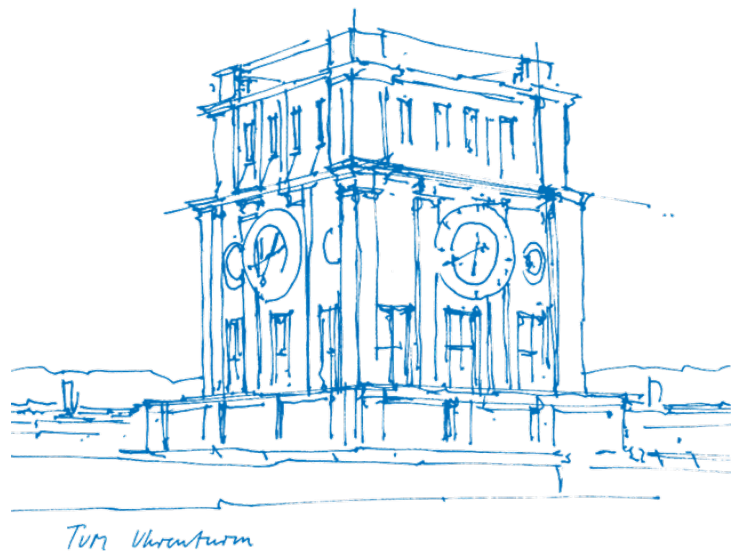


# Thesis title

Subtitle of the thesis

Author Name





# Thesis title

Subtitle of the thesis

Author Name



# Thesis title

Subtitle of the thesis

**Author Name**

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktor-Ingenieurs (Dr.-Ing.)**

genehmigten Dissertation.

**Vorsitzende(r):**

Prof. Franz X. Gabelsberger

**Prüfer der Dissertation:**

1. Prof. Dr. Georg Simon Ohm
2. Prof. James Clerk Maxwell

Die Dissertation wurde am 29.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 11.07.2016 angenommen.



*To Franz X. Gabelsberger, inventor of the street named after him.*





# Abstract

The abstract of your thesis goes here.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.



# Contents

<b>Abstract</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Pathophysiology of chronic lung diseases . . . . .	1
1.1.1 Bronchopulmonary Dysplasia (BPD) . . . . .	1
1.1.2 Asthma . . . . .	1
1.1.3 Chronic Obstructive Pulmonary Disease (COPD) . . . . .	1
1.1.4 Idiopathic Pulmonary Fibrosis (IPF) . . . . .	1
1.2 Computational biology and chronic lung diseases . . . . .	1
1.2.1 Multi-omics data integration . . . . .	1
1.2.2 Clinical prediction . . . . .	1
1.2.3 Systems biology . . . . .	1
1.3 Aims . . . . .	1
<b>2 Methodology</b>	<b>3</b>
2.1 Data gathering . . . . .	4
2.1.1 Mice data . . . . .	4
2.1.2 Human data . . . . .	4
2.1.3 Public data . . . . .	4
2.2 Preprocessing . . . . .	4
2.2.1 Normalization . . . . .	4
2.2.2 Data imputation . . . . .	4
2.2.3 Batch-effect detection . . . . .	4
2.3 Differential expression analysis . . . . .	4
2.3.1 Limma . . . . .	4
2.3.2 DESeq2 . . . . .	4
2.4 Enrichment analysis . . . . .	4
2.4.1 Gene list functional enrichment analysis . . . . .	4
2.5 Multi-omics factor analysis (MOFA) . . . . .	4
2.6 Clinical data correlation . . . . .	5
2.6.1 Linear regression . . . . .	5
2.6.2 Binomial regression . . . . .	5
2.6.3 Ordinal regression . . . . .	5
2.6.4 Multinomial logistic regression . . . . .	5
2.6.5 Dirichlet regression . . . . .	5
2.7 Benchmarking of Lasso models dealing with missing values . . . . .	5
2.7.1 Knowledge guided multi-level network inference . . . . .	5
2.7.2 Two-steps based models . . . . .	5
2.7.3 Inverse covariance based methods . . . . .	5
2.8 Adult data correlation . . . . .	5
2.8.1 Random forest . . . . .	5
2.8.2 t-test, manova, log-reg . . . . .	5
<b>3 Summary of publications</b>	<b>7</b>
<b>4 Discussion</b>	<b>9</b>



# **1 Introduction**

This is the introduction of the thesis.

## **1.1 Pathophysiology of chronical lung diseases**

### **1.1.1 Bronchopulmonary Dysplasia (BPD)**

### **1.1.2 Asthma**

### **1.1.3 Chronic Obstructive Pulmonary Disease (COPD)**

### **1.1.4 Idiopathic Pulmonary Fibrosis (IPF)**

## **1.2 Computational biology and chronic lung diseases**

### **1.2.1 Multi-omics data integration**

### **1.2.2 Clinical prediction**

### **1.2.3 Systems biology**

## **1.3 Aims**



## **2 Methodology**

This is the methodology of the thesis.

