

TECHNISCHE UNIVERSITÄT MÜNCHEN

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und
Umwelt

Statistical data integration in translational medicine

Linda Carina Krause

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende/-r:

Prof. Dr. Dietmar Zehn

Prüfende/-r der Dissertation:

1. Prof. Dr. Dr. Fabian Theis

2. Prof. Dr. Bernhard Küster

Die Dissertation wurde am 28.03.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 27.09.2019 angenommen.

Acknowledgment

I would like to thank Fabian J. Theis for being my doctoral supervisor and inviting me to come to Munich to work as a doctoral researcher in the rich scientific environment at his institute. I am deeply grateful to my supervisor Nikola S. Mueller for her continuous guidance, her steady support and scientific input. I thank my co-supervisor Stefanie Eyerich for discussing all my results in detail, being open to my ideas and introducing me to the exciting world of immunology.

Further, I would like to thank Prof. Dr. Bernhard Küster for being in my thesis committee and Prof. Dr. Dietmar Zehn for being the chair of this committee.

I would like to thank all former and current members of the computational cell maps group for great discussion rounds, interesting journal clubs and also for the fun we had together during our events and retreats. A special thanks to all the long-term great room mates I had over the years: beginning with Ivan Kondoversky, Julia Söllner and Norbert Krautenbacher, followed by Eudes Barbosa and Karolina Worf and finally Christoph Ogris joined us in room 022. I always enjoyed going to the office and an important part of me being so happy there, were all you! Thank you for tolerating my aromatic teas and urge for fresh air.

I would like to thank all former and current members of the Institute of Computational Biology for making the place a warm and inspiring place to work. Especially I would like to thank my daily “lunch round” girls Karolina Worf, Kiki Do, Elisa Benedetti, Alida Kindt, Anna Fiedler and Carolin Loos. A special thanks to the inofficial, best thesis writers’ club! I would also like to thank Bettina Knapp who helped me getting started with my PhD at the very beginning.

Further, I would like to thank all of my collaboration partners for gathering the data, coming up with exciting scientific questions and for a respectful and fruitful working environment. A special thanks to Stefanie Eyerich’s group at the ZAUM - Center of Allergy and Environment who welcomed me every week in their offices. Especially I would like to thank Anne Atenhan and Jenny Thomas for answering all my questions around T cells, biology, laboratory procedure, ... and for a great atmosphere in general. It was always a pleasure for me to come to ZAUM, even during the times when the seminar started at 7.30 am. I also would like to emphasize my gratitude towards my clinical collaborators in Munich at the Department of Dermatology and Allergy, Technical University of Munich.

Especially I would like to thank Kilian Eyerich, Natalie Garzorz-Stark, Felix Lauffer and Kerstin Pätzold for the great work we have accomplished together.

I would like to thank the three graduates schools (QBM - Quantitative Bioscience Munich, HELENA - Helmholtz Graduate School, TUM-GS the graduate school of TU Munich) I was part of for their financial support, guidance and help to find my way in the scientific world. I truly enjoyed the inspiring workshops and networking events. Especially I would like to thank those graduate schools for introducing me to some of my closest friends who helped me through the troubles of being a doctoral researcher: Simone Heber for always having a great coffee at the institute next door for me, Andrea Weinzierl and Andor Goetzendorff for our almost weekly dinners including discussions about science and more and all of my QBM Buddies for showing me what kind of science is also out there, for great dinners and fun retreats!

Further, I would like to thank all of my friends and (Beach-)Volleyball team mates for helping me keep my mind off work from time to time and for showing me that there is a life outside of the scientific world.

I also would like to acknowledge all those people and especially professors who were not part of my time as a doctoral student but who motivated me to pursue science. In particular, I would like to thank Bernhard Haubold and Angelika Börsch-Haubold for first introducing me to the scientific world and showing me how interesting life as a scientist can be.

I would like to thank Patrick Siecke for showing me what is important in life.

Finally, I would like to express my deepest gratitude to my parents Ursula Schwieder-Krause and Jörg Krause and my sister Laura Krause for their support, their belief in me throughout my life and always being there for me. I especially thank Tim Henschel for supporting me through the ups and downs of the last years and for always being by my side.

Abstract

Translational medicine describes the path from basic biomedical research to health improvement for the general public. The focus of this thesis is on bridging the gap between wet lab-focused views and clinical application across ten studies with three university hospitals by employing tailored computational tools and statistical methods. In particular, four topics in translational medicine research were thoroughly investigated.

First, two molecular classifiers as means towards unbiased diagnosis were developed. To differentiate two inflammatory skin diseases a novel classifier was validated on disease subtypes. To diagnose asthma a novel molecular classifier was established and validated by means of an independent patient cohort using penalized regression modeling.

Second, to standardize clinical characterization, serum proteins as easily accessible and minimally invasive markers were modeled jointly with clinical attributes. For disease monitoring and patient prognosis, optimized, regularized and consensus regression models were built.

Third, the application of linear mixed effects models to adjust for inter-individual variability in human gene expression data was established. This variability masks common, underlying disease characteristics in complex phenotypes.

Finally, the characterization of human T helper cell subsets was improved by identifying subset-specific molecular markers. Those markers might advance the insights into the function of T helper cell subsets within the human immune system.

In summary, the analyses improved diagnosis, monitoring and understanding of human diseases by means of statistical data integration. The work on the molecular disease classifier for inflammatory skin diseases is extended for use in local practices so might impact daily clinical diagnosis in dermatology. To identify possible, further benefits for the patients the remaining computational results still need to be validated in the labs and clinics.

Zusammenfassung

Translationale Medizin beschreibt den Pfad beginnend bei biomedizinischer Grundlagenforschung hin zur Verbesserung der allgemeinen Gesundheit der Bevölkerung. In insgesamt zehn Studien in Kollaboration mit drei Universitätskliniken lag der Fokus dieser Arbeit darauf, die Lücke zwischen experimentellen Ergebnissen und klinischer Anwendung zu schließen, indem maßgeschneiderte bioinformatische und statistische Methoden angewendet wurden. Insbesondere vier Themen der Forschung in der translationalen Medizin wurden untersucht.

Zunächst wurden zwei molekulare Klassifikatoren entwickelt, um Krankheiten objektiv zu diagnostizieren. Zur Unterscheidung zweier entzündlicher Hauterkrankungen wurde ein neuartigen Klassifikator mithilfe von Krankheitssubtypen validiert. Zur Diagnose von Asthma wurde ein neuer molekularer Klassifikator etabliert und mithilfe einer unabhängigen Patientenkohorte validiert.

Als Zweites wurden zur Standardisierung der klinischen Charakterisierung von Patienten Serumproteine als minimalinvasive Marker gemeinsam mit klinischen Attributen modelliert. Für die Überwachung von Krankheiten und der individuellen Prognose wurden optimierte, regularisierte und konsensusbasierte Regressionsmodelle entwickelt.

Als Drittes wurde die Anwendung linearer gemischter Modelle zur Korrektur interindividueller Variabilität in humanen Genexpressionsdaten etabliert. Diese Variabilität überlagert zugrunde liegende Krankheitsmerkmale in komplexen Phänotypen.

Abschließend wurde die Charakterisierung von Subpopulationen menschlicher T-Helferzellen verbessert, indem subpopulations-spezifische molekulare Marker identifiziert wurden. Diese Marker könnten die Erkenntnisse über die Funktion von T-Helferzell-Subpopulationen im menschlichen Immunsystem vertiefen.

Zusammenfassend verbesserten die auf statischer Datenintegration basierenden, analytischen Ergebnisse die Diagnose, Überwachung und das Verständnis menschlicher Erkrankungen. Der molekulare Klassifikator für entzündliche Hautkrankheiten wird momentan in niedergelassenen Arztpraxen getestet. Dies könnte die tägliche klinische Diagnose in der Dermatologie grundlegend verändern. Um weitere mögliche Vorteile für Patienten zu ermitteln, sollten auch die weiteren Forschungsergebnisse experimentell und klinisch validiert werden.

List of contributed articles

The thesis is mainly based on the results published in the following peer-reviewed publications or on currently prepared manuscripts. The publications are listed according to the chapters in which their content is discussed. The specific contributions of the thesis author in each study is described in detail in section 1.6. (* = equal contribution)

Chapter 3

- N. Garzorz-Stark*, **L. Krause***, F. Lauffer, A. Atenhan, J. Thomas, S.P. Stark, R. Franz, S. Weidinger, A. Balato, N.S. Mueller, F.J. Theis, J. Ring, C.B. Schmidt-Weber, T. Biedermann, S. Eyerich and K. Eyerich: *A novel molecular disease classifier for psoriasis and eczema*. Experimental Dermatology (2016).
- K. Milger, J. Götschke, **L. Krause**, P. Nathan, F. Alessandrini, A. Tufman, R. Fischer, S. Bartel, F.J. Theis, J. Behr, S. Dehmel, N.S. Mueller, N. Kneidinger and S. Krauss-Etschmann: *Identification of a plasma miRNA biomarker-signature for allergic asthma: a translational approach*. Allergy (2017).

Chapter 4

- **L. Krause**, V. Mourantchian, K. Brockow, F.J. Theis, C.B. Schmidt-Weber, B. Knapp, N.S. Mueller and S. Eyerich: *A computational model to predict severity of atopic eczema from 30 serum proteins*. Journal of Allergy and Clinical Immunology (2016).
- V. Baghin*, **L. Krause***, S. Eyerich, K. Eyerich, F.J. Theis, N.S. Mueller, F. Lauffer and N. Garzorz-Stark: *Predicting persistence of atopic eczema in children using serum proteins and clinical data*. - in preparation -
- K. Dehlke*, **L. Krause***, F.J. Theis, U. Klingmüller, N.S. Mueller and K. Hoffmann: *Prediction of individualized liver regeneration capacity after liver resection based on cytokine and growth factor profiling*. - in preparation -

Chapter 5

- F. Lauffer, M. Jargosch, **L. Krause**, N. Garzorz-Stark, R. Franz, S. Roenneberg, A. Böhner, N.S. Mueller, F.J. Theis, C.B. Schmidt-Weber, T. Biedermann, S. Eyerich and K. Eyerich: *Type I immune response induces keratinocyte necroptosis and is associated with interface dermatitis*. The Journal of Investigative Dermatology (2018).
- N. Garzorz-Stark*, F. Lauffer*, **L. Krause**, J. Thomas, A. Atenhan, R. Franz, S. Roenneberg, A. Böhner, M. Jargosch, R. Batra, N.S. Mueller, S. Haak, C. Groß, O. Groß, C. Traidl-Hoffmann, F.J. Theis, C.B. Schmidt-Weber, T. Biedermann, S. Eyerich and K. Eyerich: *Toll-like receptor 7/8 agonists stimulate plasmacytoid dendritic cells to initiate TH17- deviated acute contact dermatitis in human subjects*. Journal of Allergy and Clinical Immunology (2018).

