chiellini, G. Grenningloh, O. Cochet, M. Scheideler, Z. Trajanoski, G. Ailhaud, C. Dani, and E. Z. Amri. Stathmin-like 2, a developmentally-associated neuronal marker, is expressed and modulated during osteogenesis of human mesenchymal stem cells. *Biochem Biophys Res Commun*, 374(1):64–8, 2008.
- [126] M. Solarski, H. Wang, H. Wille, and G. Schmitt-Ulms. Somatostatin in alzheimer’s disease: A new role for an old player. *Prion*, 12(1):1–8, 2018.
- [127] E. Hama and T. C. Saido. Etiology of sporadic alzheimer’s disease: somatostatin, neprilysin, and amyloid beta peptide. *Med Hypotheses*, 65(3):498–500, 2005.
- [128] Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May, and et al. The reactome pathway knowledgebase. *Nucleic Acids Research*, 46(D1):D649–D655, Nov 2017.
- [129] Ling-jie He, Nan Liu, Tian-lin Cheng, Xiao-jing Chen, Yi-ding Li, You-sheng Shu, Zi-long Qiu, and Xiao-hui Zhang. Conditional deletion of mecp2 in parvalbumin-expressing

- gabaergic cells results in the absence of critical period plasticity. *Nature Communications*, 5(1), Oct 2014.
- [130] Masaru Ishii and Yoshihisa Kurachi. Muscarinic acetylcholine receptors. *Current Pharmaceutical Design*, 12(28):3573–3581, Oct 2006.
- [131] Purves D. *Neuroscience*. Sinauer Associates, 2001. Available at <https://www.ncbi.nlm.nih.gov/books/NBK10799/>.
- [132] A. Limon, J. M. Reyes-Ruiz, and R. Miledi. Loss of functional gabaa receptors in the alzheimer diseased brain. *Proceedings of the National Academy of Sciences*, 109(25):10071–10076, Jun 2012.
- [133] Michaeline Hebron, Margo Peyton, Xiaoguang Liu, Xiaokong Gao, Ruochong Wang, Irina Lonskaya, and Charbel E.-H. Moussa. Discoidin domain receptor inhibition reduces neuropathology and attenuates inflammation in neurodegeneration models. *Journal of Neuroimmunology*, 311:1–9, Oct 2017.
- [134] Paolina Crocco, Adolfo Saiardi, Miranda S. Wilson, Raffaele Maletta, Amalia C. Bruni, Giuseppe Passarino, and Giuseppina Rose. Contribution of polymorphic variation of inositol hexakisphosphate kinase 3 (ip6k3) gene promoter to the susceptibility to late onset alzheimer’s disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1862(9):1766–1773, Sep 2016.
- [135] Yuji Kajiwarra, Erming Wang, Minghui Wang, Wun Chey Sin, Kristen J. Brennand, Eric Schadt, Christian C. Naus, Joseph Buxbaum, and Bin Zhang. Gja1 (connexin43) is a key regulator of alzheimer’s disease pathogenesis. *Acta Neuropathologica Communications*, 6(1), Dec 2018.
- [136] Ryoko Yoshida, Giichi Takaesu, Hideyuki Yoshida, Fuyuki Okamoto, Tomoko Yoshioka, Yongwon Choi, Shizuo Akira, Taro Kawai, Akihiko Yoshimura, and Takashi Kobayashi. Traf6 and mekk1 play a pivotal role in the rig-i-like helicase antiviral pathway. *Journal of Biological Chemistry*, 283(52):36211–36220, Nov 2008.
- [137] Shan Shan Li, Zhengdong Qu, Matilda Haas, Linh Ngo, You Jeong Heo, Hyo Jung Kang, Joanne Maria Britto, Hayley Daniella Cullen, Hannah Kate Vanyai, Seong-Seng Tan, and et al. The hsa21 gene eurl/c21orf91 controls neurogenesis within the cerebral cortex and is implicated in the pathogenesis of down syndrome. *Scientific Reports*, 6(1), Jul 2016.
- [138] M. W. Pitts, P. M. Kremer, A. C. Hashimoto, D. J. Torres, C. N. Byrns, C. S. Williams, and M. J. Berry. Competition between the brain and testes under selenium-compromised conditions: Insight into sex differences in selenium metabolism and risk of neurodevelopmental disease. *Journal of Neuroscience*, 35(46):15326–15338, Nov 2015.
- [139] Elisabetta Menna, Andrea Disanza, Cinzia Cagnoli, Ursula Schenk, Giuliana Gelsomino, Emanuela Frittoli, Maud Hertzog, Nina Offenhauser, Corinna Sawallisch, Hans-Jürgen Kreienkamp, and et al. Correction: Eps8 regulates axonal filopodia in hippocampal neurons in response to brain-derived neurotrophic factor (bdnf). *PLOS Biology*, 13(6):e1002184, Jun 2015.

- [140] Xiuling Nie, Yu Sun, Suiwen Wan, Hui Zhao, Renyuan Liu, Xueping Li, Sichu Wu, Zuzana Nedelska, Jakub Hort, Zhao Qing, and et al. Subregional structural alterations in hippocampus and nucleus accumbens correlate with the clinical impairment in patients with alzheimer's disease clinical spectrum: Parallel combining volume and vertex-based approach. *Frontiers in Neurology*, 8, Aug 2017.
- [141] L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. J. Westendorp, E. L. E. M. Bollen, P. W. de Bruin, H. A. M. Middelkoop, M. A. van Buchem, and J. van der Grond. Strongly reduced volumes of putamen and thalamus in alzheimer's disease: an mri study. *Brain*, 131(12):3277–3285, Nov 2008.
- [142] Annalisa Nobili, Emanuele Claudio Latagliata, Maria Teresa Viscomi, Virve Cavallucci, Debora Cutuli, Giacomo Giacobazzo, Paraskevi Krashia, Francesca Romana Rizzo, Ramona Marino, Mauro Federici, et al. Dopamine neuronal loss contributes to memory and reward dysfunction in a model of alzheimer's disease. *Nature communications*, 8:14727, 2017.
- [143] Alessandro Martorana and Giacomo Koch. Is dopamine involved in alzheimer's disease? *Frontiers in aging neuroscience*, 6:252, 2014.
- [144] Cathra Halabi, Anasheh Halabi, David L Dean, Pei-Ning Wang, Adam L Boxer, John Q Trojanowski, Stephen J DeArmond, Bruce L Miller, Joel H Kramer, and William W Seeley. Patterns of striatal degeneration in frontotemporal dementia. *Alzheimer disease and associated disorders*, 27(1):74, 2013.
- [145] Alyssa A Brewer and Brian Barton. Visual cortex in aging and alzheimer's disease: changes in visual field maps and population receptive fields. *Frontiers in psychology*, 5:74, 2014.
- [146] Centers for Disease Control and Prevention. U.s. burden of alzheimer's disease, related dementias to double by 2060, 2018. Accessed: 2018-12-01.
- [147] Kyle Steenland, Felicia C. Goldstein, Allan Levey, and Whitney Wharton. A meta-analysis of alzheimer's disease incidence and prevalence comparing african-americans and caucasians. *Journal of Alzheimer's disease : JAD*, 50(1):71–76, 2016.
- [148] Tim Stuart and Rahul Satija. Integrative single-cell analysis. *Nature Reviews Genetics*, page 1, 2019.
- [149] Daojing Wang and Steven Bodovitz. Single cell analysis: the new frontier in 'omics'. *Trends in biotechnology*, 28(6):281–290, 2010.
- [150] Sid E. O'Bryant. Introduction to special issue on advances in blood-based biomarkers of alzheimer's disease. *Alzheimer's & dementia (Amsterdam, Netherlands)*, 3:110–112, 2016.
- [151] Lionel A Mandell, Richard G Wunderink, Antonio Anzueto, John G Bartlett, G Douglas Campbell, Nathan C Dean, Scott F Dowell, Thomas M File, Daniel M Musher, and Michael S Niederman. Infectious diseases society of america/american thoracic society consensus guidelines on the management of community-acquired pneumonia in adults. *Clinical infectious diseases*, 44(Supplement 2):S27–S72, 2007.

- [152] N. D. Wolfe, C. P. Dunavan, and J. Diamond. Origins of major human infectious diseases. *Nature*, 447(7142):279–83, 2007.
- [153] H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, and J. Bartlett. Bad bugs, no drugs: no escape! an update from the infectious diseases society of america. *Clin Infect Dis*, 48(1):1–12, 2009.
- [154] Mayo Clinic Staff. Infectious diseases. 2016.
- [155] Purushothama V Dasaraju and Chien Liu. Infections of the respiratory system. *Medical microbiology*, 4, 1996.
- [156] Christopher Troeger, Mohammad Forouzanfar, Puja C. Rao, Ibrahim Khalil, Alexandria Brown, Scott Swartz, Nancy Fullman, Jonathan Mosser, Robert L. Thompson, Jr. Reiner, Robert C., Amanuel Abajobir, Noore Alam, Mulubirhan Assefa Alemayohu, Azmeraw T. Amare, Carl Abelardo Antonio, Hamid Asayesh, Euripide Avokpaho, Aleksandra Barac, Muktar A. Beshir, Dube Jara Boneya, Michael Brauer, Lalit Dandona, Rakhi Dandona, Joseph R. A. Fitchett, Tsegaye Tewelde Gebrehiwot, Gessesew Buggsa Hailu, Peter J. Hotez, Amir Kasaeian, Tawfik Khoja, Niranjana Kissoon, Luke Knibbs, G. Anil Kumar, Rajesh Kumar Rai, Hassan Magdy Abd El Razek, Muktar S. K. Mohammed, Katie Nielson, Eyal Oren, Abdalla Osman, George Patton, Mostafa Qorbani, Hirbo Shore Roba, Benn Sartorius, Miloje Savic, Mika Shigematsu, Bryan Sykes, Soumya Swaminathan, Roman Topor-Madry, Kingsley Ukwaja, Andrea Werdecker, Naohiro Yonemoto, Maysaa El Sayed Zaki, Stephen S. Lim, Mohsen Naghavi, Theo Vos, Simon I. Hay, Christopher J. L. Murray, and Ali H. Mokdad. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the global burden of disease study 2015. *The Lancet Infectious Diseases*, 17(11):1133–1161, 2017.
- [157] T. Bandaranayake and A. C. Shaw. Host resistance and immune aging. *Clinics and Geriatric Medicine*, 32(3):415–32, 2016.
- [158] D. Bogaert, D. Weinberger, C. Thompson, M. Lipsitch, and R. Malley. Impaired innate and adaptive immunity to streptococcus pneumoniae and its effect on colonization in an infant mouse model. *Infect Immun*, 77(4):1613–22, 2009.
- [159] J. G. Burel, S. H. Apte, and D. L. Doolan. Systems approaches towards molecular profiling of human immunity. *Trends Immunol*, 37(1):53–67, 2016.
- [160] C. Castelo-Branco and I. Soveral. The immune system and aging: a review. *Gynecological Endocrinology*, 30(1):16–22, 2014.
- [161] M. Pinti, V. Appay, J. Campisi, D. Frasca, T. Fulop, D. Sauce, A. Larbi, B. Weinberger, and A. Cossarizza. Aging of the immune system - focus on inflammation and vaccination. *Eur J Immunol*, 2016.
- [162] D. Aw, A. B. Silva, and D. B. Palmer. Immunosenescence: emerging challenges for an ageing population. *Immunology*, 120(4):435–46, 2007.

- [163] C. Caruso, S. Buffa, G. Candore, G. Colonna-Romano, D. Dunn-Walters, D. Kipling, and G. Pawelec. Mechanisms of immunosenescence. *Immun Ageing*, 6:10, 2009.
- [164] G. K. Paterson and T. J. Mitchell. Innate immunity and the pneumococcus. *Microbiology*, 152(Pt 2):285–93, 2006.
- [165] Aras Kadioglu, Jeffrey N Weiser, James C Paton, and Peter W Andrew. The role of streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nature Reviews Microbiology*, 6(4):288–301, 2008.
- [166] Centers for Disease Control and Prevention. Manual for the surveillance of vaccine-preventable diseases. 2014.
- [167] Centers for Disease Control Prevention. Active bacterial core surveillance reports. 2016.
- [168] Centers for Disease Control and Prevention. Active bacterial core surveillance report, emerging infections program network, streptococcus pneumoniae. 2015.
- [169] Centers for Disease Control and Prevention. Abcs report: Streptococcus pneumoniae. Report, 2010.
- [170] D. Bogaert, R. de Groot, and P. W. M. Hermans. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *The Lancet Infectious Diseases*, 4(3):144–154, 2004.
- [171] Tom van der Poll and Steven M Opal. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet*, 374:1543–1556, 2009.
- [172] Centers for Disease Control and Prevention. Prevention of pneumococcal disease: Recommendations of the advisory committee on immunization practices (acip). *MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports*, 46(RR-08):1–24, 1997.
- [173] Centers for Disease Control and Prevention. Pneumococcal disease. 2015.
- [174] S. Blumental, A. Granger-Farbos, J. C. Moisi, B. Soullie, P. Leroy, B. M. Njanpop-Lafourcade, S. Yaro, B. Nacro, M. Hallin, and J. L. Koeck. Virulence factors of streptococcus pneumoniae. comparison between african and french invasive isolates and implication for future vaccines. *PLoS One*, 10(7):e0133885, 2015.
- [175] Charles Feldman and Ronald Anderson. Epidemiology, virulence factors and management of pneumococcus. *F1000Research*, 5(F1000 Faculty Rev)(2320), 2016.
- [176] M. J. Jedrzejewski. Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev*, 65(2):187–207 ; first page, table of contents, 2001.
- [177] Anthony J. Infante, Jonathan A. McCullers, and Carlos J. Orihuela. *Chapter 19 - Mechanisms of Predisposition to Pneumonia: Infants, the Elderly, and Viral Infections*, pages 363–382. Academic Press, Amsterdam, 2015.
- [178] Centers for Disease Control and Prevention. Fast stats: Leading causes of death. 2013.

- [179] Centers for Disease Control and Prevention. Influenza (flu): Flu vaccines and preventing flu illness. 2016.
- [180] American Lung Association. Trends in pneumonia and influenza morbidity and mortality. Report, 2015.
- [181] Angela E. Bridy-Pappas, Marya B. Margolis, Kimberly J. Center, and Daniel J. Isaacman. Streptococcus pneumoniae: Description of the pathogen, disease epidemiology, treatment, and prevention. *The Journal of Human Pharmacology and Drug Therapy*, 25(9):1193–1212, 2005.
- [182] Centers for Disease Control and Prevention. Epidemiology and prevention of vaccine-preventable diseases. *Washington DC: Public Health Foundation*, 12, 2011.
- [183] K Aaron Geno, Gwendolyn L Gilbert, Joon Young Song, Ian C Skovsted, Keith P Klugman, Christopher Jones, Helle B Konradsen, and Moon H Nahm. Pneumococcal capsules and their types: past, present, and future. *Clinical microbiology reviews*, 28(3):871–899, 2015.
- [184] Benjamin White, Elliott Stirling Robinson, and Laverne Almon Barnes. *The biology of pneumococcus: the bacteriological, biochemical, and immunological characters and activities of Diplococcus pneumoniae*, volume 12. Harvard University Press, 1938.
- [185] World Health Organization. 23-valent pneumococcal polysaccharide vaccine: Who position paper. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, 83(42):373–384, 2008.
- [186] Naomi Sugimoto, Yuka Yamagishi, Jun Hirai, Daisuke Sakanashi, Hiroyuki Suematsu, Naoya Nishiyama, Yusuke Koizumi, and Hiroshige Mikamo. Invasive pneumococcal disease caused by mucoid serotype 3 streptococcus pneumoniae: a case report and literature review. *BMC research notes*, 10(1):21, 2017.
- [187] Lance E Keller, D Ashley Robinson, and Larry S McDaniel. Nonencapsulated streptococcus pneumoniae: emergence and pathogenesis. *MBio*, 7(2):e01792–15, 2016.
- [188] M. Hackel, C. Lascols, S. Bouchillon, B. Hilton, D. Morgenstern, and J. Purdy. Serotype prevalence and antibiotic resistance in streptococcus pneumoniae clinical isolates among global populations. *Vaccine*, 31(42):4881–7, 2013.
- [189] D. J. Isaacman, E. D. McIntosh, and R. R. Reinert. Burden of invasive pneumococcal disease and serotype distribution among streptococcus pneumoniae isolates in young children in europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines. *International Journal of Infectious Diseases*, 14(3):e197–209, 2010.
- [190] E. Varon. Epidemiology of streptococcus pneumoniae. *Médecine et Maladies Infectieuses*, 42(8):361–5, 2012.
- [191] P. Chavanet. Pneumococcus infections: is the burden still as heavy? *Med Mal Infect*, 42(4):149–53, 2012.