## **2.1 Data gathering**

### **2.1.1 Mice data**

Transcriptomics

### **2.1.2 Human data**

Transcriptomics

Metabolomics

### **2.1.3 Public data**

Multi-omics bulk data

Neonatal single-cell transcriptomics

## **2.2 Preprocessing**

### **2.2.1 Normalization**

DESeq2

Pareto scaling

Size-effect

### **2.2.2 Data imputation**

Random-forest

knn-Imputation

### **2.2.3 Batch-effect detection**

Principal component analysis (PCA)

Hierarchical clustering

K-BET

## **2.3 Differential expression analysis**

### **2.3.1 Limma**

### **2.3.2 DESeq2**

## **2.4 Enrichment analysis**

### **2.4.1 Gene list functional enrichment analysis**

## **2.5 Multi-omics factor analysis (MOFA)**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien



est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

## **2.6 Clinical data correlation**

### **2.6.1 Linear regression**

### **2.6.2 Binomial regression**

### **2.6.3 Ordinal regression**

### **2.6.4 Multinomial logistic regression**

### **2.6.5 Dirichlet regression**

## **2.7 Benchmarking of Lasso models dealing with missing values**

### **2.7.1 Knowledge guided multi-level network inference**

### **2.7.2 Two-steps based models**

Grouped adaptive Lasso (GALasso)

Stacked adaptive Lasso (SALasso)

### **2.7.3 Inverse covariance based methods**

Convexed conditioned Lasso (CoCoLasso)

Lasso with high missing rate (HMLasso)

## **2.8 Adult data correlation**

### **2.8.1 Random forest**

Imbalanced random forest

Nested cross-validation in random forest

### **2.8.2 t-test, manova, log-reg**



### 3 Summary of publications

1. Juan Henao, Alida Kindt, Tanja Seegmüller, Kai Foerster, Andreas Flemmer<sup>3</sup>, Juergen Behr<sup>4</sup>, Nikolaus Kneidinger, Marion Frankenberger, Fabian Theis, Benjamin Schubert, Markus List, Anne Hilgendorff. Multi-omic signatures relate to the severity of pulmonary outcome in neonates traced into adult disease.

Summary: kfkfkfkfk

2. Erika Gonzalez Rodriguez<sup>1</sup>, Juan Henao<sup>2</sup>, Motaharehsadat Heydarian<sup>1</sup>, Tina Pritzke<sup>1</sup>, Alida Kindt<sup>3</sup>, Anna M. Dmitrieva<sup>1</sup>, Heiko Adler<sup>4, 5</sup>, Melanie Markmann<sup>6</sup>, Valeria Viteri-Alvarez<sup>1</sup>, Prajakta Oak<sup>1</sup>, Markus Koschlig<sup>1</sup>, Xin Zhang<sup>1</sup>, Kai M. Foerster<sup>7</sup>, Andreas Flemmer<sup>7</sup>, Hamid Hossain<sup>6,8</sup>, Xavier Pastor<sup>2</sup>, Holger Kirsten<sup>9</sup>, Peter Ahnert<sup>9</sup>, Juergen Behr<sup>10</sup>, Tushar J. Desai<sup>11</sup>, Benjamin Schubert<sup>2</sup>, Anne Hilgendorff<sup>1,12</sup>. Hyperoxia-induced cell cycle arrest drives long-term impairment of lung development and DNA repair in neonates.
3. Juan David Henao Sanchez<sup>3,14</sup>, *MustafaAbdo*<sup>1,2</sup>, MD, MSc, Benjamin Schubert<sup>3,14</sup>, PhD, Markus List<sup>4</sup>, PhD, Henrik Watz<sup>2,14</sup>, MD, Frauke Pedersen<sup>1,2,14</sup>, PhD, Alina Bauer<sup>3,15</sup>, MSc, Dominik Thiele<sup>5,14</sup>, MSc, Adam M. Chaker<sup>6,7</sup>, MD, Constanze A. Jakwerth<sup>7,15</sup>, PhD, Benjamin Waschki<sup>1,8,14</sup>, MD, Anne Kirsten<sup>2,14</sup>, MD, Markus Weckmann<sup>9,14</sup>, PhD, Oliver Fuchs<sup>9,10,14</sup>, MD, PhD, Gesine Hansen<sup>11,16</sup>, MD, Matthias V. Kopp<sup>9,14</sup>, MD, Erika v. Mutius<sup>12,13,15</sup>, MD, MSc, Inke R. König<sup>4,14</sup>, PhD, Klaus F. Rabe<sup>1,14</sup>, MD, PhD, Thomas Bahmer<sup>1,14</sup>, MD, Carsten B. Schmidt-Weber<sup>7,15</sup>, PhD, Ulrich M. Zissler<sup>7,15</sup>, PhD, and the ALLIANCE Study Group\*. Cytokines Derived from Nasal Epithelial Lining Fluid in Patients with Asthma.
4. Henao, J. D., Lauber, M., Azevedo, M., Grekova, A., Theis, F., List, M., ... & Schubert, B. (2023). Multi-omics regulatory network inference in the presence of missing data. *Briefings in Bioinformatics*, 24(5), bbad309.



## 4 Discussion

This is the discussion of the thesis.



# A Appendix

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.