Chapter 6

- **L. Krause**, J. Thomas, A. Atenhan, F.J. Theis, S. Eyerich and N.S. Mueller: *Understanding human T helper cell subsets using multiple omics levels*.
- in preparation -

Further contributed articles

Besides the above mentioned publications, the thesis author was also involved in other studies during her PhD time. Two studies were already published, one publication is under review and three manuscripts are currently prepared. Author lists and publication titles might be subject to changes for currently prepared manuscripts.

- J. Thomas*, A. Atenhan*, M. Quaranta, **L. Krause**, J. Buters, C. Ohnmacht, R. de Jong, C.B. Schmidt-Weber and S. Eyerich: *FOXO4 and AHR control a stress-induced immune-mediated host protection via secretion of IL-22*. - in preparation -
- **L. Krause***, A. Pitea*, C. Ogris, G. Eraslan, S. Sass, J. Arloth, M. Preusse and N.S. Mueller: *Holistic multi-level analysis of long non-coding RNAs using flexible graph database*. - in preparation -
- D.C. Dittlein*, C. J. Groß*, J. Smollich, D. Stößel, R. Mishra, **L. Krause**, H. Shi, K. Eyerich, B. Beutler, C. Traidl-Hoffmann and O. Groß: *Imiquimod inhibits mitochondrial complex I for K^+ efflux-independent NLRP3 inflammasome activation and cancer cell growth arrest*. - in preparation -
- F. Lauffer*, M. Jargosch*, V. Baghin, **L. Krause**, W. Kempf, M. Absmaier, M. Morelli, S. Madonna, C. Albanesi, N.S. Mueller, F.J. Theis, C. Schmidt-Weber, S. Eyerich, T. Biedermann, N. Vandeghinste, S. Steidl and K. Eyerich: *IL-17C amplifies epithelial inflammation in psoriasis and atopic eczema*. - submitted to Journal of Allergy and Clinical Immunology in January 2019 -
- J. Thomas*, M. Küpper*, R. Batra, M. Jargosch, A. Atenhan, V. Baghin, **L. Krause**, F. Lauffer, T. Biedermann, F.J. Theis, K. Eyerich, S. Eyerich and N. Garzorz-Stark: *Is the humoral immunity dispensable for the pathogenesis of psoriasis?* Journal of the European Academy of Dermatology and Venereology (2018).
- A. Graessel, S.M. Hauck, C. von Toerne, E. Kloppmann, T. Goldberg, H. Koppensteiner, M. Schindler, B. Knapp, **L. Krause**, K. Dietz, C.B. Schmidt-Weber and K. Suttner: *A Combined Omics Approach to Generate the Surface Atlas of Human Naive CD4+ T Cells during Early T-Cell Receptor Activation*. Molecular and Cellular Proteomics (2015).

Contents

1	Introduction	1
1.1	Translational medicine and its data	2
1.2	The role of the immune system in complex phenotypes	5
1.3	Challenges in translational medicine	6
1.4	Research questions	13
1.5	Overview of this thesis	14
1.6	Specific contributions of the thesis author in each study	16
2	Methods	19
2.1	Statistical analysis	19
2.2	Gene expression measurements and processing	39
2.3	Measurement and processing of secretome expression	48
2.4	Overview of studied diseases	53
2.5	Clinical data and its processing	55
2.6	Overview of studies, cohorts and data sets	57
3	Building and characterizing disease classifiers in inflammatory skin diseases and allergic asthma	61
3.1	Molecular classifier for psoriasis and eczema	63
3.2	Plasma miRNA classifier for allergic asthma	70
4	Serum proteins as easily accessible surrogates for monitoring human diseases	79
4.1	Predicting severity of atopic eczema	81
4.2	Predicting persistence of atopic eczema	87
4.3	Prediction of individualized liver regeneration capacity	95
5	Adjusting patient-bias in microarray data using linear mixed effects models	103
5.1	Application in interface dermatitis	105
5.2	Application in imiquimod-induced skin reactions	113
6	Molecular characterization of human T helper cell subsets using multiple omics levels	123
6.1	Motivation	125
6.2	Methods	128
6.3	Results	135

Contents

6.4 Discussion	170
7 Discussion and outlook	175
7.1 Discussion	175
7.2 Outlook	179
Bibliography	183

Chapter 1

Introduction

Inferring knowledge from data is the general aim in computational biology and for many scientific endeavors. Data is often plentiful available but data alone does not help to answer the urging questions around human health and disease currently posed in the field of translational medicine. To transform data to knowledge, computational tools, statistical methods and custom-tailored data integration concepts are needed. By analyzing data jointly, in an integrative manner, and not consecutively or in parallel, the gain in knowledge is increased since a more complete picture is investigated. In this thesis, available computational tools and statistical methods are tailored to the clinical or biological question at hand to enable statistical integration of different biomedical data types, from molecular measurements to clinical attributes. The aim of statistical data integration in the thesis is to improve diagnosis and characterization of human diseases by inferring robust and interpretable computational results.

In this thesis, data integration is defined in terms of statistical analysis of paired biological and clinical data and the subsequent interpretation. If large amounts of heterogeneous data are collected for the same individual, it was analyzed jointly using statistics to infer interplay between at least two data types. Data types are clinical attributes or molecular markers and measurements where understanding their relationships is crucial for translational medicine research. To perform statistical data integration, first data is cleaned up and transformed to a usable format for analysis but this is not the main focus like in data engineering. In contrast to mechanistic models which try to infer how processes are related and draw causal relationships, the focus of this thesis is in a statistical description of the interplay of measurements and data types.

In translational medicine the knowledge which should be inferred is usually centered around a specific clinical or biological question. The questions originate from physicians and biologists. So, during the whole scientific process, computational biologists, work hand in

hand with collaboration partners to integrate the available data and interpret the results in light of the originally asked question. Finally, a way of transporting the insights to physicians through visualizations and discussions is required so that integrated results can be interpreted correctly for the benefit of the patients. Pursuing to increase the benefit for the individual patient and the general public is at the core of translational medicine research.

1.1 Translational medicine and its data

Translational medicine is a broadly used term in the current literature describing the path from basic biomedical research to health improvement for the general public. In the year 2000 the term was first used in a publication and steadily increased since then up to over 4,600 papers dealing with this topic in 2013 (Cohrs et al., 2015). What exactly is understood by *translational medicine* or *translational research* is actively discussed. A recent review summarized **33 different definitions** of translational research and the author clustered them into three categories each having one originating paper which presented the definition idea first (Fort et al., 2017). The three categories include the “gap” model initiated by Sung et al. (2003), the “continuum” model from Khoury et al. (2007) and a “mixed” model by Woolf (2008).

Computational biologists focus on **bridging the gap** between wet lab-focused views and clinical application using analyses to improve understanding, diagnosing and monitoring of human diseases. This is described as the first gap or obstacle of translational medicine by Sung et al. (2003) who generalized the difficulty to the transfer of new insights gained in the laboratory to first applications in humans. However, that is not the entire picture of translational medicine. The second gap manifests in using results of clinical studies in the daily clinical practice (Sung et al., 2003). To fill the second gap different methods and approaches are needed. The aim of translational research is to bridge both gaps, also according to ideas discussed at a meeting at the Institute of Medicine, which is now called the National Academy of Medicine (The National Academies of Science, Engineering and Medicine, 2015). These two gaps also describe the different views between basic scientists and public health agencies on translational research. Basic scientists often focus on the first gap, from bench to first testing in humans (Hörig et al., 2005). Public health agencies, on the other hand, believe that translation also includes the next steps, from clinical testing to health benefit for the whole population (Centers for Disease Control and Prevention, 2007).

1.1 TRANSLATIONAL MEDICINE AND ITS DATA

The **continuum model** was developed later and describes a path through the translational continuum rather than gaps and extends the previous idea. The second gap in the model by Sung et al. (2003) is split into one part focusing on the development of evidence-based guidelines and another which describes the step from these guidelines to clinical practice. Khoury et al. (2007) extended the model by a fourth part dealing with the impact of changing clinical practice in the real world. The translational medicine model described by Woolf (2008) is a mix of both definitions. All phases of translational medicine are important. However, this thesis only shows work on the first step in translational research: data analysis and statistical modeling in the intersection between laboratory work and investigations in the clinics.

Translational medicine is a **data-rich field** (Hey et al., 2009). Data used differs from gene expression, genomic variants to electronic health records. This thesis focuses on biomedical data measured in *in vitro* cell lines or human subjects acquired by the collaborators (Fig. 1.1). In this thesis, access to two broad categories of data was provided, both important for translational medicine: molecular data and clinical attributes.

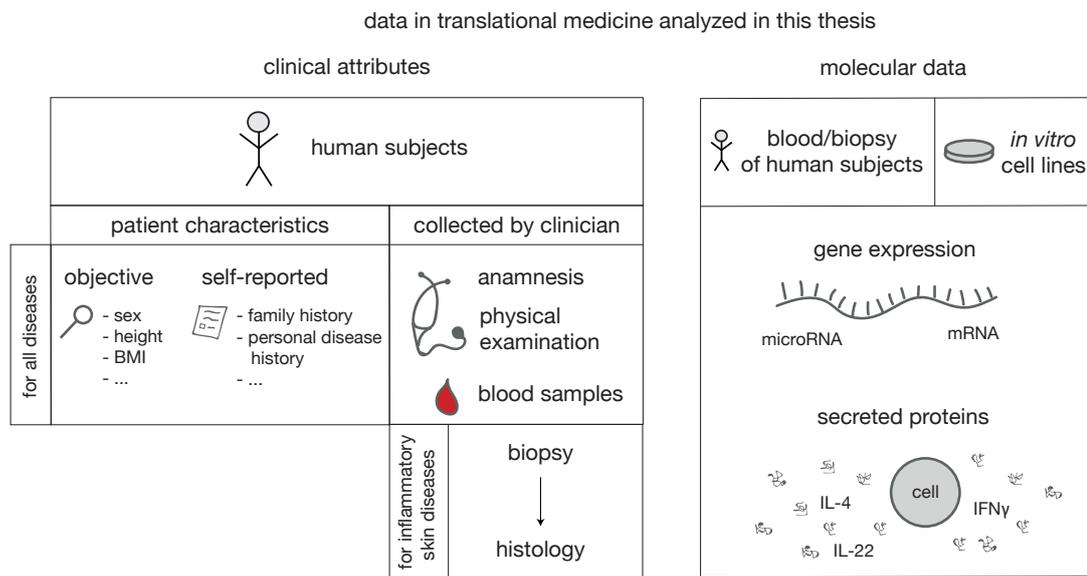


Fig. 1.1 – Biomedical data analyzed in this thesis. Data in translational medicine studies investigated in this thesis consist of clinical attributes and molecular data. For human subjects, clinical attributes are both patient characteristics and attributes collected by the clinician. Blood is routinely drawn for all investigated diseases and for inflammatory skin diseases biopsies are taken. Molecular data consist of measurements of gene expression and secreted proteins performed in blood or biopsies of patients or in *in vitro* cell lines.