- [192] P. C. Wroe, J. A. Finkelstein, G. T. Ray, J. A. Linder, K. M. Johnson, S. Rifas-Shiman, M. R. Moore, and S. S. Huang. Aging population and future burden of pneumococcal pneumonia in the united states. *Journal of Infectious Diseases*, 205(10):1589–92, 2012.
- [193] A. Torres, F. Blasi, N. Dartois, and M. Akova. Which individuals are at increased risk of pneumococcal disease and why? impact of copd, asthma, smoking, diabetes, and/or chronic heart disease on community-acquired pneumonia and invasive pneumococcal disease. *Thorax*, 70(10):984–9, 2015.
- [194] Ji A. Jung, Hirohito Kita, Barbara P. Yawn, Thomas G. Boyce, Kwang H. Yoo, Michaela E. McGree, Amy L. Weaver, Peter Wollan, Robert M. Jacobson, and Young J. Juhn. Increased risk of serious pneumococcal disease in patients with atopic conditions other than asthma. *The Journal of allergy and clinical immunology*, 125(1):217–221, 2010.
- [195] Hongxia Zhao, Cheol-In Kang, Mark S. Rouse, Robin Patel, Hirohito Kita, and Young J. Juhn. The role of il-17 in the association between pneumococcal pneumonia and allergic sensitization. *International Journal of Microbiology*, 2011:709509, 2011.
- [196] Y. Chao, L. R. Marks, M. M. Pettigrew, and A. P. Hakansson. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. *Front Cell Infect Microbiol*, 4:194, 2014.
- [197] Richard A. Adegbola, Rodrigo DeAntonio, Philip C. Hill, Anna Roca, Effua Usuf, Bernard Hoet, and Brian M. Greenwood. Carriage of streptococcus pneumoniae and other respiratory bacterial pathogens in low and lower-middle income countries: A systematic review and meta-analysis. *PLOS ONE*, 9(8):e103293, 2014.
- [198] Izabela Korona-Glowniak and Anna Malm. Characteristics of streptococcus pneumoniae strains colonizing upper respiratory tract of healthy preschool children in poland. *The Scientific World Journal*, 2012, 2012.
- [199] Laura R. Marks, G. Iyer Parameswaran, and Anders P. Hakansson. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. *Infection and Immunity*, 80(8):2744–2760, 2012.
- [200] H. C. Steel, R. Cockeran, R. Anderson, and C. Feldman. Overview of community-acquired pneumonia and the role of inflammatory mechanisms in the immunopathogenesis of severe pneumococcal disease. *Mediators of Inflammation*, 2013:490346, 2013.
- [201] Mayo Clinic Staff. Meningitis. 2016.
- [202] Centers for Disease Control and Prevention. Bacterial meningitis. 2017.
- [203] A. M. Mitchell and T. J. Mitchell. Streptococcus pneumoniae: virulence factors and variation. *Clin Microbiol Infect*, 16(5):411–8, 2010.
- [204] C. Hyams, E. Camberlein, J. M. Cohen, K. Bax, and J. S. Brown. The streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun*, 78(2):704–15, 2010.

- [205] A. Rabes, N. Suttorp, and B. Opitz. Inflammasomes in pneumococcal infection: Innate immune sensing and bacterial evasion strategies. *Curr Top Microbiol Immunol*, 397:215–27, 2016.
- [206] Aaron L Nelson, Aoife M Roche, Jane M Gould, Kannie Chim, Adam J Ratner, and Jeffrey N Weiser. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infection and immunity*, 75(1):83–90, 2007.
- [207] M. N. Khan and M. E. Pichichero. The host immune dynamics of pneumococcal colonization: implications for novel vaccine development. *Humans Vaccines and Immunotherapeutics*, 10(12):3688–99, 2015.
- [208] M. A. Zafar, Y. Wang, S. Hamaguchi, and J. N. Weiser. Host-to-host transmission of streptococcus pneumoniae is driven by its inflammatory toxin, pneumolysin. *Cell Host Microbe*, 21(1):73–83, 2017.
- [209] L. E. Keller, J. L. Bradshaw, H. Pipkins, and L. S. McDaniel. Surface proteins and pneumolysin of encapsulated and nonencapsulated streptococcus pneumoniae mediate virulence in a chinchilla model of otitis media. *Frontiers in Cellular and Infection Microbiology*, 6:55, 2016.
- [210] M. Hotomi, J. Yuasa, D. E. Briles, and N. Yamanaka. Pneumolysin plays a key role at the initial step of establishing pneumococcal nasal colonization. *Folia Microbiol (Praha)*, 61(5):375–83, 2016.
- [211] J. B. Rubins and E. N. Janoff. Pneumolysin: a multifunctional pneumococcal virulence factor. *Journal of Laboratory and Clinical Medicine*, 131(1):21–7, 1998.
- [212] J. R. Shak, H. P. Ludewick, K. E. Howery, F. Sakai, H. Yi, R. M. Harvey, J. C. Paton, K. P. Klugman, and J. E. Vidal. Novel role for the streptococcus pneumoniae toxin pneumolysin in the assembly of biofilms. *MBio*, 4(5):e00655–13, 2013.
- [213] P. Rai, F. He, J. Kwang, B. P. Engelward, and V. T. Chow. Pneumococcal pneumolysin induces dna damage and cell cycle arrest. *Sci Rep*, 6:22972, 2016.
- [214] D. A Watson, D.M Musher, and J. Verhoef. Pneumococcal virulence factors and host immune responses to them. *European Journal of Clinical Microbiology and Infectious Diseases*, 14(6):479–490, 1995.
- [215] J. C. Paton, R. A. Lock, and D. J. Hansman. Effect of immunization with pneumolysin on survival time of mice challenged with streptococcus pneumoniae. *Infect Immun*, 40(2):548–52, 1983.
- [216] C. C. Kietzman, G. Gao, B. Mann, L. Myers, and E. I. Tuomanen. Dynamic capsule restructuring by the main pneumococcal autolysin lytA in response to the epithelium. *Nat Commun*, 7:10859, 2016.
- [217] G. D. Shockman, L. Daneo-Moore, R. Kariyama, and O. Massidda. Bacterial walls, peptidoglycan hydrolases, autolysins, and autolysis. *Microb Drug Resist*, 2(1):95–8, 1996.

- [218] A. Tomasz. The role of autolysins in cell death. *Ann N Y Acad Sci*, 235(0):439–47, 1974.
- [219] R. Lopez, P. Garcia, and E. Garcia. [bacterial autolysins]. *Microbiol Esp*, 36(1-2):45–57, 1983.
- [220] Melanie Abeyta, Gail G. Hardy, and Janet Yother. Genetic alteration of capsule type but not pspa type affects accessibility of surface-bound complement and surface antigens of streptococcus pneumoniae. *Infection and Immunity*, 71(1):218–225, 2003.
- [221] S. Bergmann and S. Hammerschmidt. Versatility of pneumococcal surface proteins. *Microbiology*, 152(Pt 2):295–303, 2006.
- [222] A. H. Tu, R. L. Fulgham, M. A. McCrory, D. E. Briles, and A. J. Szalai. Pneumococcal surface protein a inhibits complement activation by streptococcus pneumoniae. *Infect Immun*, 67(9):4720–4, 1999.
- [223] Gowrisankar Rajem, Julie M. Anderton, George M. Carlone, Jacquelyn S. Sampson, and Edwin W. Ades. Pneumococcal surface adhesin a (psaa): A review. *Critical Reviews in Microbiology*, 34:131–142, 2008.
- [224] F. Iannelli, D. Chiavolini, S. Ricci, M. R. Oggioni, and G. Pozzi. Pneumococcal surface protein c contributes to sepsis caused by streptococcus pneumoniae in mice. *Infect Immun*, 72(5):3077–80, 2004.
- [225] Khoosheh K Gosink, Elizabeth R Mann, Chris Guglielmo, Elaine I Tuomanen, and H Robert Masure. Role of novel choline binding proteins in virulence of streptococcus pneumoniae. *Infection and immunity*, 68(10):5690–5695, 2000.
- [226] B. Maestro and J. M. Sanz. Choline binding proteins from streptococcus pneumoniae: A dual role as enzybiotics and targets for the design of new antimicrobials. *Antibiotics (Basel)*, 5(2), 2016.
- [227] Brian Henderson. *Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections*. John Wiley and Sons, 2017.
- [228] Rémi Terrasse, Ana Amoroso, Thierry Vernet, and Anne Marie Di Guilmi. Streptococcus pneumoniae gapdh is released by cell lysis and interacts with peptidoglycan. *PLoS ONE*, 10(4):e0125377, 2015.
- [229] Marcus Fulde and Simone Bergmann. Impact of streptococcal enolase in virulence. *Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections*, pages 245–268, 2017.
- [230] M. A. Barocchi, J. Ries, X. Zogaj, C. Hemsley, B. Albiger, A. Kanth, S. Dahlberg, J. Fernebro, M. Moschioni, V. Masignani, K. Hultenby, A. R. Taddei, K. Beiter, F. Wartha, A. von Euler, A. Covacci, D. W. Holden, S. Normark, R. Rappuoli, and B. Henriques-Normark. A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci U S A*, 103(8):2857–62, 2006.

- [231] L. Pancotto, G. De Angelis, E. Bizzarri, M. A. Barocchi, G. Del Giudice, M. Moschioni, and P. Ruggiero. Expression of the streptococcus pneumoniae pilus-1 undergoes on and off switching during colonization in mice. *Nature Scientific Reports*, 3:2040, 2013.
- [232] E. N. Janoff, J. B. Rubins, C. Fasching, D. Charboneau, J. T. Rahkola, A. G. Plaut, and J. N. Weiser. Pneumococcal iga1 protease subverts specific protection by human iga1. *Mucosal Immunol*, 7(2):249–56, 2014.
- [233] M. Proctor and P. J. Manning. Production of immunoglobulin a protease by streptococcus pneumoniae from animals. *Infect Immun*, 58(9):2733–7, 1990.
- [234] Y. C. Chi, J. T. Rahkola, A. A. Kendrick, M. J. Holliday, N. Paukovich, T. S. Roberts, E. N. Janoff, and E. Z. Eisenmesser. Streptococcus pneumoniae iga1 protease: A metalloprotease that can catalyze in a split manner in vitro. *Protein Sci*, 26(3):600–610, 2017.
- [235] R. G. Wunderink and G. W. Waterer. Clinical practice. community-acquired pneumonia. *New England Journal of Medicine*, 370(6):543–51, 2014.
- [236] Melonie P Heron. Deaths: leading causes for 2012. 2015.
- [237] National Heart, Lung, and Blood Institute. Pneumonia. 2017, 2016.
- [238] World Health Organization. World pneumonia day. 2017.
- [239] Centers for Disease Control and Prevention. Pneumococcal disease: Fast facts. 2017.
- [240] United Nations Children Fund. Pneumonia claims the lives of the world’s most vulnerable children. 2018.
- [241] Antonella F Simonetti, Diego Viasus, Carolina Garcia-Vidal, and Jordi Carratalà. Management of community-acquired pneumonia in older adults. *Therapeutic advances in infectious disease*, 2(1):3–16, 2014.
- [242] John E Stupka, Eric M Mortensen, Antonio Anzueto, and Marcos I Restrepo. Community-acquired pneumonia in elderly patients. *Aging health*, 5(6):763–774, 2009.
- [243] World Health Organization. Fact sheet: Children: Reducing mortality. 2016.
- [244] World Health Organization. Pneumonia - key facts. 2016.
- [245] Olivier le Polain de Waroux, Stefan Flasche, Adam Kucharski, Celine Langendorf, Donny Ndazima, Juliet Mwanga-Amumpaire, Rebecca Grais, Sandra Cohuet, and W. John Edmunds. Identifying human encounters that shape the transmission of streptococcus pneumoniae and other respiratory infections. *bioRxiv*, 2017.
- [246] BM Althouse, LL Hammitt, L Grant, BG Wagner, R Reid, F Larzelere-Hinton, R Weatherholtz, KP Klugman, GL Rodgers, and KL O’Brien. Identifying transmission routes of streptococcus pneumoniae and sources of acquisitions in high transmission communities. *Epidemiology and Infection*, 145(13):2750–2758, 2017.

- [247] Raquel Sá-Leão, Sónia Nunes, António Brito-Avão, Carla R Alves, João A Carriço, Joana Saldanha, Jonas S Almeida, Ilda Santos-Sanches, and Hermíniade de Lencastre. High rates of transmission of and colonization by *streptococcus pneumoniae* and *haemophilus influenzae* within a day care center revealed in a longitudinal study. *Journal of clinical microbiology*, 46(1):225–234, 2008.
- [248] Marc Lipsitch, Osman Abdullahi, Alexander D’Amour, Wen Xie, Daniel M Weinberger, Eric Tchetgen Tchetgen, and J Anthony G Scott. Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in kenya with a markov transition model. *Epidemiology (Cambridge, Mass.)*, 23(4):510, 2012.
- [249] Jonathan F Mosser, Lindsay R Grant, Eugene V Millar, Robert C Weatherholtz, Delois M Jackson, Bernard Beall, Mariddie J Craig, Raymond Reid, Mathuram Santosham, and Katherine L O’Brien. Nasopharyngeal carriage and transmission of *streptococcus pneumoniae* in american indian households after a decade of pneumococcal conjugate vaccine use. *PloS one*, 9(1):e79578, 2014.
- [250] Laura R Marks, Ryan M Reddinger, and Anders P Hakansson. Biofilm formation enhances fomite survival of *streptococcus pneumoniae* and *streptococcus pyogenes*. *Infection and immunity*, 82(3):1141–1146, 2014.
- [251] R. N. Allan, P. Skipp, J. Jefferies, S. C. Clarke, S. N. Faust, L. Hall-Stoodley, and J. Webb. Pronounced metabolic changes in adaptation to biofilm growth by *streptococcus pneumoniae*. *PLoS One*, 9(9):e107015, 2014.
- [252] Rebecca L Walsh and Andrew Camilli. *Streptococcus pneumoniae* is desiccation tolerant and infectious upon rehydration. *MBio*, 2(3):e00092–11, 2011.
- [253] B. M. Davis, A. E. Aiello, S. Dawid, P. Rohani, S. Shrestha, and B. Foxman. Influenza and community-acquired pneumonia interactions: the impact of order and time of infection on population patterns. *Am J Epidemiol*, 175(5):363–7, 2012.
- [254] Eili Y. Klein, Bradley Monteforte, Alisha Gupta, Wendi Jiang, Larissa May, Yu-Hsiang Hsieh, and Andrea Dugas. The frequency of influenza and bacterial coinfection: a systematic review and meta-analysis. *Influenza and Other Respiratory Viruses*, 10(5):394–403, 2016.
- [255] John F Brundage and G Dennis Shanks. Deaths from bacterial pneumonia during 1918–19 influenza pandemic. *Emerging infectious diseases*, 14(8):1193, 2008.
- [256] Yu-Wen Chien, Keith P Klugman, and David M Morens. Bacterial pathogens and death during the 1918 influenza pandemic. *New England Journal of Medicine*, 361(26):2582–2583, 2009.
- [257] Lauren O. Bakaletz. Viral–bacterial co-infections in the respiratory tract. *Current Opinion in Microbiology*, 35(Supplement C):30–35, 2017.
- [258] Linda S. Cauley and Anthony T. Vella. Why is co-infection with influenza virus and bacteria so difficult to control? *Discovery medicine*, 19(102):33–40, 2015.