CHAPTER 1 INTRODUCTION

In routine medical practice, **physicians collect a lot of information**. This information consists of medical history and clinical attributes which are noted during anamnesis and physical examination (Fig. 1.1). The available information also often contains blood samples which are collected during physical examinations. In the field of dermatology taking biopsies of the involved skin for histological analysis is also part of the standard procedure. All of these data types are routinely stored in hospitals. Due to data security reasons and disjoint systems, the data stored in hospitals is yet not routinely available for computational analysis but only through close collaboration with clinical partners. With the help of collaboration partners, the work presented in this thesis always started working from these routinely stored data but integrated further information gained through molecular and clinical phenotyping.

Molecular data is either measured in the tissue of interest, e.g. in biopsies of lesional skin of inflammatory skin diseases or respective tissue from surgical procedures, in blood or in *in vitro* cell lines (Fig. 1.1). Even though the involved tissue might bear more information, the blood is readily accessible, routinely collected and serves as a momentary state of the patient's condition. Molecular data is an umbrella term; the studies in this thesis only contain data on mRNA and micro RNA levels, secreted proteins and blood test results. Molecular data is assumed to be objective quantifications, with machine limitations, of true values.

Besides molecular data, **clinical attributes** fall into the second category of data analyzed in this thesis. There are at least two kinds of clinical attributes: attributes collected by clinicians who examine a patient and self-reported patient characteristics often obtained by questionnaires (Fig. 1.1). Examination results of doctors can depend on their experience and how they evaluate subjective criteria. However, critical scores like disease severity scores are assessed following a detailed procedure minimizing inter-doctor variability. For example, the SCORAD (SCORing Atopic Dermatitis) is a method to determine the severity of atopic eczema (Schallreuter et al., 1993). After assessing the variability in scoring results the scheme was further optimized to increase consistency (Kunz et al., 1997). The other part of clinical attributes, namely patient characteristics, include objective attributes like height, BMI and sex. Also personal disease and family history are important factors. The latter ones rely on what the patient reports and remembers correctly.

The depth and completeness of clinical and molecular data depends on the study set up, thus the **kind of cohort**. There are different resources for clinical and molecular information: longitudinal and national cohorts, local biobanks and investigator driven case-control studies. In longitudinal, population-based studies participants are recruited during a

1.2 THE ROLE OF THE IMMUNE SYSTEM IN COMPLEX PHENOTYPES

specific time frame and monitored over years. Local biobanks follow a different recruitment strategy. They are typically associated to a local hospital and recruit patients as they come to the clinics. As a consequence they grow permanently. Investigator-driven case-control studies follow several approaches to find suitable patients. Investigators can check whether local biobanks contain patients who fulfill the inclusion criteria. Another option is to actively enroll patients in the study solely based on its specific criteria. Biobank Biederstein is a local, Munich example situated at the Clinic of Dermatology and Allergology which stores biological samples left over after routine investigations (Quaranta et al., 2014a).

Regardless of the specific cohort used, both molecular and clinical data is analyzed to find markers that help diagnose and monitor patients. Choosing the right computational method to detect important features among all measured ones is a challenge. Moreover, careful data integration using the right statistical tools can create a deeper comprehension of the underlying pathogenesis.

1.2 The role of the immune system in complex phenotypes

Why do some people get sick and others not? The **immune system** plays a crucial role in answering this question. The immune system helps every day to fight against pathogens like viruses, bacteria or fungi. One part of the immune system is the innate immune system, which is already fully established when humans start their lives as newborns and is stable throughout their lives. This stability is used by pathogens to find ways to overcome it. Therefore a second line of defense against pathogens has evolved, the adaptive immune system. In contrast to the innate immune system, the adaptive immune system is trained throughout life. It is responsible for keeping information of pathogens which were already encountered in order to react faster when it meets them again.

An important part of the adaptive immune system are **T helper cells (Th cells)**. By secretion of small proteins which are called cytokines, they organize adaptive immune responses. The body is not only attacked by one kind of pathogen but by very diverse pathogens. This is why Th cells have specialized to accomplish various tasks which leads to efficient defense against those damaging invaders. It was shown that there are at least seven Th cell subsets in humans with specific functions in the human body: Th1, Th2, Th9, Th17, Th22, T follicular helper cells (Tfh), regulatory T cells (Tregs) (Eyerich and Zielinski, 2014).

CHAPTER 1 INTRODUCTION

Recently, single cell biology has increasing impact in improving the understanding of the immune system. Wong et al. (2016) analyzed T cell trafficking and important functional markers in humans. Through mass cytometry analysis they showed that T cell subsets defined in mouse models cannot directly be associated to human subsets (Wong et al., 2016). Zielinski et al. (2018) used this exciting data set (Wong et al., 2016) to discuss new analytical approaches for high dimensional data in single cell immunology. Hoppe et al. (2016) applied live imaging of single cells to understand haematopoiesis and challenged previous models of early myeloid lineage choice. Angelidis et al. (2019) investigated transcriptional and cell type composition changes in lung tissue of young and old mice. Lung tissue includes many immune cells and they showed that the relative amount of immune cells was higher in older mice.

Before the immune system has to take action, pathogens have to break through the first barrier of defense, epithelial surfaces such as skin, lung and gut. The **skin**, the biggest organ, is also home to Th cells, hosting almost twice as many Th cells compared to blood (Clark et al., 2006). These Th cells belong to different subsets. If the balance of Th cells subsets is distorted, **inflammatory skin diseases** can arise. In patients suffering from psoriasis an imbalance towards more Th1, Th17 and Th22 cells compared to Th2 cells was observed in the lesional skin (Nestle et al., 2009). In contrast, in patients with atopic eczema more Th2 cells were detected in affected skin in relation to Th1 cells (Bieber, 2008). Often percentages of specific subsets are compared to gain insight into disease pathogenesis (Zhan et al., 2018). However, psoriasis and atopic eczema are not only defined by the Th cell imbalance, but by several more factors, which makes them multifactorial, complex diseases.

Most human diseases do not have a single cause, but have a **complex phenotype** and pathogenesis which is entangled with the immune system's activity. A better understanding of underlying molecular mechanisms and immune cell regulations in complex diseases could lead to development of better treatment or diagnostic strategies. Translational medicine research helps providing these insights, also by performing biomedical data integration across all levels of information.

1.3 Challenges in translational medicine

Translational medicine is a promising field which can lead to improvement of diagnosis and treatment of individual patients. Some challenges of the field were identified. Here, each challenge is described in detail, solutions introduced by others are presented and

1.3 CHALLENGES IN TRANSLATIONAL MEDICINE

problems of these solutions, if they exist, are mentioned. The following challenges are described: challenges in compositions of clinical case-control cohorts, challenges in clinical characterization and monitoring of patients, challenges when investigating complex phenotypes and the challenge of finding laboratory equivalents to test hypothesis in translational medicine research.

1.3.1 Composition of clinical case-control cohorts

One challenge is the composition of clinical case-control cohorts in which translational medicine discovery studies are performed. Most volunteers taking part in clinical research studies are recruited for case cohorts in local hospitals or practices which leads to a regional, environmental bias (Andersen, 1995; Rattay et al., 2013). Especially, recruitment in hospitals can confound research results since patients are more severely affected or need special treatment if they are referred to by their local practitioner (Ommen et al., 2007). Selection of control cohorts should follow several principles according to Wacholder et al. (1992). One is the “study base principle”. It describes that cases and controls should have the same context, exposure and risks. Another principle is the “comparable accuracy principle” which implies that the degree of accuracy of cases and controls should be the same in measuring the exposure of interest. Recall or information bias distorts the equity in accuracy between cases and controls. Recall bias might especially be present if the study assesses the exposures retrospectively due to the possibly different perception of the importance of exposures by healthy and diseased individuals (Mann, 2003; Song and Chung, 2010). Instead of following the principles by Wacholder et al. (1992), control cohorts are often biased. They usually include volunteers available for recruitment and are generally not sampled following the same procedure as the cases were (Mann, 2003). In summary, small clinical case-control cohorts, are not representative of the population but rather a sub-selection biased on availability and consent (Mann, 2003). Moreover, researchers may not have access to an independent validation cohort, so they evaluate trained models again on the same data set which might lead to non-representative findings (for example Thijs et al. (2015b)). The number of studied individuals ranges from around ten to hundreds to almost 500,000 in latest genome wide association studies.

The ideal solution to overcome the problem of biased cohorts are prospective longitudinal studies. The key idea is to establish an unbiased cohort by including a large study population in a specific region or multiple regions during a recruitment period and follow the participants over time. Regular medical examinations and questionnaires are part of most prospective longitudinal cohorts. German examples are the GINI (von Berg

CHAPTER 1 INTRODUCTION

et al., 2003) and LISA (Heinrich et al., 2002) cohorts, two prospective, longitudinal birth cohorts assessing the influence of environmental factors on the immune system and allergies. GINI included 2,000 infants and for LISA 3,000 newborns were recruited. Prospective, longitudinal studies give unique possibilities to understand disease environment interactions and are good for interpretation. Yet, they are time consuming and expensive due to the need of following the participants over a long period of time (Cotter et al., 2005). There are also population-based cohorts like KORA (“Kooperative Gesundheitsforschung in der Region Augsburg”), the UK Biobank or the German national cohort (NAKO) which provide a large unbiased study population. KORA recruited since 1984 around 18,000 participant living in the Augsburg region (Holle et al., 2005). Since 2004, also biological samples have been gathered and stored to perform genetic-epidemiological research. UK Biobank started in 2006 and included 500,000 individuals aged between 40 and 69 and follows them over 30 years (Sudlow et al., 2015; Bycroft et al., 2018). Biological samples are linked to electronic health records to investigate the genetic and life-style risks for human diseases. NAKO started in 2015 and aims to recruit 200,000 participants aged between 20 and 69 years (NAKO e. V., 2019). In January 2019, 194,426 participants have already registered (NAKO e. V., 2019). The subjects undergo medical check-ups, fill questionnaires and donate biological samples upon recruitment and follow-ups are planned. The aim of NAKO is to understand the development of chronic diseases, determine risk factors and define prevention strategies (Wichmann et al., 2012).

In population-based cohorts disease tissue samples are non-existent and they only work for common diseases. To improve statistical power in case-control studies matched study designs are a possibility. In matched study designs some variables which are decided on in the beginning of the study, e.g. sex and age, are matched between cases and controls (Mann, 2003). Methodologically, this implies the usage of paired analysis methods between matched individuals. Problematic in matching is that unknown influences cannot be controlled for and matching does not entail that exposure was the same between both groups. Another problem is over-matching when matched variables are associated to the outcome of interest (Song and Chung, 2010).

1.3.2 Clinical characterization and monitoring of patients

Another key element is the clinical characterization and monitoring of patients and their progression. Medical doctors perform the necessary examinations and evaluate diagnoses and treatment strategies. The results are summarized in clinical attributes, which are important for defining diseases, interpretation of statistical modeling results and association

1.3 CHALLENGES IN TRANSLATIONAL MEDICINE

analyses. Variability in patients' characterization can arise due to individual characteristics of the doctor, like the experience level, and the patient but also subjective assessments.

Bertakis (2009) showed that gender of both patients and doctors influences their interactions during medical visits and their overall relationship. The doctor-patient relationship has an impact on their communication and can even change long-term treatment outcome (Ong et al., 1995). McKinlay et al. (2006) compared patient management of doctors from the US and UK in a video-based study (32 patient videos viewed and analyzed by 256 doctors in the US and UK). Doctors' gender and experience, the national health care system but also age, gender and overall appearance of the patient influenced the diagnosis (McKinlay et al., 2006). That experience impacts doctors' decisions was also shown in further studies (Ghassemi et al., 2018; Vogel, 2018). Especially, subjective assessments can vary widely between clinicians (study with 104 clinicians about four clinical estimates had mean standard deviation of 19.5, Dolan et al. (1986)).

Several clinical scores exist to make clinical characterization of patients more objective. For example, the "psoriasis area and severity index", short PASI, is a severeness score for psoriasis (Fredriksson and Pettersson, 1978) ranging from zero to 72 points. According to guidelines, a threshold for initiating biological treatment in patients is a PASI of ten (Smith et al., 2009). It is used as primary endpoint in clinical trials (Langley et al., 2015). Even though, clinical scores were introduced to improve objectivity in clinical characterization, variability is still present. Langley and Ellis (2004) showed that intra-personal variance in PASI score is between 1.2 and 3.2 for nine experienced and eight inexperienced raters, respectively, when evaluating 35 patients (PASI range 1 - 51, mean not given) twice. Inter-personal variance is higher between 8.1 and 9.6 PASI points, again separately calculated for experienced and inexperienced raters. A more recent study by Fink et al. (2018) where mean PASI was 8.8 (range 0.7 - 34.8) and three physicians examined 120 patients twice found that the between-rater difference was 3.3 PASI points and the difference within raters was 2.2 points (always assessed as mean absolute difference). The difference between intra-rater and inter-rater variability is even more pronounced when pathologists rate cancer volumes (standard deviation of 10 to 18% between pathologists and only 5% within one pathologist, Fiorino et al. (1998)). The SCORAD (SCORing Atopic Dermatitis) measures the severity of atopic eczema based on several criteria and defines patients' treatment (Schallreuter et al., 1993). Kunz et al. (1997) showed that objective SCORAD criteria, for example the extend of the lesion, can still range from 20 to 80 % (mean 40.8 ± 19 %) for one individual examined by 23 physicians. In summary, physicians are trained during their education to be objective but are still human actors

CHAPTER 1 INTRODUCTION

who can be influenced by non-medical factors like patient characteristics and their own nature (McKinlay et al., 1996; Eisenberg, 1979).

Clinical research studies try to tackle this problem by two different approaches. Either only one medical doctor examines all included patients to avoid inter-personal variability. Examples are present in a variety of research fields, from studies on atrial fibrillation (572 patients, Gladstone et al. (2014)), cancer (834 patients, Aghilinejad et al. (2017)), to nonalcoholic fatty liver disease (66 participants, Yoneda et al. (2010)). Or all examinations are performed by at least two clinicians to decrease variance and increase consistency. For this approach examples, where all clinical evaluation was performed by two physicians, can be found in a wide range of disciplines, from studies in adverse drug events (Morimoto et al., 2011), treatment of scars (Choi et al., 2014), medication use in elderly people (Sakuma et al., 2011) to pathology (Alessi et al., 2013). Sometimes authors even stressed that physicians acted independently and were blinded to the treatment (Le Duff et al., 2010; Fadel and Tawfik, 2015).

1.3.3 Complex phenotypes

Apart from the fact that available data sets might not be flawless due to biased cohort design and variability in clinical characterizations, another challenge in translational medicine are complex phenotypes like inflammatory skin diseases. Inflammatory skin diseases are not amongst the “top 10 global causes of death in 2016” (World Health Organisation, 2018), nonetheless, they put a huge burden on the patients. When calculating years people lose due to disability, skin diseases rank forth worldwide after low back pain, major depressive disorder and iron-deficiency anemia (Hay et al., 2014). Due to their complex nature, studying inflammatory skin diseases bears difficulties. One particular challenge are the inter-individual differences in patients. Every person’s uniqueness with regard to his or her gene expression may mask the underlying, common disease specific signatures and leads to inter-individual variable gene expression patterns (Yang et al., 2016; Ho et al., 2008; Whitney et al., 2003). This is true for complex diseases and particularly for skin diseases (Cole et al., 2001).

Complex diseases are often based on complex genetic backgrounds and genetic-environment interactions (Hunter, 2005). Genetic effects can be studied through their molecular effects e.g. on gene expression. Environment effects are more difficult to study. Especially the interactions make diagnosis and study of complex diseases challenging (Manolio et al., 2009).

1.3 CHALLENGES IN TRANSLATIONAL MEDICINE

An increase in sample size can help to overcome the challenge complex phenotypes pose. Nishino et al. (2018) estimated that a sample size of 50,000 or more cases and the same number of controls is required for a successful genome-wide association study investigating genetic effects of major depressive disorder. There are no sample size estimations for gene expression analysis in inflammatory skin diseases. Nevertheless, meta-analyses of gene expression data in psoriasis were performed and showed small overlap in results of different studies. The meta-analysis of Tian et al. (2012) included five studies on differentially expressed genes in lesional compared to noninvolved skin in psoriasis patients. Each of the five studies included between 13 and 81 patients. The intersection of gene lists contained only 100 genes. The meta-analysis of all microarrays combined, 386 paired samples of 193 patients, revealed 1,120 differentially regulated genes. This shows that an increase in sample size, here achieved via a meta-analysis, is one way of increasing detection power in complex phenotypes. The problem with meta-analysis is the harmonization of different techniques regarding how the gene expression was measured and how the prognosis or diagnosis of patients was evaluated (Ramasamy et al., 2008).

1.3.4 Laboratory equivalents for hypothesis testing

Computational analyses of translational research studies identify new hypotheses around human health and disease which need experimental testing. The results of the computational analyses are hypotheses which describe associations between genes or other molecules and clinical outcomes (Quaranta et al., 2014b; Garzorz-Stark et al., 2018; Lauffer et al., 2018). Testing the association is simplified if, for example, an approved drug exists which targets the gene or molecule of interest. In one study described in this thesis, ustekinumab was used which is an antibody that is capable of neutralizing the cytokine IL-23, to successfully test the hypothesis that IL-23 is involved in the occurrence of induced contact dermatitis upon application of imiquimod (Garzorz-Stark et al., 2018). For other hypotheses the testing has to be performed in laboratory equivalents of human diseases. A complete human model system is missing and better *ex vivo* models or more detailed *in vitro* models are needed for improved hypothesis testing.

For several diseases, established mouse models are available and used for hypothesis testing. Especially, inflammatory skin diseases are tough to study in mouse model systems due to the differences in skin architecture (Khavari, 2006) and the immune system (Mestas and Hughes, 2004) which is crucially involved in the pathogenesis. Three dimensional skin equivalents are getting popular since they were introduced by Poumay et al. (2004) in studying inflammatory skin diseases. For example, Van Den Bogaard et al. (2014)

CHAPTER 1 INTRODUCTION

established a three dimensional model including immune and skin cells and proposed to use it for preclinical drug screening. However, the combination of different cell types is still challenging in *in vitro* models due to varying culture conditions of different cell types (Duque-Fernandez et al., 2016).

Single cell studies are increasingly used to detect disease biomarkers (Zhu et al., 2014) and to validate them (Niu et al., 2016). Tirosh et al. (2016) used single cell sequencing of skin cells to understand the complex interplay between cells in melanoma tumors. Der et al. (2017) determined a biomarker for Lupus nephritis, an autoimmune disease, using single cell RNA sequencing of renal and skin biopsies. A lot of human single cell studies are performed in cancer patients (hepatocellular carcinoma, Zheng et al. (2017); ovarian cancer, Winterhoff et al. (2017); breast cancer, Demeulemeester et al. (2016); colorectal cancer, Leung et al. (2017)) or stem cells (cardiac stem cells, Liu et al. (2017); chronic myeloid leukemia stem cells, Giustacchini et al. (2017); bladder cancer stem cell, Yang et al. (2017); neural stem cells, Dulken et al. (2017)). These studies currently aim at understanding intratumoral heterogeneity and metastasis or describing cell differentiation and subpopulations. The Human Cell Atlas will provide a large resource of single cell data sets which will help to improve understanding of all cells within the human body (Regev et al., 2017). The focus of the Human Cell Atlas is samples from healthy individuals and only small cohorts of patients with relevant diseases are included (Regev et al., 2017). Potentially, single cell studies are useful for testing translational medicine hypotheses.