- [259] Matthew P. Crotty, Shelby Meyers, Nicholas Hampton, Stephanie Bledsoe, David J. Ritchie, Richard S. Buller, Gregory A. Storch, Scott T. Micek, and Marin H. Kollef. Epidemiology, co-infections, and outcomes of viral pneumonia in adults: An observational cohort study. *Medicine*, 94(50):e2332, 2015.
- [260] J. Hoffmann, D. Machado, O. Terrier, S. Pouzol, M. Messaoudi, W. Basualdo, E. E. Espinola, R. M. Guillen, M. Rosa-Calatrava, V. Picot, T. Benet, H. Endtz, G. Russomando, and G. Paranhos-Baccala. Viral and bacterial co-infection in severe pneumonia triggers innate immune responses and specifically enhances ip-10: a translational study. *Nature Scientific Reports*, 6:38532, 2016.
- [261] Huong Thi Thu Vu, Lay Myint Yoshida, Motoi Suzuki, Hien Anh Thi Nguyen, Cat Dinh Lien Nguyen, Ai Thi Thuy Nguyen, Kengo Oishi, Takeshi Yamamoto, Kiwao Watanabe, and Thiem Dinh Vu. Association between nasopharyngeal load of streptococcus pneumoniae, viral coinfection, and radiologically confirmed pneumonia in vietnamese children. *The Pediatric infectious disease journal*, 30(1):11–18, 2011.
- [262] M. N. Khan, Q. Xu, and M. E. Pichichero. Protection against streptococcus pneumoniae invasive pathogenesis by a protein-based vaccine is achieved by suppression of nasopharyngeal bacterial density during influenza a virus coinfection. *Infect Immun*, 85(2), 2017.
- [263] Jaelle C Brealey, Keith J Chappell, Sally Galbraith, Emmanuelle Fantino, Jane Gaydon, Sarah Tozer, Paul R Young, Patrick G Holt, and Peter D Sly. Streptococcus pneumoniae colonization of the nasopharynx is associated with increased severity during respiratory syncytial virus infection in young children. *Respirology*, 2017.
- [264] M. T. Goncalves, T. J. Mitchell, and J. M. Lord. Immune ageing and susceptibility to streptococcus pneumoniae. *Biogerontology*, 17(3):449–65, 2016.
- [265] Hervé Tettelin, Scott Chancey, Tim Mitchell, Dalia Denapate, Yvonne Schähle, Martin Rieger, and Regine Hakenbeck. *Chapter 5 - Genomics, Genetic Variation, and Regions of Differences A2 - Brown, Jeremy*, pages 81–107. Academic Press, Amsterdam, 2015.
- [266] Claudio Donati, N. Luisa Hiller, Hervé Tettelin, Alessandro Muzzi, Nicholas J. Croucher, Samuel V. Angiuoli, Marco Oggioni, Julie C. Dunning Hotopp, Fen Z. Hu, David R. Riley, Antonello Covacci, Tim J. Mitchell, Stephen D. Bentley, Morgens Kilian, Garth D. Ehrlich, Rino Rappuoli, E. Richard Moxon, and Vega Masignani. Structure and dynamics of the pan-genome of streptococcus pneumoniae and closely related species. *Genome Biology*, 11(10):R107–R107, 2010.
- [267] N. Luisa Hiller, Azad Ahmed, Evan Powell, Darren P. Martin, Rory Eutsey, Josh Earl, Benjamin Janto, Robert J. Boissy, Justin Hogg, Karen Barbadora, Rangarajan Sampath, Shaun Lonergan, J. Christopher Post, Fen Z. Hu, and Garth D. Ehrlich. Generation of genic diversity among streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLOS Pathogens*, 6(9):e1001108, 2010.
- [268] Jeffrey N Weiser, Daniela M Ferreira, and James C Paton. Streptococcus pneumoniae: transmission, colonization and invasion. *Nature Reviews Microbiology*, page 1, 2018.

- [269] Jean-Pierre Claverys, Marc Prudhomme, and Bernard Martin. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annual review of microbiology*, 60, 2006.
- [270] John A Lees, Nicholas J. Croucher, David Goldblatt, Francois Nosten, Julian Parkhill, Claudia Turner, Paul Turner, and Stephen D. Bentley. Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration. *Elife*, 2017.
- [271] Florian Wartha, Katharina Beiter, Barbara Albiger, Jenny Fernebro, Arturo Zychlinsky, Staffan Normark, and Birgitta Henriques-Normark. Capsule and d-alanylated lipoteichoic acids protect streptococcus pneumoniae against neutrophil extracellular traps. *Cellular Microbiology*, 9(5):11162–11171, 2007.
- [272] G. L. Rodgers, A. Arguedas, R. Cohen, and R. Dagan. Global serotype distribution among streptococcus pneumoniae isolates causing otitis media in children: potential implications for pneumococcal conjugate vaccines. *Vaccine*, 27(29):3802–10, 2009.
- [273] R. Wilson, J. M. Cohen, R. J. Jose, C. de Vogel, H. Baxendale, and J. S. Brown. Protection against streptococcus pneumoniae lung infection after nasopharyngeal colonization requires both humoral and cellular immune responses. *Mucosal Immunol*, 8(3):627–39, 2015.
- [274] Musher Daniel M. Pneumococcal vaccination in adults. *UpToDate*, 2016, 2015.
- [275] Yuan Li, Daniel M Weinberger, Claudette M Thompson, Krzysztof Trzciński, and Marc Lipsitch. Surface charge of streptococcus pneumoniae predicts serotype distribution. *Infection and immunity*, 81(12):4519–4524, 2013.
- [276] Bin Chang, Akiyoshi Nariai, Tsuyoshi Sekizuka, Yukihiro Akeda, Makoto Kuroda, Kazunori Oishi, and Makoto Ohnishi. Capsule switching and antimicrobial resistance acquired during repeated streptococcus pneumoniae pneumonia episodes. *Journal of Clinical Microbiology*, 53(10):3318–3324, 2015.
- [277] Kelly L. Wyres, Lotte M. Lambertsen, Nicholas J. Croucher, Lesley McGee, Anne von Gottberg, Josefina Liñares, Michael R. Jacobs, Karl G. Kristinsson, Bernard W. Beall, Keith P. Klugman, Julian Parkhill, Regine Hakenbeck, Stephen D. Bentley, and Angela B. Brueggemann. Pneumococcal capsular switching: A historical perspective. *The Journal of Infectious Diseases*, 207(3):439–449, 2013.
- [278] In Ho Park, Kyung-Hyo Kim, Ana Lucia Andrade, David E Briles, Larry S McDaniel, and Moon H Nahm. Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene pspk. *MBio*, 3(3):e00035–12, 2012.
- [279] Nicolas Gisch, Katharina Peters, Ulrich Zähringer, and Waldemar Vollmer. *Chapter 8 - The Pneumococcal Cell Wall A2 - Brown, Jeremy*, pages 145–167. Academic Press, Amsterdam, 2015.
- [280] Nhat Khai Bui, Alice Eberhardt, Daniela Vollmer, Thomas Kern, Catherine Bougault, Alexander Tomasz, Jean-Pierre Simorre, and Waldemar Vollmer. Isolation and analysis of

- cell wall components from streptococcus pneumoniae. *Analytical Biochemistry*, 421(2):657–666, 2012.
- [281] R. A. Hirst, A. Kadioglu, C. O’Callaghan, and P. W. Andrew. The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin Exp Immunol*, 138(2):195–201, 2004.
- [282] J. C. Paton, P. W. Andrew, G. J. Boulnois, and T. J. Mitchell. Molecular analysis of the pathogenicity of streptococcus pneumoniae: the role of pneumococcal proteins. *Annu Rev Microbiol*, 47:89–115, 1993.
- [283] J. E. Marshall, B. H. Faraj, A. R. Gingras, R. Lonnen, M. A. Sheikh, M. El-Mezgueldi, P. C. Moody, P. W. Andrew, and R. Wallis. The crystal structure of pneumolysin at 2.0 a resolution reveals the molecular packing of the pre-pore complex. *Sci Rep*, 5:13293, 2015.
- [284] J. K. Lemon and J. N. Weiser. Degradation products of the extracellular pathogen streptococcus pneumoniae access the cytosol via its pore-forming toxin. *MBio*, 6(1), 2015.
- [285] Benjamin P Best. Nuclear dna damage as a direct cause of aging. *Rejuvenation research*, 12(3):199–208, 2009.
- [286] Daniel R. Neill, Timothy J. Mitchell, and Aras Kadioglu. *Chapter 14 - Pneumolysin A2 - Brown, Jeremy*, pages 257–275. Academic Press, Amsterdam, 2015.
- [287] Johanna MC Jefferies, Calum HG Johnston, Lea-Ann S Kirkham, Graeme JM Cowan, Kirsty S Ross, Andrew Smith, Stuart C Clarke, Angela B Brueggemann, Robert C George, and Bruno Pichon. Presence of nonhemolytic pneumolysin in serotypes of streptococcus pneumoniae associated with disease outbreaks. *Journal of Infectious Diseases*, 196(6):936–944, 2007.
- [288] Robert A Lock, Qing Yang Zhang, Anne M Berry, and James C Paton. Sequence variation in the streptococcus pneumoniae pneumolysin gene affecting haemolytic activity and electrophoretic mobility of the toxin. *Microbial pathogenesis*, 21(2):71–83, 1996.
- [289] Lea-Ann S Kirkham, Johanna MC Jefferies, Alison R Kerr, Yu Jing, Stuart C Clarke, Andrew Smith, and Tim J Mitchell. Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *Journal of clinical microbiology*, 44(1):151–159, 2006.
- [290] PJ Morgan, PW Andrew, and TJ Mitchell. Thiol-activated cytolysins. *Reviews in Medical Microbiology*, 7(4):221–230, 1996.
- [291] FK Saunders, TJ Mitchell, JA Walker, PW Andrew, and GJ Boulnois. Pneumolysin, the thiol-activated toxin of streptococcus pneumoniae, does not require a thiol group for in vitro activity. *Infection and immunity*, 57(8):2547–2552, 1989.
- [292] H. J. Rogers and C. W. Forsberg. Role of autolysins in the killing of bacteria by some bactericidal antibiotics. *J Bacteriol*, 108(3):1235–43, 1971.
- [293] D. R. Cundell, B. J. Pearce, J. Sandros, A. M. Naughton, and H. R. Masure. Peptide permeases from streptococcus pneumoniae affect adherence to eucaryotic cells. *Infect Immun*, 63(7):2493–8, 1995.

- [294] A. Dintilhac, G. Alloing, C. Granadel, and J. P. Claverys. Competence and virulence of streptococcus pneumoniae: Adc and psaa mutants exhibit a requirement for zn and mn resulting from inactivation of putative abc metal permeases. *Mol Microbiol*, 25(4):727–39, 1997.
- [295] C. Rosenow, P. Ryan, J. N. Weiser, S. Johnson, P. Fontan, A. Ortqvist, and H. R. Masure. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of streptococcus pneumoniae. *Mol Microbiol*, 25(5):819–29, 1997.
- [296] I Pérez-Dorado, S Galan-Bartual, and JA Hermoso. Pneumococcal surface proteins: when the whole is greater than the sum of its parts. *Molecular oral microbiology*, 27(4):221–245, 2012.
- [297] Aldert Zomer, Peter WM Hermans, and Hester J Bootsma. *Non-adhesive surface proteins of Streptococcus pneumoniae*, pages 231–244. Elsevier, 2015.
- [298] Sergio Galán-Bartual, Inmaculada Pérez-Dorado, Pedro García, and Juan A Hermoso. *Structure and Function of Choline-Binding Proteins*, pages 207–230. Elsevier, 2015.
- [299] Sven Hammerschmidt, Gesina Bethe, Petra H Remane, and Gursharan S Chhatwal. Identification of pneumococcal surface protein a as a lactoferrin-binding protein of streptococcus pneumoniae. *Infection and immunity*, 67(4):1683–1687, 1999.
- [300] Mirza Shaper, Susan K Hollingshead, William H Benjamin, and David E Briles. Pspa protects streptococcus pneumoniae from killing by apolactoferrin, and antibody to psps enhances killing of pneumococci by apolactoferrin. *Infection and immunity*, 72(9):5031–5040, 2004.
- [301] Sven Hammerschmidt, Melanie P Tillig, Sonja Wolff, Jean-Pierre Vaerman, and Gursharan S Chhatwal. Species-specific binding of human secretory component to psps protein of streptococcus pneumoniae via a hexapeptide motif. *Molecular microbiology*, 36(3):726–736, 2000.
- [302] Sandhya Dave, Stephanie Carmicle, Sven Hammerschmidt, Michael K Pangburn, and Larry S McDaniel. Dual roles of pspc, a surface protein of streptococcus pneumoniae, in binding human secretory iga and factor h. *The Journal of Immunology*, 173(1):471–477, 2004.
- [303] Alison R Kerr, Gavin K Paterson, Jackie McCluskey, Francesco Iannelli, Marco R Oggioni, Gianni Pozzi, and Tim J Mitchell. The contribution of pspc to pneumococcal virulence varies between strains and is accomplished by both complement evasion and complement-independent mechanisms. *Infection and immunity*, 74(9):5319–5324, 2006.
- [304] Francesco Iannelli, Marco R. Oggioni, and Gianni Pozzi. Allelic variation in the highly polymorphic locus pspc of streptococcus pneumoniae. *Gene*, 284(1):63–71, 2002.
- [305] Antoine Dieudonné-Vatran, Stefanie Krentz, Anna M. Blom, Seppo Meri, Birgitta Henriques-Normark, Kristian Riesbeck, and Barbara Albiger. Clinical isolates of streptococcus pneumoniae bind the complement inhibitor c4b-binding protein in a pspc allele-dependent fashion. *The Journal of Immunology*, 182(12):7865, 2009.

- [306] Anna Martner, Susann Skovbjerg, James C Paton, and Agnes E Wold. Streptococcus pneumoniae autolysis prevents phagocytosis and production of phagocyte-activating cytokines. *Infection and immunity*, 77(9):3826–3837, 2009.
- [307] Anna Martner, Claes Dahlgren, James C. Paton, and Agnes E. Wold. Pneumolysin released during streptococcus pneumoniae autolysis is a potent activator of intracellular oxygen radical production in neutrophils. *Infection and Immunity*, 76(9):4079–4087, 2008.
- [308] P. Mellroth, R. Daniels, A. Eberhardt, D. Ronnlund, H. Blom, J. Widengren, S. Normark, and B. Henriques-Normark. Lyta, major autolysin of streptococcus pneumoniae, requires access to nascent peptidoglycan. *J Biol Chem*, 287(14):11018–29, 2012.
- [309] Peter Mellroth, Tatyana Sandalova, Alexey Kikhney, Francisco Vilaplana, Dusan Heseck, Mijoon Lee, Shahriar Mobashery, Staffan Normark, Dmitri Svergun, and Birgitta Henriques-Normark. Structural and functional insights into peptidoglycan access for the lytic amidase lyta of streptococcus pneumoniae. *MBio*, 5(1):e01120–13, 2014.
- [310] Elisa Ramos-Sevillano, Ana Urzainqui, Susana Campuzano, Miriam Moscoso, Fernando González-Camacho, Mirian Domenech, Santiago Rodríguez de Córdoba, Francisco Sánchez-Madrid, Jeremy S Brown, and Ernesto García. Pleiotropic effects of cell wall amidase lyta on streptococcus pneumoniae sensitivity to the host immune response. *Infection and immunity*, 83(2):591–603, 2015.
- [311] Xiao-Hui Bai, Hui-Jie Chen, Yong-Liang Jiang, Zhensong Wen, Yubin Huang, Wang Cheng, Qiong Li, Lei Qi, Jing-Ren Zhang, and Yuxing Chen. Structure of pneumococcal peptidoglycan hydrolase lytb reveals insights into the bacterial cell wall remodeling and pathogenesis. *Journal of Biological Chemistry*, 289(34):23403–23416, 2014.
- [312] P. Garcia, M. P. Gonzalez, E. Garcia, R. Lopez, and J. L. Garcia. Lytb, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol*, 31(4):1275–81, 1999.
- [313] E. Ramos-Sevillano, M. Moscoso, P. Garcia, E. Garcia, and J. Yuste. Nasopharyngeal colonization and invasive disease are enhanced by the cell wall hydrolases lytb and lytc of streptococcus pneumoniae. *PLoS One*, 6(8):e23626, 2011.
- [314] Mirian Domenech, Susana Ruiz, Miriam Moscoso, and Ernesto García. In vitro biofilm development of streptococcus pneumoniae and formation of choline-binding protein–dna complexes. *Environmental microbiology reports*, 7(5):715–727, 2015.
- [315] Inmaculada Pérez-Dorado, Ana González, María Morales, Reyes Sanles, Waldemar Striker, Waldemar Vollmer, Shahriar Mobashery, José L García, Martín Martínez-Ripoll, and Pedro García. Insights into pneumococcal fratricide from the crystal structures of the modular killing factor lytc. *Nature Structural and Molecular Biology*, 17(5):576, 2010.
- [316] Vegard Eldholm, Ola Johnsborg, Kristine Haugen, Hilde Solheim Ohnstad, and Leiv Sigve Håvarstein. Fratricide in streptococcus pneumoniae: contributions and role of the cell wall hydrolases cbpd, lyta and lytc. *Microbiology*, 155(7):2223–2234, 2009.