Skin organoids grown from induced pluripotent stem cells (kidney organoids, Forbes et al. (2018); neural organoids, Hartley and Brennand (2017)) or biopsies (gut organoids, Sato et al. (2011); colon cancer organoids, Boehnke et al. (2016)) pose another way to test drugs and perform validation experiments *ex vivo*. In summary, a lot of translational medicine hypotheses including diseases pathogenesis and treatment options can be tested in cell culture experiments. Fully understanding them is crucial for drawing the correct conclusions.

1.4 Research questions

The overall aim of this thesis is to improve diagnosis and characterization of human diseases by jointly analyzing clinical and molecular data using statistical data integration methods. The focus is mainly on the fields of immunology and dermatology. During the statistical analysis, an emphasis was put on robustness and interpretability of computational results to increase likelihood of translating wet lab based results into clinical practice.

The aim was approached from different angles depending on available resources and specific clinical interest of the overall 14 different studies. Of these, eight are already published papers and for the six remaining studies manuscripts are currently prepared or submitted (see section about scientific publication on pages (ix-xi) for more details). This thesis describes a subset of eight studies of which five are among the published ones. The entire research of these eight studies in this thesis answers four main questions:

- Can robust and interpretable molecular disease classifiers for unbiased patient diagnosis in inflammatory skin diseases and allergic asthma be found by means of disease subtypes and independent patient cohorts?
- Can serum proteins as easily accessible and minimally invasive markers for disease monitoring and prognosis be used to standardize clinical characterization in atopic eczema and liver resection surgery?
- Can inter-individual variability in patients which masks common, underlying disease characteristics in complex phenotypes be adjusted using the example of inflammatory skin diseases?
- Can new marker genes describing T helper cell subsets be obtained to better characterize their phenotypes and their role in the immune system?

1.5 Overview of this thesis

The chapters of this thesis are organized according to research questions, each giving one or more examples of how the challenges were met and overcome. Each subchapter explaining one example is structured in the same way. The study is first motivated, second methods specific for this study are explained which also include a description of the patient cohort, if applicable, third results are presented and finally the results are discussed. **Chapter 1** introduces the topics and main concepts used in the thesis. **Chapter 2** provides background information on data types, statistical methods shared among several studies and analyzed cohorts.

The following chapters are centered around human cells which were assessed for inside-cell gene expression and secreted proteins. This data is measured in *in vivo* and *in vitro* studies investigating human diseases and the immune system. Clinical features are important counterparts used for defining diseases, interpreting disease classifiers and calculating associations to *in vivo* markers (Fig. 1.2).

Chapter 3 describes an approach on defining and characterizing suitable molecular disease markers to fulfill the need for unbiased patient diagnosis. Two published disease classifiers for inflammatory skin diseases and allergic asthma are discussed in detail (Garzorz-Stark et al., 2016; Milger et al., 2017). After determining disease-specific molecular markers, they were associated to all available personal, clinical and laboratory attributes for better interpretability.

Patient serum is easily accessible and only requires a minimally invasive routine intervention performed by medical doctors. The focus was on finding markers in serum as surrogates for disease monitoring and prognosis, alongside of three applications explained in **chapter 4**. A combination of serum proteins which best predict the severity of atopic eczema in adult patients was determined (Krause et al., 2016). A common set of variables was defined which predict whether children loose atopic eczema when they grow up or have a persistent disease course (publication in preparation). Finally, time series data from patients who underwent liver resection surgery was analyzed (publication in preparation).

Every patient is unique and so is his or her gene expression profile. Particularly, when analyzing patients who share the same disease background clinically, inter-individual variability may mask the underlying, common disease characteristics. In **chapter 5**, the application of linear mixed effects models to adjust for patient-bias in microarray gene expression data was proposed. The approach was applied to interface dermatitis diseases

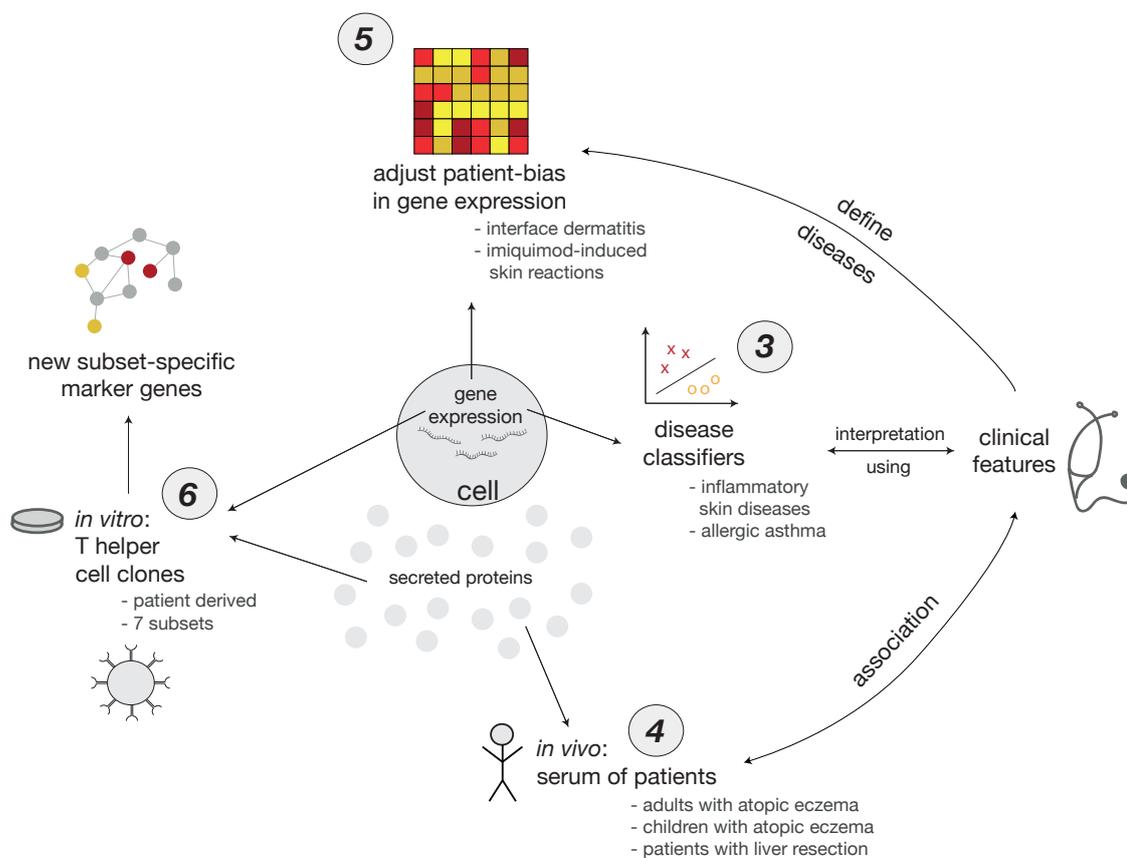


Fig. 1.2 – Overview of this thesis. The cell, its gene expression and secreted proteins are in the center and used in different *in vivo* and *in vitro* studies. Clinical features are used to define diseases, understand molecular classifiers and be associated to serum measurements. Numbers refer to chapters.

(Lauffer et al., 2018) and imiquimod-induced skin reactions in humans (Garzorz-Stark et al., 2018).

The last three chapters provide deeper molecular understanding of human diseases which is necessary and important. Translational medicine research generates hypotheses about how diseases develop and why people get sick. Testing these hypotheses is a crucial part often done in model systems like *in vitro* cell line experiments. **Chapter 6** summarizes work on characterizing the phenotype of T helper cell subsets using two omics levels measured in *in vitro* experiments (publication in preparation).

1.6 Specific contributions of the thesis author in each study

This section gives an overview of the specific contributions and responsibilities of the thesis author in the studies which are described in this thesis. Analysis ideas and approaches were initiated and always performed by the thesis author but discussed with her supervisor Nikola S. Mueller. Results were first interpreted by the thesis author, then further discussed with Nikola S. Mueller and respective collaboration partners. The thesis author did not perform any medical examinations or laboratory experiments in any of the describes studies. The text passages written here can be found again directly before the “x.y.1 Motivation” subsection of each subchapter and chapter, respectively.

In the study described in section 3.1 the thesis author was responsible for data management, data processing and all statistical analyses were performed by her including modeling and association analyses. All figures presented in this section, except the photographs of patients’ skin, were created and designed by the thesis author. The paper Garzorz-Stark et al. (2016) was written jointly together with Natalie Garzorz-Stark. Here, the part on how the molecular classifier was built is described in more detail than in the published paper.

In the study described in section 3.2 the thesis author was responsible for the processing and analysis of all data gathered in humans, starting with data management and quality control of molecular and clinical data. Part of the analysis was building of the human miRNA based asthma classifier using regularized regression modeling. Further associations to clinical attributes were calculated by her. The thesis author was further responsible for interpreting the computational results and designing and creating all figures presented here. The mouse data was analyzed by Nikola S. Mueller and results are summarized here in order to make the study more understandable. All text in the publication Milger et al. (2017) concerning statistical analysis, interpretation and description of results in the human data were originally written by the thesis author. The statistical methods are described here in more detail than in the publication.

In the study described in section 4.1 the thesis author was responsible for and performed all data processing and all statistical analyses including handling of missing data, correlation analyses, differential testing and regression modeling. All figures were designed and created by her. The thesis author wrote all text concerning statistical analyses and interpretations for the original manuscript Krause et al. (2016), its correction and the reply Krause et al. (2017).

1.6 SPECIFIC CONTRIBUTIONS OF THE THESIS AUTHOR IN EACH STUDY

In the study described in section 4.2 the thesis author was responsible for data processing including imputation of missing values and normalization. Further, she came up with the idea to use several modeling approaches and compare the results. The thesis author performed all statistical analyses herself, including all modeling. All figures were designed and created by her. The results were interpreted together with supervisors and collaborators.

In the study described in section 4.3 the thesis author was responsible for all data processing including data normalization, imputation of missing values and batch correction. All statistical analyses including time series clustering, pairwise correlation analyses and regularized regression modeling were designed and performed by the thesis author. Possible networks visualizations were discussed with Nikola S. Mueller and implemented by the thesis author. All figures presented here were designed and created by the thesis author. Results were discussed and interpreted jointly with supervisors and collaborators.