- [317] Rafael Molina, Ana González, Meike Stelter, Inmaculada Pérez-Dorado, Richard Kahn, María Morales, Susana Campuzano, Nuria E. Campillo, Shahriar Mobashery, José L. García, Pedro García, and Juan A. Hermoso. Crystal structure of cbpf, a bifunctional choline-binding protein and autolysis regulator from streptococcus pneumoniae. *EMBO Reports*, 10(3):246–251, 2009.
- [318] Sébastien Guiral, Tim J. Mitchell, Bernard Martin, and Jean-Pierre Claverys. Competence-programmed predation of noncompetent cells in the human pathogen streptococcus pneumoniae: Genetic requirements. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24):8710–8715, 2005.
- [319] Jeremy S. Brown, Sarah M. Gilliland, Javier Ruiz-Albert, and David W. Holden. Characterization of pit, a streptococcus pneumoniae iron uptake abc transporter. *Infection and Immunity*, 70(8):4389–4398, 2002.
- [320] Xiao-Yan Yang, Bin Sun, Liang Zhang, Nan Li, Junlong Han, Jing Zhang, Xuesong Sun, and Qing-Yu He. Chemical interference with iron transport systems to suppress bacterial growth of streptococcus pneumoniae. *PLoS ONE*, 9(8):e105953, 2014.
- [321] Evelyne Deplazes, Stephanie L Begg, Jessica H Van Wonderen, Rebecca Campbell, Bostjan Kobe, James C Paton, Fraser MacMillan, Christopher A McDevitt, and Megan L O’Mara. Characterizing the conformational dynamics of metal-free psaa using molecular dynamics simulations and electron paramagnetic resonance spectroscopy. *Biophysical chemistry*, 207:51–60, 2015.
- [322] Jason W. Johnston, Lisa E. Myers, Martina M. Ochs, William H. Benjamin, David E. Briles, and Susan K. Hollingshead. Lipoprotein psaa in virulence of streptococcus pneumoniae: Surface accessibility and role in protection from superoxide. *Infection and Immunity*, 72(10):5858–5867, 2004.
- [323] A. M. Berry and J. C. Paton. Sequence heterogeneity of psaa, a 37-kilodalton putative adhesin essential for virulence of streptococcus pneumoniae. *Infection and Immunity*, 64(12):5255–5262, 1996.
- [324] Rachael H Whalan, Simon GP Funnell, Lucas D Bowler, Michael J Hudson, Andrew Robinson, and Christopher G Dowson. Distribution and genetic diversity of the abc transporter lipoproteins piua and piaa within streptococcus pneumoniae and related streptococci. *Journal of bacteriology*, 188(3):1031–1038, 2006.
- [325] Rachael H Whalan, Simon GP Funnell, Lucas D Bowler, Michael J Hudson, Andrew Robinson, and Christopher G Dowson. Piua and piaa, iron uptake lipoproteins of streptococcus pneumoniae, elicit serotype independent antibody responses following human pneumococcal septicemia. *FEMS Immunology and Medical Microbiology*, 43(1):73–80, 2005.
- [326] Jeremy S Brown, A David Ogunniyi, Matthew C Woodrow, David W Holden, and James C Paton. Immunization with components of two iron uptake abc transporters protects mice against systemic streptococcus pneumoniae infection. *Infection and immunity*, 69(11):6702–6706, 2001.

- [327] Jeremy S Brown, Sarah M Gilliland, and David W Holden. A streptococcus pneumoniae pathogenicity island encoding an abc transporter involved in iron uptake and virulence. *Molecular microbiology*, 40(3):572–585, 2001.
- [328] Erin S. Honsa, Michael D. L. Johnson, and Jason W. Rosch. The roles of transition metals in the physiology and pathogenesis of streptococcus pneumoniae. *Frontiers in Cellular and Infection Microbiology*, 3:92, 2013.
- [329] Wang Cheng, Qiong Li, Yong-Liang Jiang, Cong-Zhao Zhou, and Yuxing Chen. Structures of streptococcus pneumoniae piaa and its complex with ferrichrome reveal insights into the substrate binding and release of high affinity iron transporters. *PLOS ONE*, 8(8):e71451, 2013.
- [330] Douglas I Johnson. *Bacterial Pathogens and Their Virulence Factors*. Springer, 2017.
- [331] Liang Zhang, Nan Li, Kun Cao, Xiao-Yan Yang, Guandi Zeng, Xuesong Sun, and Qing-Yu He. Crucial residue trp158 of lipoprotein piaa stabilizes the ferrichrome-piaa complex in streptococcus pneumoniae. *Journal of Inorganic Biochemistry*, 167(Supplement C):150–156, 2017.
- [332] S. S. Tai, C. J. Lee, and R. E. Winter. Hemin utilization is related to virulence of streptococcus pneumoniae. *Infection and Immunity*, 61(12):5401–5405, 1993.
- [333] Stanley S. Tai, Chialin Yu, and Janice K. Lee. A solute binding protein of streptococcus pneumoniae iron transport. *FEMS Microbiology Letters*, 220(2):303–308, 2003.
- [334] Xiao-Yan Yang, Ke He, Gaofei Du, Xiaohui Wu, Guangchuang Yu, Yunlong Pan, Gong Zhang, Xuesong Sun, and Qing-Yu He. Integrated translomics with proteomics to identify novel iron-transporting proteins in streptococcus pneumoniae. *Frontiers in Microbiology*, 7(78), 2016.
- [335] Jos Boekhorst, Mark W. H. J. de Been, Michiel Kleerebezem, and Roland J. Siezen. Genome-wide detection and analysis of cell wall-bound proteins with lpxtg-like sorting motifs. *Journal of Bacteriology*, 187(14):4928–4934, 2005.
- [336] Luciano A Marraffini, Andrea C DeDent, and Olaf Schneewind. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiology and Molecular Biology Reviews*, 70(1):192–221, 2006.
- [337] Arun S. Kharat and Alexander Tomasz. Inactivation of the srta gene affects localization of surface proteins and decreases adhesion of streptococcus pneumoniae to human pharyngeal cells in vitro. *Infection and Immunity*, 71(5):2758–2765, 2003.
- [338] Sergio Galán-Bartual, Inmaculada Pérez-Dorado, Pedro García, and Juan A. Hermoso. Chapter 11 - Structure and Function of Choline-Binding Proteins A2 - Brown, Jeremy, pages 207–230. Academic Press, Amsterdam, 2015.
- [339] Nadja Noske, Ulrike Kämmerer, Manfred Rohde, and Sven Hammerschmidt. Pneumococcal interaction with human dendritic cells: phagocytosis, survival, and induced adaptive immune response are manipulated by pava. *The Journal of Immunology*, 183(3):1952–1963, 2009.

- [340] Jan Kolberg, Audun Aase, Simone Bergmann, Tove K Herstad, Gunnhild Rødal, Ronald Frank, Manfred Rohde, and Sven Hammerschmidt. Streptococcus pneumoniae enolase is important for plasminogen binding despite low abundance of enolase protein on the bacterial cell surface. *Microbiology*, 152(5):1307–1317, 2006.
- [341] Àngels Díaz-Ramos, Anna Roig-Borrellas, Ana García-Melero, and Roser López-Alemany. α -enolase, a multifunctional protein: Its role on pathophysiological situations. *Journal of Biomedicine and Biotechnology*, 2012:156795, 2012.
- [342] Vaibhav Agarwal, Sven Hammerschmidt, Sven Malm, Simone Bergmann, Kristian Riesbeck, and Anna M Blom. Enolase of streptococcus pneumoniae binds human complement inhibitor c4b-binding protein and contributes to complement evasion. *The journal of immunology*, 189(7):3575–3584, 2012.
- [343] Yuka Mori, Masaya Yamaguchi, Yutaka Terao, Shigeyuki Hamada, Takashi Ooshima, and Shigetada Kawabata. α -enolase of streptococcus pneumoniae induces formation of neutrophil extracellular traps. *The Journal of Biological Chemistry*, 287(13):10472–10481, 2012.
- [344] Simone Bergmann, Manfred Rohde, and Sven Hammerschmidt. Glyceraldehyde-3-phosphate dehydrogenase of streptococcus pneumoniae is a surface-displayed plasminogen-binding protein. *Infection and Immunity*, 72(4):2416–2419, 2004.
- [345] Gabriella De Angelis, Monica Moschioni, Alessandro Muzzi, Alfredo Pezzicoli, Stefano Censini, Isabel Delany, Morena Lo Sapio, Antonia Sinisi, Claudio Donati, Vega Masignani, and Michèle A. Barocchi. The streptococcus pneumoniae pilus-1 displays a biphasic expression pattern. *PLOS ONE*, 6(6):e21269, 2011.
- [346] Fabio Bagnoli, Monica Moschioni, Claudio Donati, Valentina Dimitrovska, Ilaria Ferlenghi, Claudia Facciotti, Alessandro Muzzi, Fabiola Giusti, Carla Emolo, and Antonella Sinisi. A second pilus type in streptococcus pneumoniae is prevalent in emerging serotypes and mediates adhesion to host cells. *Journal of bacteriology*, 190(15):5480–5492, 2008.
- [347] Sandra Muschiol, Simon Erlendsson, Marie-Stephanie Aschtgen, Vitor Oliveira, Peter Schmieder, Casper de Lichtenberg, Kaare Teilum, Thomas Boesen, Umit Akbey, and Birgitta Henriques-Normark. Structure of the competence pilus major pilin comgc in streptococcus pneumoniae. *Journal of Biological Chemistry*, 292(34):14134–14146, 2017.
- [348] Raphaël Laurenceau, Gérard Péhau-Arnaudet, Sonia Baconnais, Joseph Gault, Christian Malosse, Annick Dujeancourt, Nathalie Campo, Julia Chamot-Rooke, Eric Le Cam, Jean-Pierre Claverys, and Rémi Fronzes. A type iv pilus mediates dna binding during natural transformation in streptococcus pneumoniae. *PLoS Pathogens*, 9(6):e1003473, 2013.
- [349] JM Woof and MW Russell. Structure and function relationships in iga. *Mucosal immunology*, 4(6):590–597, 2011.
- [350] Prashant Rai, Marcus Parrish, Ian Jun Jie Tay, Na Li, Shelley Ackerman, Fang He, Jimmy Kwang, Vincent T. Chow, and Bevin P. Engelward. Streptococcus pneumoniae secretes hydrogen peroxide leading to dna damage and apoptosis in lung cells. *Proceedings of the National Academy of Sciences of the United States of America*, 112(26):E3421–E3430, 2015.

- [351] Christopher D. Pericone, Karin Overweg, Peter W. M. Hermans, and Jeffrey N. Weiser. Inhibitory and bactericidal effects of hydrogen peroxide production by streptococcus pneumoniae on other inhabitants of the upper respiratory tract. *Infection and Immunity*, 68(7):3990–3997, 2000.
- [352] Maria Loose, Martina Hudel, Klaus-Peter Zimmer, Ernesto Garcia, Sven Hammerschmidt, Rudolf Lucas, Trinad Chakraborty, and Helena Pillich. Pneumococcal hydrogen peroxide–induced stress signaling regulates inflammatory genes. *The Journal of Infectious Diseases*, 211(2):306–316, 2015.
- [353] Stella Pesakhov, Rachel Benisty, Noga Sikron, Zvi Cohen, Pavel Gomelsky, Inna Khozin-Goldberg, Ron Dagan, and Nurith Porat. Effect of hydrogen peroxide production and the fenton reaction on membrane composition of streptococcus pneumoniae. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(3):590–597, 2007.
- [354] Hasan Yesilkaya, Vahid Farshchi Andisi, Peter W Andrew, and Jetta JE Bijlsma. Streptococcus pneumoniae and reactive oxygen species: an unusual approach to living with radicals. *Trends in microbiology*, 21(4):187–195, 2013.
- [355] Christopher D Pericone, Sunny Park, James A Imlay, and Jeffrey N Weiser. Factors contributing to hydrogen peroxide resistance in streptococcus pneumoniae include pyruvate oxidase (spxb) and avoidance of the toxic effects of the fenton reaction. *Journal of bacteriology*, 185(23):6815–6825, 2003.
- [356] Ulrich Dobrindt, Bianca Hochhut, Ute Hentschel, and Jörg Hacker. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology*, 2(5):414, 2004.
- [357] Herbert Schmidt and Michael Hensel. Pathogenicity islands in bacterial pathogenesis. *Clinical Microbiology Reviews*, 17(1):14–56, 2004.
- [358] Pooja Shivshankar, Carlos Sanchez, Lloyd F. Rose, and Carlos J. Orihuela. The streptococcus pneumoniae adhesin psrp binds to keratin 10 on lung cells. *Molecular microbiology*, 73(4):663–679, 2009.
- [359] Anel Lizcano, Ramya Akula Suresh Babu, Anukul T. Shenoy, Alison Maren Saville, Nikhil Kumar, Adonis D’Mello, Cecilia A. Hinojosa, Ryan P. Gilley, Jesus Segovia, Timothy J. Mitchell, Hervé Tettelin, and Carlos J. Orihuela. Transcriptional organization of pneumococcal psrp-secy2a2 and impact of gtfa and gtfb deletion on psrp-associated virulence properties. *Microbes and Infection*, 19(6):323–333, 2017.
- [360] R. M. Donlan. Biofilms: microbial life on surfaces. *Emerg Infect Dis*, 8(9):881–90, 2002.
- [361] Melissa B. Oliver and W. Edward Swords. *Chapter 16 - Pneumococcal Biofilms and Bacterial Persistence During Otitis Media Infections A2 - Brown, Jeremy*, pages 293–308. Academic Press, Amsterdam, 2015.

- [362] Laura R Marks, Ryan M Reddinger, and Anders P Hakansson. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in streptococcus pneumoniae. *MBio*, 3(5):e00200–12, 2012.
- [363] César de la Fuente-Núñez, Fany Reffuveille, Lucía Fernández, and Robert E. W. Hancock. Bacterial biofilm development as a multicellular adaptation: antibiotic resistance and new therapeutic strategies. *Current Opinion in Microbiology*, 16(5):580–589, 2013.
- [364] C. J. Sanchez, P. Shivshankar, K. Stol, S. Trakhtenbroit, P. M. Sullam, K. Sauer, P. W. Hermans, and C. J. Orihuela. The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. *PLoS Pathog*, 6(8):e1001044, 2010.
- [365] Hong Wu, Claus Moser, Heng-Zhuang Wang, Niels Højby, and Zhi-Jun Song. Strategies for combating bacterial biofilm infections. *International Journal of Oral Science*, 7(1):1–7, 2015.
- [366] Manfred Fliegau, Andreas F. P. Sonnen, Bernhard Kremer, and Philipp Henneke. Mucociliary clearance defects in a murine in vitro model of pneumococcal airway infection. *PLoS ONE*, 8(3):e59925, 2013.
- [367] Kenneth Murphy and Casey Weaver. *Innate Immunity: The First Line of Defense*, book section 02. Garland Science, 9th edition, 2017.
- [368] A. Kadioglu and P. W. Andrew. The innate immune response to pneumococcal lung infection: the untold story. *Trends Immunol*, 25(3):143–9, 2004.
- [369] C.A. Jr. Janeway, P. Travers, and M. Walport. *The complement system and innate immunity*. Garland Science, 2001.
- [370] J. A. Whitsett and T. Alenghat. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nat Immunol*, 16(1):27–35, 2015.
- [371] MB Antunes and NA Cohen. Mucociliary clearance—a critical upper airway host defense mechanism and methods of assessment. *Current Opinion in allergy and clinical immunology*, 7(1):5–10, 2007.
- [372] G. O. André, W. R. Politano, S. Mirza, T. R. Converso, L. F. C. Ferraz, L. C. C. Leite, and M. Darrieux. Combined effects of lactoferrin and lysozyme on streptococcus pneumoniae killing. *Microbial Pathogenesis*, 89:7–17, 2015.
- [373] Kathryn L. Nawrocki, Emily K. Crispell, and Shonna M. McBride. Antimicrobial peptide resistance mechanisms of gram-positive bacteria. *Antibiotics*, 3(4):461–492, 2014.
- [374] Márta Kovács, Alexander Halfmann, Iris Fedtke, Manuel Heintz, Andreas Peschel, Walde-mar Vollmer, Regine Hakenbeck, and Reinhold Brückner. A functional dlt operon, encoding proteins required for incorporation of d-alanine in teichoic acids in gram-positive bacteria, confers resistance to cationic antimicrobial peptides in streptococcus pneumoniae. *Journal of bacteriology*, 188(16):5797–5805, 2006.

- [375] David H. Dockrell and Jeremy S. Brown. *Chapter 21 - Streptococcus pneumoniae Interactions with Macrophages and Mechanisms of Immune Evasion*, pages 401–422. Academic Press, Amsterdam, 2015.
- [376] JH Shelhamer, Z Marom, C Logun, and M Kaliner. Human respiratory mucous glycoproteins. *Experimental lung research*, 7(2):149–162, 1984.
- [377] Hasan Yesilkaya, Sonia Manco, Aras Kadioglu, Vanessa S Terra, and Peter W Andrew. The ability to utilize mucin affects the regulation of virulence gene expression in streptococcus pneumoniae. *FEMS microbiology letters*, 278(2):231–235, 2007.
- [378] Kimberly M. Davis, Henry T. Akinbi, Alistair J. Standish, and Jeffrey N. Weiser. Resistance to mucosal lysozyme compensates for the fitness deficit of peptidoglycan modifications by streptococcus pneumoniae. *PLOS Pathogens*, 4(12):e1000241, 2008.
- [379] J. N. Weiser, R. Austrian, P. K. Sreenivasan, and H. R. Masure. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infection and Immunity*, 62(6):2582–2589, 1994.
- [380] Barry B. Mook-Kanamori, Madelijn Geldhoff, Tom van der Poll, and Diederik van de Beek. Pathogenesis and pathophysiology of pneumococcal meningitis. *Clinical Microbiology Reviews*, 24(3):557–591, 2011.
- [381] Adam Wanner, Matthias Salathé, and Thomas G O’Riordan. Mucociliary clearance in the airways. *American journal of respiratory and critical care medicine*, 154(6):1868–1902, 1996.
- [382] Barbara R. Grubb, Alessandra Livraghi-Butrico, Troy D. Rogers, Weining Yin, Brian Button, and Lawrence E. Ostrowski. Reduced mucociliary clearance in old mice is associated with a decrease in muc5b mucin. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 310(9):L860–L867, 2016.
- [383] A. Craig, J. Mai, S. Cai, and S. Jeyaseelan. Neutrophil recruitment to the lungs during bacterial pneumonia. *Infect Immun*, 77(2):568–75, 2009.
- [384] E. Kolaczkowska and P. Kubes. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol*, 13(3):159–75, 2013.
- [385] Judith Falloon and John I Gallin. Neutrophil granules in health and disease. *Journal of Allergy and Clinical Immunology*, 77(5):653–662, 1986.
- [386] E. E. Gardiner and R. K. Andrews. Neutrophil extracellular traps (nets) and infection-related vascular dysfunction. *Blood Reviews*, 26(6):255–9, 2012.
- [387] A. K. Simon, G. A. Hollander, and A. McMichael. Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci*, 282(1821):20143085, 2015.

- [388] Claudia Nussbaum, Anna Gloning, Monika Pruenster, David Frommhold, Susanne Bierschenk, Orsolya Genzel-Boroviczény, Ulrich H. von Andrian, Elizabeth Quackenbush, and Markus Sperandio. Neutrophil and endothelial adhesive function during human fetal ontogeny. *Journal of Leukocyte Biology*, 93(2):175–184, 2013.
- [389] Athanasios Filias, Georgios L. Theodorou, Sofia Mouzopoulou, Anastasia A. Varvarigou, Stephanos Mantagos, and Marina Karakantza. Phagocytic ability of neutrophils and monocytes in neonates. *BMC Pediatrics*, 11:29–29, 2011.
- [390] Elizabeth Sapey, Hannah Greenwood, Georgia Walton, Elizabeth Mann, Alexander Love, Natasha Aaronson, Robert H Insall, Robert A Stockley, and Janet M Lord. Phosphoinositide 3-kinase inhibition restores neutrophil accuracy in the elderly: toward targeted treatments for immunosenescence. *Blood*, 123(2):239–248, 2014.
- [391] Birgit Simell, Arja Vuorela, Nina Ekström, Arto Palmu, Antti Reunanen, Seppo Meri, Helena Käyhty, and Merja Väkeväinen. Aging reduces the functionality of anti-pneumococcal antibodies and the killing of streptococcus pneumoniae by neutrophil phagocytosis. *Vaccine*, 29(10):1929–1934, 2011.
- [392] G. Arango Duque and A. Descoteaux. Macrophage cytokines: involvement in immunity and infectious diseases. *Frontiers in Immunology*, 5:491, 2014.
- [393] C. D. Gregory and A. Devitt. The macrophage and the apoptotic cell: an innate immune interaction viewed simplistically? *Immunology*, 113(1):1–14, 2004.
- [394] Georg Kraal, Luc JW van der Laan, Outi Elomaa, and Karl Tryggvason. The macrophage receptor marco. *Microbes and infection*, 2(3):313–316, 2000.
- [395] Elisabeth Förster-Waldl, Kambis Sadeghi, Dietmar Tamandl, Bernadette Gerhold, Ulrike Hallwirth, Klaudia Rohrmeister, Michael Hayde, Andrea R Prusa, Kurt Herkner, and George Boltz-Nitulescu. Monocyte toll-like receptor 4 expression and lps-induced cytokine production increase during gestational aging. *Pediatric research*, 58(1):121, 2005.
- [396] K. Takeda and S. Akira. Toll-like receptors in innate immunity. *Int Immunol*, 17(1):1–14, 2005.
- [397] A. Anas, T. van der Poll, and A. F. de Vos. Role of cd14 in lung inflammation and infection. *Critical Care*, 14(2):209, 2010.
- [398] Nicole Fitzner, Sigrid Clauberg, Frank Essmann, Joerg Liebmann, and Victoria Kolb-Bachofen. Human skin endothelial cells can express all 10 tlr genes and respond to respective ligands. *Clinical and Vaccine Immunology*, 15(1):138–146, 2008.
- [399] L. A. O’Neill, C. E. Bryant, and S. L. Doyle. Therapeutic targeting of toll-like receptors for infectious and inflammatory diseases and cancer. *Pharmacol Rev*, 61(2):177–97, 2009.
- [400] M. C. Dessing, M. Schouten, C. Draing, M. Levi, S. von Aulock, and T. van der Poll. Role played by toll-like receptors 2 and 4 in lipoteichoic acid-induced lung inflammation and coagulation. *J Infect Dis*, 197(2):245–52, 2008.

- [401] S. Kohler, F. Voss, A. Gomez Mejia, J. S. Brown, and S. Hammerschmidt. Pneumococcal lipoproteins involved in bacterial fitness, virulence, and immune evasion. *FEBS Lett*, 590(21):3820–3839, 2016.
- [402] Gillian Tomlinson, Suneeta Chimalapati, Tracey Pollard, Thabo Lapp, Jonathan Cohen, Emilie Camberlein, Sian Stafford, Jimstan Periselneris, Christine Aldridge, Waldemar Vollmer, Capucine Picard, Jean-Laurent Casanova, Mahdad Noursadeghi, and Jeremy Brown. Tlr-mediated inflammatory responses to streptococcus pneumoniae are highly dependent on surface expression of bacterial lipoproteins. *The Journal of Immunology Author Choice*, 193(7):3736–3745, 2014.
- [403] Aimee L Richard, Steven J Siegel, Jan Erikson, and Jeffrey N Weiser. Tlr2 signaling decreases transmission of streptococcus pneumoniae by limiting bacterial shedding in an infant mouse influenza a co-infection model. *PLoS pathogens*, 10(8):e1004339, 2014.
- [404] Srivastava A, Henneke P, Visintin A, Morse SC, Martiv V, Watkins C, Paton JC, Wessels MR, Golenbock DT, and Malley R. The apoptotic response to pneumolysin is toll-like receptor 4 dependent and protects against pneumococcal disease. *Infect Immun*, 73(10):6479–6487, 2005.
- [405] T. Kawai and S. Akira. The roles of tlrs, rlrs and nlrs in pathogen recognition. *Int Immunol*, 21(4):317–37, 2009.
- [406] P. J. Godowski. A smooth operator for lps responses. *Nature Immunology*, 6(6):544–6, 2005.
- [407] Neil Warner and Gabriel Núñez. Myd88: a critical adaptor protein in innate immunity signal transduction. *The Journal of Immunology*, 190(1):3–4, 2013.
- [408] Adeeb H Rahman, Devon K Taylor, and Laurence A Turka. The contribution of direct tlr signaling to t cell responses. *Immunologic research*, 45(1):25–36, 2009.
- [409] Jodie L Simpson, Vanessa M McDonald, Katherine J Baines, Kevin M Oreo, Fang Wang, Philip M Hansbro, and Peter G Gibson. Influence of age, past smoking, and disease severity on tlr2, neutrophilic inflammation, and mmp-9 levels in copd. *Mediators of inflammation*, 2013, 2013.
- [410] Alejandro Ramirez, Vijay Rathinam, Katherine A Fitzgerald, Douglas T Golenbock, and Anuja Mathew. Defective pro-il-1 β responses in macrophages from aged mice. *Immunity and Ageing*, 9(1):27, 2012.
- [411] Soo Jung Cho, Kristen Rooney, Augustine MK Choi, and Heather W Stout-Delgado. Nlrp3 inflammasome activation in aged macrophages is diminished during streptococcus pneumoniae infection. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 314(3):L372–L387, 2017.
- [412] Angela R Boyd, Pooja Shivshankar, Shoulei Jiang, Michael T Berton, and Carlos J Orihuela. Age-related defects in tlr2 signaling diminish the cytokine response by alveolar macrophages during murine pneumococcal pneumonia. *Experimental gerontology*, 47(7):507–518, 2012.

- [413] Peter J Murray. Beyond peptidoglycan for nod2. *Nature immunology*, 10(10):1053–1054, 2009.
- [414] L. Franchi, N. Warner, K. Viani, and G. Nunez. Function of nod-like receptors in microbial recognition and host defense. *Immunol Rev*, 227(1):106–28, 2009.
- [415] Stephen E Girardin, Ivo G Boneca, Jérôme Viala, Mathias Chamaillard, Agnes Labigne, Gilles Thomas, Dana J Philpott, and Philippe J Sansonetti. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (mdp) detection. *Journal of Biological Chemistry*, 278(11):8869–8872, 2003.
- [416] Kimberly M. Davis, Shigeki Nakamura, and Jeffrey N. Weiser. Nod2 sensing of lysozyme-digested peptidoglycan promotes macrophage recruitment and clearance of s. pneumoniae colonization in mice. *The Journal of Clinical Investigation*, 121(9):3666–3676, 2011.
- [417] M. C. Dessing, S. Knapp, S. Florquin, A. F. de Vos, and T. van der Poll. Cd14 facilitates invasive respiratory tract infection by streptococcus pneumoniae. *Am J Respir Crit Care Med*, 175(6):604–11, 2007.
- [418] Janeway CA Jr, Travers P, and Walport M. *Immunobiology: The Immune System in Health and Disease*. Garland Science, 5th edition, 2001.
- [419] L. Zhang, Z. Li, Z. Wan, A. Kilby, J. M. Kilby, and W. Jiang. Humoral immune responses to streptococcus pneumoniae in the setting of hiv-1 infection. *Vaccine*, 33(36):4430–6, 2015.
- [420] Jeffrey N Weiser, Deborah Bae, Claudine Fasching, Ronald W Scamurra, Adam J Ratner, and Edward N Janoff. Antibody-enhanced pneumococcal adherence requires iga1 protease. *Proceedings of the National Academy of Sciences*, 100(7):4215–4220, 2003.
- [421] Sergio Romagnani. Th1/th2 cells. *Inflammatory bowel diseases*, 5(4):285–294, 1999.
- [422] Richard Malley, Krzysztof Trzcinski, Amit Srivastava, Claudette M. Thompson, Porter W. Anderson, and Marc Lipsitch. Cd4(+) t cells mediate antibody-independent acquired immunity to pneumococcal colonization. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13):4848–4853, 2005.
- [423] Edwin Hoe, Jeremy Anderson, Jordan Nathanielsz, Zheng Quan Toh, Rachel Marimla, Anne Balloch, and Paul V Licciardi. The contrasting role of th17 immunity in human health and disease. *Microbiology and immunology*, 2017.
- [424] Taj Azarian, Lindsay R Grant, Maria Georgieva, Laura L Hammitt, Raymond Reid, Stephen D Bentley, David Goldblatt, Mathuran Santosham, Robert Weatherholtz, and Paula Burbidge. Association of pneumococcal protein antigen serology with age and antigenic profile of colonizing isolates. *The Journal of infectious diseases*, 215(5):713–722, 2017.
- [425] E. AlonsoDeVelasco, A. F. Verheul, J. Verhoef, and H. Snippe. Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines. *Microbiological Reviews*, 59(4):591–603, 1995.

- [426] Barry M Gray, George M Converse III, and Hugh C Dillon Jr. Epidemiologic studies of streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *Journal of Infectious Diseases*, 142(6):923–933, 1980.
- [427] Barry M Gray and Hugh C Dillon Jr. Epidemiological studies of streptococcus pneumoniae in infants: antibody to types 3, 6, 14, and 23 in the first two years of life. *Journal of Infectious Diseases*, 158(5):948–955, 1988.
- [428] Brandon Coder and Dong-Ming Su. Thymic involution beyond t-cell insufficiency. *Onco-target*, 6(26):21777, 2015.
- [429] Kornelis SM van der Geest, Wayel H Abdulahad, Sarah M Tete, Pedro G Lorencetti, Gerda Horst, Nicolaas A Bos, Bart-Jan Kroesen, Elisabeth Brouwer, and Annemieke MH Boots. Aging disturbs the balance between effector and regulatory cd4+ t cells. *Experimental gerontology*, 60:190–196, 2014.
- [430] Alberto Baroja-Mazo, Fatima Martín-Sánchez, Ana I Gomez, Carlos M Martínez, Joaquín Amores-Iniesta, Vincent Compan, Maria Barberà-Cremades, Jordi Yagüe, Estibaliz Ruiz-Ortiz, and Jordi Antón. The nlrp3 inflammasome is released as a particulate danger signal that amplifies the inflammatory response. *Nature immunology*, 15(8):738–748, 2014.
- [431] Edel A McNeela, Áine Burke, Daniel R Neill, Cathy Baxter, Vitor E Fernandes, Daniela Ferreira, Sarah Smeaton, Rana El-Rachkidy, Rachel M McLoughlin, and Andres Mori. Pneumolysin activates the nlrp3 inflammasome and promotes proinflammatory cytokines independently of tlr4. *PLoS Pathogens*, 6(11):e1001191, 2010.
- [432] Martin Witzenrath, Florence Pache, Daniel Lorenz, Uwe Koppe, Birgitt Gutbier, Christoph Tabeling, Katrin Reppe, Karolin Meixenberger, Anca Dorhoi, and Jiangtao Ma. The nlrp3 inflammasome is differentially activated by pneumolysin variants and contributes to host defense in pneumococcal pneumonia. *The Journal of Immunology*, 187(1):434–440, 2011.
- [433] Jeremy S Brown, Tracy Hussell, Sarah M Gilliland, David W Holden, James C Paton, Michael R Ehrenstein, Mark J Walport, and Marina Botto. The classical pathway is the dominant complement pathway required for innate immunity to streptococcus pneumoniae infection in mice. *Proceedings of the National Academy of Sciences*, 99(26):16969–16974, 2002.
- [434] Youssif M Ali, Nicholas J Lynch, Kashif S Haleem, Teizo Fujita, Yuichi Endo, Soren Hansen, Uffe Holmskov, Kazue Takahashi, Gregory L Stahl, and Thomas Dudler. The lectin pathway of complement activation is a critical component of the innate immune response to pneumococcal infection. *PLoS Pathogens*, 8(7):e1002793, 2012.
- [435] A. R. Kerr, G. K. Paterson, A. Riboldi-Tunnicliffe, and T. J. Mitchell. Innate immune defense against pneumococcal pneumonia requires pulmonary complement component c3. *Infect Immun*, 73(7):4245–52, 2005.
- [436] Charles A Davis, Enrique H Vallota, and Judith Forristal. Serum complement levels in infancy: age related changes. *Pediatric research*, 13(9):1043, 1979.