In the study described in section 5.1 the thesis author was responsible for all data analysis of clinical and whole-genome gene expression data. In detail, she processed clinical data and performed correlation analysis of histological attributes of mixed data types to determine objective criteria mostly representing subjective criteria. For the whole-genome gene expression analysis, she analyzed microarray data by performing the whole workflow: data processing, differential expression analysis, pathways analysis using model based gene set analysis and visualization of the results. For differential gene expression analysis the thesis author came up with the idea of using and implemented the usage of linear mixed effect models for microarray data. In particular, she designed the analysis so that the specific medical question could be answered by implementing two models and comparing results. She further integrated *in vitro* gene expression measurements. All figures presented here were designed and created by the thesis author in discussion with her supervisors. The thesis author wrote the original text about analysis and interpretation of computation results for the publication Lauffer et al. (2018).

In the study described in section 5.2 the thesis author was responsible for gene expression analysis. Gene expression was measured with microarrays. Gene expression analysis for this study contained extensive data preprocessing, dimension reduction in a non-standard way and differential gene expression analysis using linear mixed effects models. There are two main differences in the analysis of gene expression data in this study compared to the study described in section 5.1 about interface dermatitis. First, the microarray preprocessing was done more stringently and included adjustment for unknown sources of variability. Second, the linear mixed effects model setup was more complex since more groups were compared. The thesis author further designed and created all figures shown

CHAPTER 1 INTRODUCTION

in this section with a special focus on visualizing pathway analysis results comparing several conditions. The text passages concerning computational methods and results were originally written by the thesis author for the publication Garzorz-Stark et al. (2018).

In the study described in chapter 6 the thesis author was responsible for all data analyses including analysis of secretome and whole-genome gene expression data measured by microarrays. She was also responsible for interpreting the results. Further, she performed clustering analyses using five different clustering algorithms. After going through the full microarray preprocessing workflow, the thesis author performed differential gene expression analysis using six different methods, some of them not commonly used for this task. The thesis author further came up with an elaborate approach to determine most differentially regulated genes per method and how to find consensus top hits. She also performed gene set enrichment analysis and mapping to a protein-protein interaction network. All figures presented in this section were designed and created by the thesis author. The thesis author did not perform any of the laboratory experiments. The biological interpretations of the results were researched and written by the thesis author and approved by her supervisors and biological collaborators.

Chapter 2

Methods

This chapter gives an overview of the applied statistical methods shared among several studies, used data types, studied diseases and available cohorts. All analyses were performed with the software R (R Core Team, 2019). For experimental methods, the laboratory process is summarized here but the focus lies on the resulting data and how different aspects of the data influenced further analytical decisions.

2.1 Statistical analysis

Statistical analysis is useful to understand relationships between variables, detect their dependencies in order to answer biological or clinical questions. It is important to choose the right analytical method depending on the data's characteristics regarding type and implicit biological or clinical background. Here, an overview of statistical analyses, tests and models used in this thesis is presented. Reference for this whole section is Sachs and Hedderich (2009) if not stated otherwise. The analyses were performed with default R packages unless the specific package name is given (R Core Team, 2019). Each chapter and section points to the specific test or modeling approach used for the respective study and indicates if other R packages are used than those described here.

2.1.1 Statistical terminology

A data set consists of samples for which one or more attributes were measured or collected. The samples can be anything from individual people to the collection of cells from a cell line. The attributes are variables which were, throughout this thesis, assessed in the lab, by a clinician or through questionnaires. All variables, also from possibly several time

CHAPTER 2 METHODS

points, taken together for one sample represent one **observation**. For modeling, the variables are split into one **response variable**, which is the variable of interest, and the remaining variables which are called **predictor variables**. Depending on the research question several variables in one data set can be, one by one, the variable of interest which should be understood further and thus becomes the response in modeling.

Observations consist of p -tuples containing information for the outcome of interest and predictor variables, $(y_i, x_{i1}, x_{i2}, \dots, x_{i(p-1)})$ for each of n observations $i \in 1, \dots, n$. Taking the observations together, the response variable $y \in \mathbb{R}^{n \times 1}$ is the outcome of interest which should be modeled using $p - 1$ predictor variables $x_j \in \mathbb{R}^{n \times 1}$. The predictor variables together with the intercept $x_0 = (1, 1, \dots, 1) \in \{1\}^n$ are summarized in a design matrix $X = \{x_0, x_1, x_2, \dots, x_{p-1}\} \in \mathbb{R}^{n \times p}$.

Hypothesis testing

One part of statistical analyses is hypothesis testing. In hypothesis testing, a **null hypothesis** H_0 and a complementary **alternative hypothesis** H_1 is postulated. The alternative hypothesis is formulated so that it includes what the research questions tries to show. For example, in a hypothesis test around the question of height differences between women and men the H_1 is “men are taller than women” and H_0 is “women are taller than men”. Next, data according to the hypothesis is obtained, in this case measurements of the height of a group of women and men is taken. From the collected data, a test statistic specific for the hypothesis is calculated which follows known properties and is directly linked to a **p-value**. The p-value is the probability of obtaining the measured data or more extreme results given that the null hypothesis H_0 is true, also called the type 1 error rate. If the p-value is below a predefined significance level α , often chosen to be 5%, the null hypothesis can be rejected. The result of one hypothesis testing and corresponding error probabilities can be represented in a matrix, where the rows represent the unknown truth and the columns the test’s result:

	test is declared non-significant	test is declared significant
null hypothesis (H_0) is true	correct inference prob. = $1 - \alpha$	type 1 error prob. = α
alternative hypothesis (H_1) is true	type 2 error prob. = β	correct inference prob. = $1 - \beta$

If the hypothesis test does not lead to a correct inference, either a type 1 or a type 2 error has manifested. The probability of a type 1 error is equivalent to the p-value. The type 2 error is represented by β and directly related to the power of a statistical test. The **power** is defined as $1 - \beta$ and is the probability of having a significant test result if the alternative hypothesis is true.

Odds-ratio

Measured variables are numerical or categorical. A categorical variable with only two categories, or levels, is called a **dichotomous** variable. Any two dichotomous variables can be visualized with a contingency table. The table shows an example of one variable describing disease status of an individual in the columns with D representing diseased and H representing healthy. The rows of the table show the other variable attributed to whether the individual was exposed to a specific risk factor where the index E represents exposed and N not exposed. So D_E are some individuals which are diseased and were exposed.

	diseased	healthy
exposed	D_E	H_E
not exposed	D_N	H_N

The **odds-ratio (OR)** is then defined as the fraction of the odds of being a diseased individual in the exposed group divided by the odds of being a diseased individual in the non-exposed group:

$$OR = \frac{D_E/H_E}{D_N/H_N}.$$

The odds-ratio here represents an association between exposure (risk present or not) and outcome (healthy/disease). An odds-ratio smaller than one is attributed to exposures which lower the odds for being diseased. Whereas, an odds-ratio larger than one describes variables whose presence increase the odds for the individual to be diseased. In general, the odds-ratio describes the relationship of two dichotomous variables. In case-control studies and retrospective studies the odds-ratio should be calculated since the incidence (fraction of diseased individuals in the whole population) is not given or known. In prospective studies, or cohort studies, however, the risk ratio (RR) can be calculated (Andrade, 2015)

which is a fraction of probabilities. Since studies in this thesis do not include prospective studies, only the formula for its calculation is given here:

$$RR = \frac{D_E/(H_E + D_E)}{D_N/(H_N + D_N)}.$$

2.1.2 Adjustment for multiple comparisons

When performing several statistical tests on the same data set the false positive rate is increased which should be controlled by the type 1 error called α . There are four possible outcomes in hypothesis testings (see section 2.1.1). If a family of m hypotheses are tested, which is equivalent to m tested comparisons, they sum up as represented in the following table (adapted from Benjamini and Hochberg (1995)):

	test is declared non-significant	test is declared significant	total
null hypothesis (H_0) is true	U	V	m_0
alternative hypothesis (H_1) is true	T	S	$m - m_0$
total	$m - R$	R	m

In this table m is the known number of tested hypotheses and R is observed after those m hypotheses were tested. U , V , T and S are random variables which cannot be observed. The total number true null hypotheses is m_0 . V is a random variable representing the number of tested hypotheses where the null hypothesis is true but the test is declared significant, also called number of false discoveries.

There are four error rates in multiple testing: per-comparison error rate (PCER), family-wise error rate (FWER), per-family error rate (PFER) and false discovery rate (FDR, Benjamini and Hochberg (1995)). Per-comparison error rate ignores the increase in type 1 error in multiple testing and is the standard error rate without any adjustment. If each of the m hypotheses is tested at levels α then:

$$PCER = E(V/m) = \frac{E(V)}{m} \leq \alpha.$$

The family-wise error rate is defined as the probability that m tested hypotheses reveal one or more false discoveries V (Tukey, 1953):

$$\text{FWER} = P(V \geq 1).$$

Bonferroni correction is a method which controls the family-wise error rate by dividing the significance level α by the number of tested hypotheses m (Dunn, 1961). If each hypothesis $i \in 1, 2, \dots, m$ is tested against the level $\frac{\alpha}{m}$, a strong control at level α of the overall family-wise error rate is guaranteed:

$$\text{FWER} = P(V \geq 1) \leq \alpha.$$

Per-family error rate corresponds to the expected number of false positive results among m tested hypothesis (Tukey, 1953) which refers to summing the type 1 error probability for all m hypotheses.

$$\text{PFER} = E(V).$$

It was shown that the control of the per-family error rate also always controls the family-wise error rate and the per-comparison error rate for a fixed significance level α (Dudoit et al., 2003). In other words, a control on the level of per-family error rate is more conservative than on the family-wise error rate which itself is more conservative than the per-comparison error rate.

$$\text{PCER} \leq \text{FWER} \leq \text{PFER}.$$

The false discovery rate is defined as the expected proportion of false discoveries V among all significantly declared tests $V + S = R$ (Benjamini and Hochberg, 1995).

$$\text{FDR} = E(V/(V + S)) = E(V/R).$$

Benjamini and Hochberg (1995) proposed the following procedure to control the false discovery rate:

- m hypotheses H_1, H_2, \dots, H_m are tested with resulting p-values p_1, p_2, \dots, p_m

CHAPTER 2 METHODS

- p-values are ordered to fulfill $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$, where $H_{(i)}$ is the hypothesis for p-value $p_{(i)}$
- let k be the largest i for which $p_{(i)} \leq \frac{i}{m}q^*$, then reject all $H_{(i)}$ for $i \in \{1, 2, \dots, k\}$

They proved that the procedure controls the false discovery rate at level q^* . Comparing the control of false discovery rate and family-wise error rate, Benjamini and Hochberg (1995) showed that if all null hypotheses are true ($m_0 = m$), both are equivalent. In all other cases ($m_0 < m$) the false discovery rate is smaller than or equal to the family-wise error rate. That implies that any correction which controls the per-family error rate or the family-wise error rate also controls the false discovery rate (Hofner et al., 2015).

$$\text{FDR} \leq \text{FWER} \leq \text{PFER}.$$

The false discovery rate rejects at least as many hypotheses as methods controlling the family-wise error rate and therefore has greater power. Controlling the per-family error rate is the most conservative form of multiple testing correction.