- [437] AS Grumach, ME Ceccon, R Rutz, A Fertig, and M Kirschfink. Complement profile in neonates of different gestational ages. *Scandinavian journal of immunology*, 79(4):276–281, 2014.
- [438] Rob Roy MacGregor and Meir Shalit. Neutrophil function in healthy elderly subjects. *Journal of gerontology*, 45(2):M55–M60, 1990.
- [439] Abul K Abbas, Andrew HH Lichtman, and Shiv Pillai. *Cellular and molecular immunology*. Elsevier Health Sciences, 2014.
- [440] Reshmi Mukerji, Shaper Mirza, Aoife M Roche, Rebecca W Widener, Christina M Croney, Dong-Kwon Rhee, Jeffrey N Weiser, Alexander J Szalai, and David E Briles. Pneumococcal surface protein a inhibits complement deposition on the pneumococcal surface by competing with the binding of c-reactive protein to cell-surface phosphocholine. *The Journal of Immunology*, 189(11):5327–5335, 2012.
- [441] Damon P Eisen, Melinda M Dean, Marja A Boermeester, Katy J Fidler, Anthony C Gordon, Gitte Kronborg, Jürgen FJ Kun, Yu Lung Lau, Antonis Payeras, and Helgi Valdimarsson. Low serum mannose-binding lectin level increases the risk of death due to pneumococcal infection. *Clinical Infectious Diseases*, 47(4):510–516, 2008.
- [442] J. S. Bradley, C. L. Byington, S. S. Shah, B. Alverson, E. R. Carter, C. Harrison, S. L. Kaplan, S. E. Mace, Jr. McCracken, G. H., M. R. Moore, S. D. St Peter, J. A. Stockwell, J. T. Swanson, Society Pediatric Infectious Diseases, and America the Infectious Diseases Society of. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the pediatric infectious diseases society and the infectious diseases society of america. *Clin Infect Dis*, 53(7):e25–76, 2011.
- [443] M. S. Niederman, L. A. Mandell, A. Anzueto, J. B. Bass, W. A. Broughton, G. D. Campbell, N. Dean, T. File, M. J. Fine, P. A. Gross, F. Martinez, T. J. Marrie, J. F. Plouffe, J. Ramirez, G. A. Sarosi, A. Torres, R. Wilson, V. L. Yu, and Society American Thoracic. Guidelines for the management of adults with community-acquired pneumonia. diagnosis, assessment of severity, antimicrobial therapy, and prevention. *Am J Respir Crit Care Med*, 163(7):1730–54, 2001.
- [444] L. K. Grossman and S. E. Caplan. Clinical, laboratory, and radiological information in the diagnosis of pneumonia in children. *Annals of Emergency Medicine*, 17(1):43–6, 1988.
- [445] K. Mulholland and C. Satzke. Serotype replacement after pneumococcal vaccination. *Lancet*, 379(9824):1387; author reply 1388–9, 2012.
- [446] A. M. Werno and D. R. Murdoch. Medical microbiology: laboratory diagnosis of invasive pneumococcal disease. *Clin Infect Dis*, 46(6):926–32, 2008.
- [447] J. G. Gardner, D. R. Bhamidipati, A. M. Rueda, D. T. M. Nguyen, E. A. Graviss, and D. M. Musher. White blood cell counts, alcoholism, and cirrhosis in pneumococcal pneumonia. *Open Forum Infect Dis*, 4(2):ofx034, 2017.

- [448] Aaron M Harris, Susan E Beekmann, Philip M Polgreen, and Matthew R Moore. Rapid urine antigen testing for streptococcus pneumoniae in adults with community-acquired pneumonia: clinical use and barriers. *Diagnostic microbiology and infectious disease*, 79(4):454–457, 2014.
- [449] M. D’Amato, G. Rea, V. Carnevale, M. A. Grimaldi, A. R. Saponara, E. Rosenthal, M. M. Maggi, L. Dimitri, and M. Sperandeo. Assessment of thoracic ultrasound in complementary diagnosis and in follow up of community-acquired pneumonia (cap). *BMC Med Imaging*, 17(1):52, 2017.
- [450] Sorin Claudiu Man, Otilia Fufezan, Valentina Sas, and Cristina Schnell. Performance of lung ultrasonography for the diagnosis of communityacquired pneumonia in hospitalized children. *Medical ultrasonography*, 19(3):276–281, 2017.
- [451] S. Jun, B. Park, J. B. Seo, S. Lee, and N. Kim. Development of a computer-aided differential diagnosis system to distinguish between usual interstitial pneumonia and non-specific interstitial pneumonia using texture- and shape-based hierarchical classifiers on hrct images. *J Digit Imaging*, 2017.
- [452] G. Alcoba, K. Keitel, V. Maspoli, L. Lacroix, S. Manzano, M. Gehri, R. Tabin, A. Gervaix, and A. Galetto-Lacour. A three-step diagnosis of pediatric pneumonia at the emergency department using clinical predictors, c-reactive protein, and pneumococcal pcr. *Eur J Pediatr*, 176(6):815–824, 2017.
- [453] Joon Young Song, Byung Wook Eun, and Moon H Nahm. Diagnosis of pneumococcal pneumonia: current pitfalls and the way forward. *Infection and chemotherapy*, 45(4):351–366, 2013.
- [454] Chiara Azzari, Maria Moriondo, Giuseppe Indolfi, Martina Cortimiglia, Clementina Canessa, Laura Beccioli, Francesca Lippi, Maurizio de Martino, and Massimo Resti. Realtime pcr is more sensitive than multiplex pcr for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. *PLoS ONE*, 5(2):e9282, 2010.
- [455] Lindsay Kim, Lesley McGee, Sara Tomczyk, and Bernard Beall. Biological and epidemiological features of antibiotic-resistant streptococcus pneumoniae in pre-and post-conjugate vaccine eras: a united states perspective. *Clinical microbiology reviews*, 29(3):525–552, 2016.
- [456] Siang Yong Tan and Yvonne Tatsumura. Alexander fleming (1881–1955): Discoverer of penicillin. *Singapore medical journal*, 56(7):366, 2015.
- [457] Centers for Disease Control and Prevention. Antibiotic resistance threats in the united states, 2013. 2014.
- [458] Lesley McGee, Mathias W. Pletz, John P. Fobiwe, and Keith P. Klugman. *Chapter 2 - Antibiotic Resistance of Pneumococci A2 - Brown, Jeremy*, pages 21–40. Academic Press, Amsterdam, 2015.

- [459] Dustin T King, Solmaz Sobhanifar, and Natalie CJ Strynadka. The mechanisms of resistance to β -lactam antibiotics. *Handbook of Antimicrobial Resistance*, pages 177–201, 2017.
- [460] R. Cherazard, M. Epstein, T. L. Doan, T. Salim, S. Bharti, and M. A. Smith. Antimicrobial resistant streptococcus pneumoniae: Prevalence, mechanisms, and clinical implications. *Am J Ther*, 24(3):e361–e369, 2017.
- [461] J. E. Cornick and S. D. Bentley. Streptococcus pneumoniae: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. *Microbes Infect*, 14(7-8):573–83, 2012.
- [462] RR Reinert. The antimicrobial resistance profile of streptococcus pneumoniae. *Clinical Microbiology and Infection*, 15(s3):7–11, 2009.
- [463] J. C. Butler, R. F. Breiman, J. F. Campbell, H. B. Lipman, C. V. Broome, and R. R. Facklam. Pneumococcal polysaccharide vaccine efficacy. an evaluation of current recommendations. *JAMA*, 270(15):1826–31, 1993.
- [464] P. Smit, D. Oberholzer, S. Hayden-Smith, H. J. Koornhof, and M. R. Hilleman. Protective efficacy of pneumococcal polysaccharide vaccines. *JAMA*, 238(24):2613–6, 1977.
- [465] US Food and Drug Administration. Pneumovax 23 prescribing information. 2014.
- [466] A. J. Pollard, K. P. Perrett, and P. C. Beverley. Maintaining protection against invasive bacteria with protein-polysaccharide conjugate vaccines. *Nat Rev Immunol*, 9(3):213–20, 2009.
- [467] M. A. Westerink, Jr. Schroeder, H. W., and M. H. Nahm. Immune responses to pneumococcal vaccines in children and adults: Rationale for age-specific vaccination. *Aging and Disease*, 3(1):51–67, 2012.
- [468] US Food and Drug Administration. Prevnar 13 prescribing information. 2016.
- [469] M. J. Bonten, S. M. Huijts, M. Bolkenbaas, C. Webber, S. Patterson, S. Gault, C. H. van Werkhoven, A. M. van Deursen, E. A. Sanders, T. J. Verheij, M. Patton, A. McDonough, A. Moradoghli-Haftvani, H. Smith, T. Mellelieu, M. W. Pride, G. Crowther, B. Schmoelle-Thoma, D. A. Scott, K. U. Jansen, R. Lobatto, B. Oosterman, N. Visser, E. Caspers, A. Smorenburg, E. A. Emini, W. C. Gruber, and D. E. Grobbee. Polysaccharide conjugate vaccine against pneumococcal pneumonia in adults. *N Engl J Med*, 372(12):1114–25, 2015.
- [470] A. H. van den Biggelaar, W. Pomat, A. Bosco, S. Phuanukoonnon, C. J. Devitt, M. A. Nadal-Sims, P. M. Siba, P. C. Richmond, D. Lehmann, and P. G. Holt. Pneumococcal conjugate vaccination at birth in a high-risk setting: no evidence for neonatal t-cell tolerance. *Vaccine*, 29(33):5414–20, 2011.
- [471] Nick J Andrews, Pauline A Waight, Polly Burbidge, Emma Pearce, Lucy Roalfe, Marta Zancolli, Mary Slack, Shamez N Ladhani, Elizabeth Miller, and David Goldblatt. Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study. *The Lancet infectious diseases*, 14(9):839–846, 2014.

- [472] David L. Woodland. Jump-starting the immune system: prime–boosting comes of age. *Trends in Immunology*, 25(2):98–104, 2004.
- [473] Daniel M. Musher, Adriana M. Rueda, Moon H. Nahm, Edward A. Graviss, and Maria C. Rodriguez-Barradas. Response to pneumococcal polysaccharide and protein-conjugate vaccines singly or sequentially in adults who have recovered from pneumococcal pneumonia. *The Journal of infectious diseases*, 198(7):1019–1027, 2008.
- [474] Abdulmonam Ali, Amal Milad, Mohammad Taleb, Yousef Al Ahwel, Rahman Shehnaz, and Dan Olson. Use of pcv13 and ppsv23 in adult population: Are we following the advisory committee on immunization practices recommendations? *Chest*, 150(4, Supplement):604A, 2016.
- [475] Shan Lu. Heterologous prime-boost vaccination. *Current opinion in immunology*, 21(3):346–351, 2009.
- [476] Centers for Disease Control and Prevention. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among children aged 6–18 years with immunocompromising conditions: recommendations of the advisory committee on immunization practices (acip). *MMWR. Morbidity and mortality weekly report*, 62(25):521, 2013.
- [477] Centers for Disease Control and Prevention. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged 65 years: recommendations of the advisory committee on immunization practices (acip). *MMWR. Morbidity and mortality weekly report*, 63(37):822, 2014.
- [478] C. Sadlier, S. O’Dea, K. Bennett, J. Dunne, N. Conlon, and C. Bergin. Immunological efficacy of pneumococcal vaccine strategies in hiv-infected adults: a randomized clinical trial. *Scientific Reports*, 6:32076, 2016.
- [479] D. M. Weinberger, Z. B. Harboe, and E. D. Shapiro. Developing better pneumococcal vaccines for adults. *JAMA Internal Med*, 177(3):303–304, 2017.
- [480] Centers for Disease Control and Prevention. Updated recommendations for prevention of invasive pneumococcal disease among adults using the 23-valent pneumococcal polysaccharide vaccine (ppsv23). *Morbidity and Mortality Weekly Report*, 59(34):1102–6, 2010.
- [481] T. Shiri, S. Datta, J. Madan, A. Tsertsvadze, P. Royle, M. J. Keeling, N. D. McCarthy, and S. Petrou. Indirect effects of childhood pneumococcal conjugate vaccination on invasive pneumococcal disease: a systematic review and meta-analysis. *Lancet Global Health*, 5(1):e51–e59, 2017.
- [482] M. Suzuki, B. G. Dhoubhadel, T. Ishifuji, M. Yasunami, M. Yaegashi, N. Asoh, M. Ishida, S. Hamaguchi, M. Aoshima, K. Ariyoshi, K. Morimoto, and Group-Japan Adult Pneumonia Study. Serotype-specific effectiveness of 23-valent pneumococcal polysaccharide vaccine against pneumococcal pneumonia in adults aged 65 years or older: a multicentre, prospective, test-negative design study. *Lancet Infect Dis*, 2017.

- [483] C. Y. Lu and L. M. Huang. Is pneumococcal serotype replacement impending? *Pediatrics and Neonatology*, 57(5):363–364, 2016.
- [484] Richard McFetridge, Ajoke Sobanjo-ter Meulen, Steven D Folkerth, John A Hoekstra, Michael Dallas, Patricia A Hoover, Rocio D Marchese, Donna M Zacholski, Wendy J Watson, and Jon E Stek. Safety, tolerability, and immunogenicity of 15-valent pneumococcal conjugate vaccine in healthy adults. *Vaccine*, 33(24):2793–2799, 2015.
- [485] Tamara Pilishvili, Catherine Lexau, Monica M Farley, James Hadler, Lee H Harrison, Nancy M Bennett, Arthur Reingold, Ann Thomas, William Schaffner, and Allen S Craig. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *Journal of Infectious Diseases*, 201(1):32–41, 2010.
- [486] JT Granton and RF Grossman. Community-acquired pneumonia in the elderly patient. clinical features, epidemiology, and treatment. *Clinics in chest medicine*, 14(3):537, 1993.
- [487] L. P. Hayden, B. D. Hobbs, R. T. Cohen, R. A. Wise, W. Checkley, J. D. Crapo, C. P. Hersh, and C. OPDGene Investigators. Childhood pneumonia increases risk for chronic obstructive pulmonary disease: the copdgene study. *Respiratory Research*, 16:115, 2015.
- [488] Karalanglin Tiewsoh, Rakesh Lodha, Ravindra M Pandey, Shobha Broor, M Kalaivani, and Sushil K Kabra. Factors determining the outcome of children hospitalized with severe pneumonia. *BMC pediatrics*, 9(1):15, 2009.
- [489] Arto A. Palmu, Jukka Jokinen, Heta Nieminen, Hanna Rinta-Kokko, Esa Ruokokoski, Taneli Puumalainen, Marta Moreira, Lode Schuerman, Dorota Borys, and Terhi M. Kilpi. Vaccine-preventable disease incidence of pneumococcal conjugate vaccine in the finnish invasive pneumococcal disease vaccine trial. *Vaccine*, 36(14):1816–1822, 2018.
- [490] J. E. Bourcier, J. Paquet, M. Seinger, E. Gallard, J. P. Redonnet, F. Cheddadi, D. Garnier, J. M. Bourgeois, and T. Geeraerts. Performance comparison of lung ultrasound and chest x-ray for the diagnosis of pneumonia in the ed. *Am J Emerg Med*, 32(2):115–8, 2014.
- [491] Miguel A Chavez, Navid Shams, Laura E Ellington, Neha Naithani, Robert H Gilman, Mark C Steinhoff, Mathuram Santosham, Robert E Black, Carrie Price, and Margaret Gross. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. *Respiratory research*, 15(1):50, 2014.
- [492] Ling Long, Hao-Tian Zhao, Zhi-Yang Zhang, Guang-Ying Wang, and He-Ling Zhao. Lung ultrasound for the diagnosis of pneumonia in adults: A meta-analysis. *Medicine*, 96(3), 2017.
- [493] Sandra K Al-Tarawneh, Michael B Border, Christopher F Dibble, and Sompop Bencharit. Defining salivary biomarkers using mass spectrometry-based proteomics: a systematic review. *Omics: a journal of integrative biology*, 15(6):353–361, 2011.
- [494] R. Schnabel, R. Fijten, A. Smolinska, J. Dallinga, M. L. Boumans, E. Stobberingh, A. Boots, P. Roekaerts, D. Bergmans, and F. J. van Schooten. Analysis of volatile organic compounds

- in exhaled breath to diagnose ventilator-associated pneumonia. *Nature Scientific Reports*, 5:17179, 2015.
- [495] J. P. Leeming, K. Cartwright, R. Morris, S. A. Martin, M. D. Smith, and Group South-West Pneumococcus Study. Diagnosis of invasive pneumococcal infection by serotype-specific urinary antigen detection. *J Clin Microbiol*, 43(10):4972–6, 2005.
 - [496] Amar Safdar, Samuel A Shelburne, Scott E Evans, and Burton F Dickey. Inhaled therapeutics for prevention and treatment of pneumonia. *Expert opinion on drug safety*, 8(4):435–449, 2009.
 - [497] A. C. Berical, D. Harris, C. S. Dela Cruz, and J. D. Possick. Pneumococcal vaccination strategies. an update and perspective. *Annals of the American Thoracic Society*, 13(6):933–44, 2016.
 - [498] Centers for Disease Control and Prevention. Vaccine price list. Report, 2017.
 - [499] Lucia H Lee, Xin-Xing Gu, and Moon H Nahm. Towards new broader spectrum pneumococcal vaccines: the future of pneumococcal disease prevention. *Vaccines*, 2(1):112–128, 2014.
 - [500] K. Moffitt and R. Malley. Rationale and prospects for novel pneumococcal vaccines. *Human Vaccines and Immunotherapeutics*, 12(2):383–92, 2016.
 - [501] Gail L Rodgers and Keith P Klugman. The future of pneumococcal disease prevention. *Vaccine*, 29:C43–C48, 2011.
 - [502] Calvin C Daniels, P David Rogers, and Chasity M Shelton. A review of pneumococcal vaccines: Current polysaccharide vaccine recommendations and future protein antigens. *The Journal of Pediatric Pharmacology and Therapeutics*, 21(1):27–35, 2016.
 - [503] Sharon E Frey, Kathleen R Lottenbach, Heather Hill, Tamara P Blevins, Yinyi Yu, Ying Zhang, Karen E Brennenman, Sandra M Kelly-Aehle, Caitlin McDonald, and Angela Jansen. A phase i, dose-escalation trial in adults of three recombinant attenuated salmonella typhi vaccine vectors producing streptococcus pneumoniae surface protein antigen pspa. *Vaccine*, 31(42):4874–4880, 2013.
 - [504] Thierry Kamtchoua, Monica Bologa, Robert Hopfer, David Neveu, Branda Hu, Xiaohua Sheng, Nicolas Corde, Catherine Pouzet, Gloria Zimmermann, and Sanjay Gurunathan. Safety and immunogenicity of the pneumococcal pneumolysin derivative plyd1 in a single-antigen protein vaccine candidate in adults. *Vaccine*, 31(2):327–333, 2013.
 - [505] F. Khan, M. A. Khan, N. Ahmed, M. I. Khan, H. Bashir, S. Tahir, and A. U. Zafar. Molecular characterization of pneumococcal surface protein a (pspa), serotype distribution and antibiotic susceptibility of streptococcus pneumoniae strains isolated from pakistan. *Infect Dis Ther*, 2018.
 - [506] Amy Sarah Ginsburg, Moon H Nahm, Farukh M Khambaty, and Mark R Alderson. Issues and challenges in the development of pneumococcal protein vaccines. *Expert review of vaccines*, 11(3):279–285, 2012.

- [507] Harm HogenEsch, Anisa Dunham, Bethany Hansen, Kathleen Anderson, Jean-Francois Maisonneuve, and Stanley L Hem. Formulation of a killed whole cell pneumococcus vaccine-effect of aluminum adjuvants on the antibody and il-17 response. *Journal of immune based therapies and vaccines*, 9(1):5, 2011.
- [508] Tamara Pilishvili and Nancy M Bennett. Pneumococcal disease prevention among adults: strategies for the use of pneumococcal vaccines. *Vaccine*, 33:D60–D65, 2015.
- [509] X. Zhang, J. Cui, Y. Wu, H. Wang, J. Wang, Y. Qiu, Y. Mo, Y. He, X. Zhang, Y. Yin, and W. Xu. Streptococcus pneumoniae attenuated strain spy1 with an artificial mineral shell induces humoral and th17 cellular immunity and protects mice against pneumococcal infection. *Front Immunol*, 8:1983, 2017.
- [510] Xiuyu Xu, Hong Wang, Yusi Liu, Yiping Wang, Lingbing Zeng, Kaifeng Wu, Jianmin Wang, Feng Ma, Wenchun Xu, Yibing Yin, and Xuemei Zhang. Mucosal immunization with the live attenuated vaccine spy1 induces humoral and th2-th17-regulatory t cell cellular immunity and protects against pneumococcal infection. *Infection and Immunity*, 83(1):90–100, 2015.
- [511] Mayo Clinic Staff. Bronchitis, 2019, (Accessed: 2019-06-02). Available at <https://www.mayoclinic.org/diseases-conditions/bronchitis/symptoms-causes/syc-20355566>.
- [512] Mayo Clinic Staff. Emphysema, 2019, (Accessed: 2019-06-02). Available at <https://www.mayoclinic.org/diseases-conditions/emphysema/symptoms-causes/syc-20355555>.
- [513] American Lung Association. Chronic obstructive pulmonary disease (copd), 2019, (Accessed: 2019-06-02). Available at <https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/copd>.
- [514] World Health Organization. Chronic obstructive pulmonary disease (copd), 2019, (Accessed: 2019-06-02). Available at <https://www.who.int/respiratory/copd/en/>.
- [515] World Health Organization. Chronic obstructive pulmonary disease (copd), 2017, (Accessed: 2019-06-02). Available at [https://www.who.int/en/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/en/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)).
- [516] Shireen Mirza, Ryan D. Clay, Matthew A. Koslow, and Paul D. Scanlon. Copd guidelines: A review of the 2018 gold report. *Mayo Clinic Proceedings*, 93(10):1488 – 1502, 2018.
- [517] Peter J. Barnes, Peter G. J. Burney, Edwin K. Silverman, Bartolome R. Celli, Jørgen Vestbo, Jadwiga A. Wedzicha, and Emiel F. M. Wouters. Chronic obstructive pulmonary disease. *Nature Reviews Disease Primers*, 1(1), Dec 2015.
- [518] Watz H Rabe KF. Chronic obstructive pulmonary disease. *The lancet*, 389:1931–1940, 2017.
- [519] SA Quaderi and JR Hurst. The unmet global burden of copd. *Global health, epidemiology and genomics*, 3, 2018.

- [520] Marina Saetta, Antonino Di Stefano, Piero Maestrelli, Alberto Ferraresso, Riccardo Drigo, Alfredo Potena, Adalberto Ciaccia, and Leonardo M Fabbri. Activated t-lymphocytes and macrophages in bronchial mucosa of subjects with chronic bronchitis. *American review of respiratory disease*, 147:301–301, 1993.
- [521] Rafael Laniado-Laborín. Smoking and chronic obstructive pulmonary disease (copd). parallel epidemics of the 21st century. *International journal of environmental research and public health*, 6(1):209–224, 2009.
- [522] A Agusti, W MacNee, K Donaldson, and M Cosio. Hypothesis: does copd have an autoimmune component?, 2003.
- [523] Steven R Rutgers, Dirkje S Postma, Nick HT ten Hacken, Henk F Kauffman, Thomas W van der Mark, Gerard H Koëter, and Wim Timens. Ongoing airway inflammation in patients with copd who do not currently smoke. *Thorax*, 55(1):12–18, 2000.
- [524] Timothy M Bahr, Grant J Hughes, Michael Armstrong, Rick Reisdorph, Christopher D Coldren, Michael G Edwards, Christina Schnell, Ross Kedl, Daniel J LaFlamme, Nichole Reisdorph, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *American journal of respiratory cell and molecular biology*, 49(2):316–323, 2013.
- [525] Yale Chang, Kimberly Glass, Yang-Yu Liu, Edwin K Silverman, James D Crapo, Ruth Tal-Singer, Russ Bowler, Jennifer Dy, Michael Cho, and Peter Castaldi. Copd subtypes identified by network-based clustering of blood gene expression. *Genomics*, 107(2-3):51–58, 2016.
- [526] Lavidia RK Brooks and George I Mias. Data-driven analysis of age, sex, and tissue effects on gene expression variability in alzheimer’s disease. *Frontiers in Neuroscience*, 13:392, 2019.
- [527] Timothy M Bahr, Grant J Hughes, Michael Armstrong, Rick Reisdorph, Christopher D Coldren, Michael G Edwards, Christina Schnell, Ross Kedl, Daniel J LaFlamme, Nichole Reisdorph, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *American journal of respiratory cell and molecular biology*, 49(2):316–323, 2013.
- [528] Nick Fishbane, Yunlong Nie, Virginia Chen, Zsuzsanna Hollander, Scott J Tebbutt, Yohan Bossé, Raymond T Ng, Bruce E Miller, Bruce McManus, Stephen Rennard, et al. The effect of statins on blood gene expression in copd. *PloS one*, 10(10):e0140022, 2015.
- [529] Dave Singh, Steven M Fox, Ruth Tal-Singer, Stewart Bates, John H Riley, and Bartolome Celli. Altered gene expression in blood and sputum in copd frequent exacerbators in the eclipse cohort. *PloS one*, 9(9):e107381, 2014.
- [530] F Martin, M Talikka, J Hoeng, and MC Peitsch. Identification of gene expression signature for cigarette smoke exposure response—from man to mouse. *Human & experimental toxicology*, 34(12):1200–1211, 2015.

- [531] Subhashini Arimilli, Behrouz Madahian, Peter Chen, Kristin Marano, and GL Prasad. Gene expression profiles associated with cigarette smoking and moist snuff consumption. *BMC genomics*, 18(1):156, 2017.
- [532] Sunirmal Paul and Sally A Amundson. Differential effect of active smoking on gene expression in male and female smokers. *Journal of carcinogenesis & mutagenesis*, 5, 2014.
- [533] Sunirmal Paul and Sally A Amundson. Gene expression signatures of radiation exposure in peripheral white blood cells of smokers and non-smokers. *International journal of radiation biology*, 87(8):791–801, 2011.
- [534] Benilton S. Carvalho and Rafael A. Irizarry. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, 26(19):2363–2367, 08 2010.
- [535] James W. MacDonald. affycoretools: Functions useful for those doing repetitive analyses with affymetrix genechips, 2018, (Accessed: 2019-03-30). Available at <https://www.bioconductor.org/packages/release/bioc/html/affycoretools.html/>, Accessed: 2019-03-30.
- [536] Matthew E. Ritchie, Jeremy Silver, Alicia Oshlack, Melissa Holmes, Dileepa Diyagama, Andrew Holloway, and Gordon K. Smyth. A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, 23(20):2700–2707, 08 2007.
- [537] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.
- [538] George Mias. *Mathematica for Bioinformatics: A Wolfram Language Approach to Omics*, chapter Chapter 9: Machine Learning, pages 283–296. Springer International Publishing, Cham, 2018.
- [539] Megan Hardin and Edwin K Silverman. Chronic obstructive pulmonary disease genetics: a review of the past and a look into the future. *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation*, 1(1):33, 2014.
- [540] Melvin Berger, Bob Geng, D William Cameron, Ladonna M Murphy, and Edward S Schulman. Primary immune deficiency diseases as unrecognized causes of chronic respiratory disease. *Respiratory medicine*, 132:181–188, 2017.
- [541] Karthik D Nath, Julie G Burel, Viswanathan Shankar, Antonia L Pritchard, Michelle Towers, David Looke, Janet M Davies, and John W Upham. Clinical factors associated with the humoral immune response to influenza vaccination in chronic obstructive pulmonary disease. *International journal of chronic obstructive pulmonary disease*, 9:51, 2014.
- [542] Brian N McCullagh, Alejandro P Comellas, Zuhair K Ballas, John D Newell Jr, M Bridget Zimmerman, and Antoine E Azar. Antibody deficiency in patients with frequent exacerbations of chronic obstructive pulmonary disease (copd). *PloS one*, 12(2):e0172437, 2017.

- [543] Byunghyuk Lee, Eunhee Ko, Jiyeon Lee, Yuna Jo, Hyunju Hwang, Tae Sik Goh, Myungsoo Joo, and Changwan Hong. Soluble common gamma chain exacerbates copd progress through the regulation of inflammatory t cell response in mice. *International journal of chronic obstructive pulmonary disease*, 12:817, 2017.
- [544] Jonas F Ludvigsson, Martin Neovius, and Lennart Hammarström. Risk of infections among 2100 individuals with iga deficiency: a nationwide cohort study. *Journal of clinical immunology*, 36(2):134–140, 2016.
- [545] Paola Panina-Bordignon, Alberto Papi, Margherita Mariani, Pietro Di Lucia, Gianluca Casoni, Cinzia Bellettato, Cecilia Buonsanti, Deborah Miotto, Cristina Mapp, Antonello Villa, et al. The cc chemokine receptors ccr4 and ccr8 identify airway t cells of allergen-challenged atopic asthmatics. *The Journal of clinical investigation*, 107(11):1357–1364, 2001.
- [546] Martina Kvist Reimer, Charlotte Brange, and Alexander Rosendahl. Ccr8 signaling influences toll-like receptor 4 responses in human macrophages in inflammatory diseases. *Clin. Vaccine Immunol.*, 18(12):2050–2059, 2011.
- [547] Yasuo Sekine, Hideki Katsura, Eitetsu Koh, Kenzo Hiroshima, and Takehiko Fujisawa. Early detection of copd is important for lung cancer surveillance. *European Respiratory Journal*, 39(5):1230–1240, 2012.
- [548] Mingxing Yang, Maxie Kohler, Tina Heyder, Helena Forsslund, Hilde K Garberg, Reza Karimi, Johan Grunewald, Frode S Berven, Sven Nyrén, C Magnus Sköld, et al. Proteomic profiling of lung immune cells reveals dysregulation of phagocytotic pathways in female-dominated molecular copd phenotype. *Respiratory research*, 19(1):39, 2018.
- [549] Renat Shaykhiev, Fouad Otaki, Prince Bonsu, David T Dang, Matthew Teater, Yael Strulovici-Barel, Jacqueline Salit, Ben-Gary Harvey, and Ronald G Crystal. Cigarette smoking reprograms apical junctional complex molecular architecture in the human airway epithelium in vivo. *Cellular and Molecular Life Sciences*, 68(5):877–892, 2011.
- [550] Jennifer L Perret, Melanie C Matheson, Lyle C Gurrin, David P Johns, John A Burgess, Bruce R Thompson, Adrian J Lowe, James Markos, Stephen S Morrison, Christine F McDonald, et al. Childhood measles contributes to post-bronchodilator airflow obstruction in middle-aged adults: A cohort study. *Respirology*, 23(8):780–787, 2018.
- [551] Kim Hoenderdos and Alison Condliffe. The neutrophil in chronic obstructive pulmonary disease. too little, too late or too much, too soon? *American journal of respiratory cell and molecular biology*, 48(5):531–539, 2013.
- [552] Irfan Rahman. Antioxidant therapies in copd. *International journal of chronic obstructive pulmonary disease*, 1(1):15, 2006.
- [553] Leah Persaud, Dayenny De Jesus, Oliver Brannigan, Maria Richiez-Paredes, Jeannette Huanman, Giselle Alvarado, Linda Riker, Gisette Mendez, Jordan Dejoie, and Moira Sauane. Mechanism of action and applications of interleukin 24 in immunotherapy. *International journal of molecular sciences*, 17(6):869, 2016.