All p-values in this thesis were adjusted for multiple testing if required by study designs. If a more stringent adjustment for multiple testing was necessary, the family-wise error rate was controlled with Bonferroni correction. Otherwise, the false discovery rate was controlled due to the increase in power. Unless stated otherwise, all p-values, adjusted and unadjusted ones, are shown in figures with stars indicating * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

2.1.3 Pairwise analysis – finding associations and differences

To analyze associations and differences among all combination of pairs of variables, there is a group of statistical tests to choose from depending on the data type. Variables are broadly differentiated into categorical and numerical variables. For categorical variables, a further separation into dichotomous variables, which have two levels, and those with more categories is important. Numerical variables were investigated whether they can be treated as normally distributed or not and grouped accordingly. Table 2.1 gives an overview of applied statistical tests.

To test for association between two categorical variables on nominal scale applied Fisher's exact test for count data was applied (Fisher, 1935). For two by two tables this test is

Tab. 2.1 – Applied statistical tests. † followed by Dunn’s test (Dunn, 1964). ‡ followed by Tukey’s “Honest Significant Difference” method (Tukey, 1949)

scale of variable 1	scale of variable 2	test name	reference	
categorical	categorical	Fisher’s test	Fisher (1935)	
categorical (2 levels)	numerical (not-normal)	Wilcoxon’s test	Wilcoxon (1945)	
categorical (>2 levels)	numerical (not-normal)	Kruskal Wallis test	Kruskal and Wallis (1952)	†
categorical (2 levels)	numerical (normal)	Welch’s t-test	Welch (1947)	
categorical (>2 levels)	numerical (normal)	analysis of variance (ANOVA)	Fisher (1925a)	‡
numerical (not-normal)	numerical (not-normal)	Spearman’s correlation coefficient	Spearman (1904)	
numerical (normal)	numerical (normal)	Pearson’s correlation coefficient	Pearson (1895)	

equivalent to testing the hypothesis that the odds-ratio between both variables is equal to one and the p-value is directly acquired from the hypergeometric distribution. If one or both variables have more than two levels, p-values are calculated using internal C implementations of FORTRAN routines from Mehta and Patel (1983) which are included in the applied R version.

For testing associations between categorical and numerical variables four different statistical tests were used. If the numerical variable was not normally distributed or had values above or below the detection limit, two different tests depending on the categorical variable were applied. If the categorical variable was dichotomous, Wilcoxon’s rank sum test was used (Wilcoxon, 1945). Otherwise, Kruskal Wallis rank sum test was applied (Kruskal and Wallis, 1952). A significant association between categorical and numerical variable in the Kruskal Wallis test only tells, that there are at least two categories between which the location parameter is different. To determine between which pairs of categories the difference is hidden, the corresponding post-hoc test was performed, Dunn’s test (Dunn, 1964) implemented in the R package *dunn.test* (Dinno, 2017). For numerical variables which are

CHAPTER 2 METHODS

normally distributed Welch’s two-sample t-test was applied (Welch, 1947), if the categorical variable showed two categories, or otherwise analysis of variance (ANOVA, Fisher (1925a)). Of note, it is no longer recommended to first test for equal variances between two groups with the Levene test and then decide to choose either Student’s t-test or Welch’s t-test but instead to directly apply Welch’s t-test (Zimmerman, 2004). Similarly to the case of Kruskal Wallis test, also the analysis of variance can be followed by an appropriate post-hoc test to extract between which two groups there is a difference. Tukey’s ‘Honest Significant Difference’ method (TukeyHSD, Tukey (1949)) was applied which performs pairwise tests while correcting for the family-wise error rate.

The concept of correlation was first introduced 130 years ago by Sir Francis Galton (Galton (1889), historic overview in Stigler (1989)) and developed further by Pearson (1895). In order to compare two approximately normally distributed numerical variables Pearson’s product moment correlation coefficient was applied. It is calculated by dividing the covariance between the variables by the product of their standard deviations and is abbreviated with the letter r . If both variables are numerical, but not normally distributed or contain values beyond the detection limit, they are compared with Spearman’s rho statistic which estimates a rank-based measure of association (Spearman, 1904).

Network visualization of pairwise associations

For better interpretability, results of pairwise statistical tests can be visualized in form of networks. After adjusting all comparisons for multiple testing, a network consisting of nodes and edges is plotted. Nodes represent measured variables in the data set. Edges connect two nodes if the pairwise statistical test between both variables represented by the respective nodes revealed a significant association between both. The level of significance is coded through edge thickness, meaning the stronger the association between two variables, the thicker the edge between them.

2.1.4 Linear regression

To interrogate the linear interplay of several variables with regard to one numerical outcome of interest, linear regression analyses were applied. Linear regression was introduced by Galton (1886) who compared heights between individuals and their parents and first proposed “regression towards the mean”. In linear regression, the response $y \in \mathbb{R}^{n \times 1}$ is a numeric variable. The response y is modeled as a linear combination of predictor variables

summarized in the design matrix $X = x_0, x_1, x_2, \dots, x_{p-1} \in \mathbb{R}^{n \times p}$ and coefficients β_j for $j \in 0, 1, \dots, p-1$ and additive noise ϵ .

For one observation y_i the linear model is

$$\begin{aligned} y_i &= \beta_0 \cdot x_{i0} + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_{(p-1)} \cdot x_{i(p-1)} + \epsilon_i \\ &= \sum_{j=0}^{p-1} \beta_j x_{ij} + \epsilon_i \\ &= x_i \beta + \epsilon_i \end{aligned}$$

for $y_i, \beta_j, x_{ij}, \epsilon_i \in \mathbb{R}^{1 \times 1}$, $x_i \in \mathbb{R}^{1 \times p}$ with $i \in \{1, 2, \dots, n\}$, $j \in \{0, 1, \dots, (p-1)\}$
and $\beta \in \mathbb{R}^{p \times 1}$.

Taking n observations together is leads to

$$\begin{aligned} y &= \beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_{(p-1)} \cdot x_{(p-1)} + \epsilon \\ &= \sum_{j=0}^{p-1} \beta_j x_j + \epsilon \\ &= X \beta + \epsilon \end{aligned}$$

for $\beta_j \in \mathbb{R}^{1 \times 1}$ and $x_j \in \mathbb{R}^{n \times 1}$ with $j \in \{0, 1, \dots, (p-1)\}$, and
 $y \in \mathbb{R}^{n \times 1}$, $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^{p \times 1}$, $\epsilon \in \mathbb{R}^{n \times 1}$.

There are four main assumptions in linear regression. First, the relationship between response and predictors is linear in the parameters β . Second, dispersion of y is independent of X . Third, the observations $(y_i, x_{i1}, x_{i2}, \dots, x_{i(p-1)})$ are independent and consequently, the residuals ϵ_i and ϵ_j are independent for $i \neq j$. Fourth, the residuals ϵ_i are normally distributed with mean zero (summarized in Sachs and Hedderich (2009)). Following from these assumptions and with E being the expected value it holds:

$$E(y) = X\beta.$$

The coefficients β_i in linear regression are not known but estimated from the observations. If the predictor variables are independent of each other, ordinary least squares can be used to estimate the coefficients. For n observations, not more than n variables can be

independent. So solving linear regression problems with ordinary least squares is only possible if the number of predictors p is less or equal to the number of observations n . Ordinary least squares was first published by Legendre (1805), however, Carl Friedrich Gauss claims to have used it before in his paper in 1799 (Stigler, 1981) and summarized the mathematical background in Gauss (1809). Ordinary least squares minimize the residual squared errors $\epsilon^\top \epsilon$ and gives an unbiased estimator of the true, underlying coefficients.

Regularized regression models

The coefficients β in linear regression are only estimable with ordinary least squares if the predictor variables are independent. In the case of more predictors p than observations n , the requirement is not fulfilled. One mathematical approach to overcome this problem are regularized regression models like ridge regression (Hoerl and Kennard, 1970), the least absolute shrinkage and selection operator (the “lasso”, Tibshirani (1996)) and combinations thereof like the elastic net (Zou and Hastie, 2005). All three methods apply a regularization term on the estimated coefficients β and solve for regularized linear regression the equation

$$\min_{\beta} \left(\sum_{i=1}^n \left(y_i - \sum_{j=0}^{p-1} \beta_j x_{ij} \right)^2 \right) \quad \text{subject to } f(\beta) \leq t,$$

where $t \geq 0$ is a tuning parameter which controls the amount of shrinkage applied to the estimated coefficients. For ridge regression the penalty is an L^2 -norm, for the lasso it is the L^1 -norm and elastic net combines both L^1 - and L^2 -norm:

$$\begin{aligned} \text{for ridge regression:} \quad & f(\beta) = \|\beta\|_2^2 = \sum_j \beta_j^2 \\ \text{for the lasso:} \quad & f(\beta) = \|\beta\|_1 = \sum_j |\beta_j| \\ \text{for elastic net:} \quad & f(\beta) = \frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \\ & = \frac{1-\alpha}{2} \sum_j \beta_j^2 + \alpha \sum_j |\beta_j|. \end{aligned}$$

The lasso produces coefficients which are exactly zero (Tibshirani, 1996), whereas ridge regression shrinks all coefficients to a value larger than zero. The elastic net parameter $0 \leq \alpha \leq 1$ determines the amount of mixing between lasso and ridge regularization.

2.1.5 Linear mixed effects models

Linear models can be extended to linear mixed effects models which include not only fixed effects β but also random effects b (Fisher, 1919). They are particularly useful in the case of repeated measurements, so if the observations (y_i, x_i) are not independent. X and Z are design matrices relating y to β and b . The model can be represented as:

$$y = X\beta + Zb + \epsilon \quad \epsilon \sim N(0, \sigma^2 I) \quad b \sim N(0, \tau^2 I)$$

with $y \in \mathbb{R}^{n \times 1}$, $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^{p \times 1}$, $Z \in \mathbb{R}^{n \times s}$, $b \in \mathbb{R}^{s \times 1}$, $\epsilon \in \mathbb{R}^{n \times 1}$.

Since $b \sim N(0, \tau^2 I)$ the expected value of y is still $X\beta$.