- [554] Haixia Zhou, Angelika Brekman, Wu-Lin Zuo, Xuemei Ou, Renat Shaykhiev, Francisco J Agosto-Perez, Rui Wang, Matthew S Walters, Jacqueline Salit, Yael Strulovici-Barel, et al. Pou2af1 functions in the human airway epithelium to regulate expression of host defense genes. *The Journal of Immunology*, 196(7):3159–3167, 2016.
- [555] Cécile M Bidan, Annemiek C Veldsink, Herman Meurs, and Reinoud Gosens. Airway and extracellular matrix mechanics in copd. *Frontiers in physiology*, 6:346, 2015.
- [556] James C Hogg and Wim Timens. The pathology of chronic obstructive pulmonary disease. *Annual Review of Pathological Mechanical Disease*, 4:435–459, 2009.
- [557] Peter J Barnes. Sex differences in chronic obstructive pulmonary disease mechanisms, 2016.
- [558] Shambhu Aryal, Enrique Diaz-Guzman, and David M Mannino. Copd and gender differences: an update. *Translational Research*, 162(4):208–218, 2013.
- [559] Mayo Clinic Staff. Chronic obstructive pulmonary disease (copd), 2019, (Accessed: 2019-06-02). Available at <https://www.mayoclinic.org/diseases-conditions/copd/symptoms-causes/syc-20353679>.
- [560] Yoichiro Kaku, Haruki Imaoka, Yoshitaka Morimatsu, Yoshihiro Komohara, Koji Ohnishi, Hanako Oda, Shinichi Takenaka, Masanobu Matsuoka, Tomotaka Kawayama, Motohiro Takeya, et al. Overexpression of cd163, cd204 and cd206 on alveolar macrophages in the lungs of patients with severe chronic obstructive pulmonary disease. *PloS one*, 9(1):e87400, 2014.
- [561] Katrina Steiling, Marc E Lenburg, and Avrum Spira. Airway gene expression in chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 6(8):697–700, 2009.
- [562] Veronika Cheplygina, Isabel Pino Pena, Jesper Holst Pedersen, David A Lynch, Lauge Sørensen, and Marleen de Bruijne. Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE journal of biomedical and health informatics*, 22(5):1486–1496, 2017.
- [563] Cristóbal Esteban, Javier Moraza, Cristóbal Esteban, Fernando Sancho, Myriam Aburto, Amaia Aramburu, Begona Goiria, Amaia Garcia-Loizaga, and Alberto Capelastegui. Machine learning for copd exacerbation prediction. *European Respiratory Journal*, 46(suppl 59), 2015.
- [564] Habiboulaye Amadou Boubacar and Joëlle Texereau. Ensemble machine learning for the early detection of copd exacerbations. *European Respiratory Journal*, 50(suppl 61), 2017.
- [565] Jeffery K Taubenberger and David M Morens. The pathology of influenza virus infections. *Annu. Rev. pathmechdis. Mech. Dis.*, 3:499–522, 2008.
- [566] Satoshi Fukuyama and Yoshihiro Kawaoka. The pathogenesis of influenza virus infections: the contributions of virus and host factors. *Current opinion in immunology*, 23(4):481–486, 2011.

- [567] Hembly Rivas, Summer Schmaling, and Marta Gaglia. Shutoff of host gene expression in influenza a virus and herpesviruses: similar mechanisms and common themes. *Viruses*, 8(4):102, 2016.
- [568] Yijie Zhai, Luis M Franco, Robert L Atmar, John M Quarles, Nancy Arden, Kristine L Bucasas, Janet M Wells, Diane Nino, Xueqing Wang, Gladys E Zapata, et al. Host transcriptional response to influenza and other acute respiratory viral infections—a prospective cohort study. *PLoS pathogens*, 11(6):e1004869, 2015.
- [569] Robert M Krug. Functions of the influenza a virus ns1 protein in antiviral defense. *Current opinion in virology*, 12:1–6, 2015.
- [570] Ham Ching Lam, Xuan Bi, Srinand Sreevatsan, and Daniel Boley. Evolution and vaccination of influenza virus. *Journal of Computational Biology*, 24(8):787–798, 2017.
- [571] Slobodan Paessler and Veljko Veljkovic. Using electronic biology based platform to predict flu vaccine efficacy for 2018/2019. *F1000Research*, 7, 2018.
- [572] Lisa A Grohskopf, Leslie Z Sokolow, Karen R Broder, Emmanuel B Walter, Alicia M Fry, and Daniel B Jernigan. Prevention and control of seasonal influenza with vaccines: recommendations of the advisory committee on immunization practices—united states, 2018–19 influenza season. *MMWR Recommendations and Reports*, 67(3):1, 2018.
- [573] Cincy Tech. Blue water vaccines, developing universal flu vaccine, closes \$7 million led by cincytech. 2019. Available at https://www.eurekalert.org/pub_releases/2019-07/c-bwv070519.php, Accessed: 2019-07-09.
- [574] Business Wire. Vaxart enters into research collaboration with janssen to evaluate oral universal influenza vaccine. 2019. Available at <https://www.businesswire.com/news/home/20190709005100/en/>, Accessed: 2019-07-09.
- [575] Margarita M Gomez Lorenzo and Matthew J Fenton. Immunobiology of influenza vaccines. *Chest*, 143(2):502–510, 2013.
- [576] Peter C Soema, Ronald Kompier, Jean-Pierre Amorij, and Gideon FA Kersten. Current and next generation influenza vaccines: formulation and production strategies. *European Journal of Pharmaceutics and Biopharmaceutics*, 94:251–263, 2015.
- [577] Asuncion Mejias, Blerta Dimo, Nicolas M Suarez, Carla Garcia, M Carmen Suarez-Arrabal, Tuomas Jartti, Derek Blankenship, Alejandro Jordan-Villegas, Monica I Ardura, Zhaohui Xu, et al. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. *PLoS medicine*, 10(11):e1001549, 2013.
- [578] Ioannis Ioannidis, Beth McNally, Meredith Willette, Mark E Peeples, Damien Chaussabel, Joan E Durbin, Octavio Ramilo, Asuncion Mejias, and Emilio Flaño. Plasticity and virus specificity of the airway epithelial cell immune response during respiratory virus infection. *Journal of virology*, 86(10):5422–5436, 2012.

- [579] Vipin Narang, Yanxia Lu, Crystal Tan, Xavier Francois Noel Camous, Shwe Zin Nyunt, Christophe Carre, Esther Wing Hei Mok, Glenn Wong, Brian Abel, Nicolas Burdin, et al. Influenza vaccine-induced antibody responses are not impaired by frailty in the community-dwelling elderly with natural influenza exposure. *Frontiers in immunology*, 9:2465, 2018.
- [580] Jake Dunning, Simon Blankley, Long T Hoang, Mike Cox, Christine M Graham, Philip L James, Chloe I Bloom, Damien Chaussabel, Jacques Banchereau, Stephen J Brett, et al. Progression of whole-blood transcriptional signatures from interferon-induced to neutrophil-associated patterns in severe influenza. 2018.
- [581] Jan-Erik Berdal, Tom E Mollnes, Torgun Wæhre, Ole K Olstad, Bente Halvorsen, Thor Ueland, Jon H Laake, May T Furuseth, Anne Maagaard, Harald Kjekshus, et al. Excessive innate immune response and mutant d222g/n in severe a (h1n1) pandemic influenza. *Journal of Infection*, 63(4):308–316, 2011.
- [582] Helder I Nakaya, Jens Wrämmert, Eva K Lee, Luigi Racioppi, Stephanie Marie-Kunze, W Nicholas Haining, Anthony R Means, Sudhir P Kasturi, Nooruddin Khan, Gui-Mei Li, et al. Systems biology of vaccination for seasonal influenza in humans. *Nature immunology*, 12(8):786, 2011.
- [583] John S Tsang, Pamela L Schwartzberg, Yuri Kotliarov, Angelique Biancotto, Zhi Xie, Ronald N Germain, Ena Wang, Matthew J Olnes, Manikandan Narayanan, Hana Golding, et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell*, 157(2):499–513, 2014.
- [584] Gerlinde Obermoser, Scott Presnell, Kelly Domico, Hui Xu, Yuanyuan Wang, Esperanza Anguiano, LuAnn Thompson-Snipes, Rajaram Ranganathan, Brad Zeitner, Anna Bjork, et al. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity*, 38(4):831–844, 2013.
- [585] Mitsuru Tsuge, Takashi Oka, Nobuko Yamashita, Yukie Saito, Yosuke Fujii, Yoshiharu Nagaoka, Masato Yashiro, Hirokazu Tsukahara, and Tsuneo Morishima. Gene expression analysis in children with complex seizures due to influenza a (h1n1) pdm09 or rotavirus gastroenteritis. *Journal of neurovirology*, 20(1):73–84, 2014.
- [586] Raquel G Cao, Nicolas M Suarez, Gerlinde Obermoser, Santiago MC Lopez, Emilio Flano, Sara E Mertz, Randy A Albrecht, Adolfo García-Sastre, Asuncion Mejias, Hui Xu, et al. Differences in antibody responses between trivalent inactivated influenza vaccine and live attenuated influenza vaccine correlate with the kinetics and magnitude of interferon signaling in children. *The Journal of infectious diseases*, 210(2):224–233, 2014.
- [587] Helder I Nakaya, Thomas Hagan, Sai S Duraisingham, Eva K Lee, Marcin Kwissa, Nadine Rouphael, Daniela Frasca, Merril Gersten, Aneesh K Mehta, Renaud Gaujoux, et al. Systems analysis of immunity to influenza vaccination across multiple years and in diverse populations reveals shared molecular signatures. *Immunity*, 43(6):1186–1198, 2015.
- [588] Meghali Goswami, Gabrielle Prince, Angelique Biancotto, Susan Moir, Lela Kardava, Brian H Santich, Foo Cheung, Yuri Kotliarov, Jinguo Chen, Rongye Shi, et al. Impaired b cell

- immunity in acute myeloid leukemia patients after chemotherapy. *Journal of translational medicine*, 15(1):155, 2017.
- [589] David Furman, Vladimir Jojic, Brian Kidd, Shai Shen-Orr, Jordan Price, Justin Jarrell, Tiffany Tse, Huang Huang, Peder Lund, Holden T Maecker, et al. Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Molecular systems biology*, 9(1):659, 2013.
- [590] Juilee Thakar, Subhasis Mohanty, A Phillip West, Samit R Joshi, Ikuyo Ueda, Jean Wilson, Hailong Meng, Tamara P Blevins, Sui Tsang, Mark Trentalange, et al. Aging-dependent alterations in gene expression and a mitochondrial signature of responsiveness to human influenza vaccination. *Aging (Albany NY)*, 7(1):38, 2015.
- [591] Luis M Franco, Kristine L Bucasas, Janet M Wells, Diane Niño, Xueqing Wang, Gladys E Zapata, Nancy Arden, Alexander Renwick, Peng Yu, John M Quarles, et al. Integrative genomic analysis of the human immune response to influenza vaccination. *Elife*, 2:e00299, 2013.
- [592] Vladimir Brusic, Raphael Gottardo, Steven H Kleinstein, Mark M Davis, David A Hafler, Helen Quill, A Karolina Palucka, Gregory A Poland, Bali Pulendran, Ellis L Reinherz, et al. Computational resources for high-dimensional immune analysis from the human immunology project consortium. *Nature biotechnology*, 32(2):146, 2014.
- [593] Sanchita Bhattacharya, Patrick Dunn, Cristel G Thomas, Barry Smith, Henry Schaefer, Jieming Chen, Zicheng Hu, Kelly A Zalocusky, Ravi D Shankar, Shai S Shen-Orr, et al. Import, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific data*, 5:180015, 2018.
- [594] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Fabian Scheipl, and Gabor Grothendieck. Package ‘lme4’. 2019. Available at <https://cran.r-project.org/web/packages/lme4/lme4.pdf>, Accessed: 2019-06-09.
- [595] Catharine I Paules, Sheena G Sullivan, Kanta Subbarao, and Anthony S Fauci. Chasing seasonal influenza—the need for a universal influenza vaccine. *New England Journal of Medicine*, 378(1):7–9, 2018.
- [596] Jun-Xia Wang and Peter Nigrovic. Cd177 participates in a novel mechanism for regulating neutrophil recruitment (p3093). *The Journal of Immunology*, 190(1 Supplement):43.9–43.9, 2013.
- [597] Yong Yu, Cui Wang, Simon Clare, Juexuan Wang, Song-Choon Lee, Cordelia Brandt, Shannon Burke, Liming Lu, Daqian He, Nancy A Jenkins, et al. The transcription factor bcl11b is specifically expressed in group 2 innate lymphoid cells and is essential for their development. *Journal of Experimental Medicine*, 212(6):865–874, 2015.
- [598] Dorothée Moisy, Sergiy V Avilov, Yves Jacob, Brid M Laoide, Xingyi Ge, Florence Baudin, Nadia Naffakh, and Jean-Luc Jestin. Hmgb1 protein binds to influenza virus nucleoprotein and promotes viral replication. *Journal of virology*, 86(17):9122–9133, 2012.

- [599] Stephan Ludwig and Oliver Planz. Influenza viruses and the nf- κ b signaling pathway—towards a novel concept of antiviral therapy. *Biological chemistry*, 389(10):1307–1312, 2008.
- [600] Naveen Kumar, Zhong-tao Xin, Yuhong Liang, Hinh Ly, and Yuying Liang. Nf- κ b signaling differentially regulates influenza virus rna synthesis. *Journal of virology*, 82(20):9880–9889, 2008.
- [601] Anwar M Hashem, Caroline Gravel, Ze Chen, Yinglei Yi, Monika Tocchi, Bozena Jaentschke, Xingliang Fan, Changgui Li, Michael Rosu-Myles, Alexander Pereboev, et al. Cd40 ligand preferentially modulates immune response and enhances protection against influenza virus. *The Journal of Immunology*, 193(2):722–734, 2014.
- [602] Caterina Hatzifoti and Andrew W Heath. Cd40-mediated enhancement of immune responses against three forms of influenza vaccine. *Immunology*, 122(1):98–106, 2007.
- [603] Liset Westera, Alisha M Jennings, Jad Maamary, Martin Schwemmle, Adolfo García-Sastre, and Eric Bortz. Poly-adp ribosyl polymerase 1 (parp1) regulates influenza a virus polymerase. *Advances in virology*, 2019, 2019.
- [604] Yueh-Ming Loo and Michael Gale Jr. Immune signaling by rig-i-like receptors. *Immunity*, 34(5):680–692, 2011.
- [605] Antonio Chiaretti, Silvia Pulitanò, Giovanni Barone, Pietro Ferrara, Valerio Romano, Domenico Capozzi, and Riccardo Riccardi. Il-1 β and il-6 upregulation in children with h1n1 influenza virus infection. *Mediators of inflammation*, 2013, 2013.
- [606] Feng Wen, Jinyue Guo, Zhili Li, and Shujian Huang. Sex-specific patterns of gene expression following influenza vaccination. *Scientific reports*, 8(1):13517, 2018.
- [607] Ronit Avitsur, Jacqueline W Mays, and John F Sheridan. Sex differences in the response to influenza virus infection: modulation by stress. *Hormones and behavior*, 59(2):257–264, 2011.
- [608] Azadeh Bahadoran, Sau H Lee, Seok M Wang, Rishya Manikam, Jayakumar Rajarajeswaran, Chandramathi S Raju, and Shamala D Sekaran. Immune responses to influenza virus and its correlation to age and inherited factors. *Frontiers in microbiology*, 7:1841, 2016.
- [609] Barnaby Young, Sapna Sadarangani, Lili Jiang, Annelies Wilder-Smith, and Mark I-Cheng Chen. Duration of influenza vaccine effectiveness: a systematic review, meta-analysis, and meta-regression of test-negative design case-control studies. *The Journal of infectious diseases*, 217(5):731–741, 2017.
- [610] Maria R Castrucci. Factors affecting immune responses to the influenza vaccine. *Human vaccines & immunotherapeutics*, 14(3):637–646, 2018.
- [611] Lauren C Ramsay, Sarah A Buchan, Robert G Stirling, Benjamin J Cowling, Shuo Feng, Jeffrey C Kwong, and Bryna F Warshawsky. The impact of repeated vaccination on influenza vaccine effectiveness: a systematic review and meta-analysis. *BMC medicine*, 17(1):9, 2019.

- [612] Daniel F Hoft, Kathleen R Lottenbach, Azra Blazevic, Aldin Turan, Tamara P Blevins, Thomas P Pacatte, Yinyi Yu, Michelle C Mitchell, Stella G Hoft, and Robert B Belshe. Comparisons of the humoral and cellular immune responses induced by live attenuated influenza vaccine and inactivated influenza vaccine in adults. *Clin. Vaccine Immunol.*, 24(1):e00414–16, 2017.
- [613] Kristine L Bucasas, Luis M Franco, Chad A Shaw, Molly S Bray, Janet M Wells, Diane Niño, Nancy Arden, John M Quarles, Robert B Couch, and John W Belmont. Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *The Journal of infectious diseases*, 203(7):921–929, 2011.