One application of linear mixed effects models is the analysis of gene expression data where individuals are measured repeatedly for several conditions. Expression of one gene is modeled as numeric response y , the conditions shared among individuals are fixed effects β and random effects b are the individuals. Random effects b , in this case, are random intercepts which represent the baseline expression of one gene in one particular individual. The fixed effects β are shared effects among individuals, which are corrected for the baseline expression of each individual, for each of the $p - 1$ predictors.

In this thesis, linear mixed effects models were estimated using the *lme4* package in R (Bates et al., 2015). P-values for each coefficient were obtained using “mixed” function from the *afex* R-package (Singmann et al., 2018). For the degrees of freedom it applies the Kenward-Roger approximation (Kenward and Roger, 1997). Individuals were represented by random effects and shared conditions as fixed effects.

2.1.6 Generalized linear models

The linear regression concept can be generalized to a larger class of models called generalized linear models (Nelder and Wedderburn, 1972) which can be written in the following form

$$g(E(y)) = X\beta.$$

In the equation $g(\cdot)$ is the link function which connects the expectation of the response value $E(y)$ to the linear predictor $X\beta$. In case of linear regression the link function equals the identity function.

Logistic regression

If the response variable is not numeric, but dichotomous, it follows a Binomial distribution with success probability π and the model is called logistic regression (Cox, 1958). The aim in logistic regression is to estimate the unknown success probability π . The linear model approach $\pi = \sum_j \beta_j x_j$ is transformed by the logistic transformation to fulfill the requirement that probabilities are bound by zero and one:

$$\pi(X) = \frac{e^{\sum_j \beta_j x_j}}{1 + e^{\sum_j \beta_j x_j}} = \frac{e^{X\beta}}{1 + e^{X\beta}}.$$

Application of the logistic transformation is justified by its following properties: if $\beta_j > 0$ the function value increases monotonically with increasing x_j ; if $\beta_j < 0$ the function value is monotonically decreasing with decreasing x_j ; if $\beta_j = 0$ then x_j does not have an influence on the success probability π . Comparing the logistic transformation to the form of generalized linear models, the logit transformation is used as the link function since it connects the success probability of the response y to the linear predictor,

$$\text{logit}(\pi(X)) = \log\left(\frac{\pi(X)}{1 + \pi(X)}\right) = \log(\text{odds}(\pi)) = X\beta$$

$$\text{with } \text{odds}(x) = \frac{x}{1 + x}.$$

The cumulative normal distribution is another possibility for a link function in case of a Binomial response (Bliss, 1935). The regression is called “probit” regression. Both differ in the correct interpretation of the estimated coefficients. In logistic regression the coefficients can be calculated back to odds-ratios, whereas in probit regression the coefficients are interpreted as differences in standard normal values.

In linear regression the optimal coefficients β are directly estimated from the observations. In logistic regression this is not possible, so coefficients are calculated as maximum likelihood estimates (Wilks, 1938) using iteratively reweighted least squares, also called Fisher scoring algorithm (Fisher, 1925b; Green, 1984).

There are more application of generalized linear models, e.g. Poisson regression. This overview is limited to the applied methods in this thesis

2.1.7 Random forest

One ensemble learning technique with increasing popularity is random forests (Breiman, 2001). Random forests take a different approach to classification, compared to regression analyses (sections 2.1.4 - 2.1.5). The core idea is to grow many different decision trees and combine them in a random forest. To obtain different decision trees a random vector is necessary which brings variability to each tree. Examples are bagging (Breiman, 1996), where for every tree only a subset of observations are used to grow the tree, and random split selection (Dietterich, 1998), where on every node in the decision tree the split is randomly chosen among k best splits. Predictions in random forests are the majority vote of all trees in the random forest. Besides classification, random forests can also solve regression tasks with regression forests.

2.1.8 Model evaluations

Once a model is set up, its performance needs to be assessed with quality-based error estimations. In linear regression the **residual sum of squares** R^2 describes the fraction of variance in the response explained by the model and is used for model evaluation:

$$R^2 = 1 - \frac{\sum_{i=1}^n \epsilon_i^2}{\sum_{i=1}^n (y_i - \frac{1}{n} \sum_{i=1}^n y_i)^2}.$$

The closer R^2 is to one, the more variance is explained by the predictor variables. The more predictor variables are included in the model, the larger R^2 becomes. The **adjusted residual sum of squares** $R_{\text{adj.}}^2$ corrects for the number of predictors $p - 1$ in the model (Theil, 1961):

$$R_{\text{adj.}}^2 = 1 - \frac{n-1}{n-p} \cdot (1 - R^2),$$

CHAPTER 2 METHODS

where R_{adj}^2 equals R^2 if there are no predictors in the model, otherwise it is smaller than R^2 . Another criterion for the evaluation of a model is the **Akaike information criterion** (AIC, Akaike (1974)):

$$AIC = -2 \log(\text{likelihood}) + 2 \cdot p,$$

where the likelihood is the probability of the data given the estimated coefficients. The aim is to determine a model with a small value for the AIC. The equation balances between number of predictors $p - 1$ and goodness of fit of the model.

In logistic regression and other binary classification tasks the model performance can be visualized with confusion matrices comparing true and predicted classes.

		true class	
		class 1	class 0
predicted class	class 1	true positive (TP)	false positive (FP)
	class 0	false negative (FN)	true negative (TN)

From the confusion matrix, different diagnostics can be calculated:

$$\text{sensitivity} = \frac{TP}{TP + FN} = \text{true positive rate}$$

$$\text{specificity} = \frac{TN}{TN + FP} = \text{true negative rate}$$

$$1 - \text{specificity} = \frac{FP}{FP + TN} = \text{false positive rate.}$$

To determine the predicted class, a threshold has to be applied to the predicted value since it is numeric not categorical. A way to choose the optimal threshold is the receiver operator characteristics (ROC) curve. For the ROC curve, true positive rate is plotted against false positive rate for different thresholds (example in Fig. 3.6 A). The optimal threshold is where the true positive rate is the highest and the false positive rate the smallest. To assess a model's performance, the **area under the ROC curve (AUC)** is calculated. A binary classification model performs better the closer the AUC is to one. The AUC can be calculated for any binary classifier.

In the optimal case, the performance of the model is tested in an independent cohort. AUC values are stated for binary classification and root mean squared prediction errors

for linear regression are calculated with the model applied on data from the independent cohort. If there is no access to an independent cohort, **cross validation** (CV) can be applied. In k -fold cross validation, the observations are first split into k groups. Then, for each k the first $k - 1$ groups are combined into the training set and the k -th group is left as test set. The model is trained using only the observations in the training set. The model is evaluated on the left-out test observations. Comparing true and predicted values in the test sets is used to assess the model performance. If the model is evaluated with AUC, the result of k -fold cross validation are k AUC values. One particular variant of cross validation is leave-one-out cross validation (LOO-CV), where only one observation is left out in every cross validation fold.

2.1.9 Feature selection procedures

In many studies more predictors were measured than those that influence the outcome. To determine those predictors which are important for the correct prediction of the outcome, **feature selection** can be performed. Selected features are always a subset of all available predictors. There are different approaches to feature selection, those applied in this thesis are described here, separately per model setup.

In linear regression the model can be optimized by **all subset regression**. The method builds separate models for all possible combinations of predictors and calculates the adjusted residual sum of squares R_{adj}^2 . It is implemented in the *leaps* package in R (Lumley, 2017). The optimal combination of predictors is determined by the model with the highest R_{adj}^2 . This approach is only suitable for the case of more observations n than predictors $p - 1$.

For logistic regression with more observations n than predictors $p - 1$ **AIC-based stepwise optimization** is an option (Chambers et al., 1992). In every step, the method chooses to include or remove the one predictor which results in the largest decrease of AIC. It is important to note, that in each step only one predictor is included in the model or removed from the model. The method stops if there is no more decrease in AIC. Stepwise optimization can either start with an intercept-only model or with the full model containing all predictors and move from these starting points to an AIC-optimized model. All predictors remaining in the optimized model are selected features.

For regularized regression models with the lasso or the elastic net penalty, several coefficients are set to zero due to regularization constraints. So all predictors with **non-zero coefficients** represent selected features. Cross-validation is applied to determine the regularization parameter in lasso and elastic net via *glmnet* package in R (Friedman

CHAPTER 2 METHODS

et al., 2010). This leads to slightly different results in different runs. To obtain a stable result, stability selection can be employed (Meinshausen and Bühlmann, 2010). It selects influential variables while controlling the per-family error rate PFER. The R package *stabs* was used to perform stability selection (Hofner and Hothorn, 2017).

Random forests were not designed to perform variable selection but to build accurate classifiers (Breiman, 2001). Nevertheless, Breiman (2001) introduced measures to rank the **predictors by importance**. Strobl et al. (2007) showed that these measures are biased towards predictors with more categories. Altmann et al. (2010) overcame this limitation by correcting for the bias with a heuristic based on permutation tests called “permutation importance” and reported a p-value for each predictor. The p-values are calculated by permuting the response and estimating the variable importance per predictor in a setting where the predictor is non-informative. Degenhardt et al. (2017) compared feature selection methods for random forests applied on biomedical data sets. Their recommendations are Vita (Janitzka et al., 2015) for high dimensional data and Boruta (Kursa et al., 2010) for low dimensional ones. The Boruta approach determines all relevant predictors by adding “shadow variables”, which are permuted versions of the original predictors, to the set of predictors. The importance attributed to the shadow variables are used as a reference to detect relevant predictors. The results showed variability due to fluctuations so results of three different runs were included.

2.1.10 Clustering methods

Clustering aims at grouping observations together into clusters based on similarities in the observed values while maximizing dissimilarity to other clusters. It was first applied by Driver and Kroeber (1932) who worked in the field of anthropology and subsequently employed in psychology (Zubin, 1938; Tryon, 1939). The principle idea is that within a cluster observations are more alike to each other than to members of other clusters. Clusters are useful for the discovery of unknown structures in the data or identification of subgroups of predictors or observations. Clustering is a broad concept and not defined by one specific algorithm. The choice of method and parameters depend on the data set and the designated use of the clustering result.

In clustering analysis three main questions need to be answered. How many clusters are in the data set? How to define similarity between observations which implies the choice of distance function? Which clustering algorithm is best suited to answer the biological question?

Number of clusters

To determine the number of clusters the silhouette approach introduced by Rousseeuw (1987) was followed. It is a measure which combines tightness within a cluster and separation between clusters. This approach should be used if the aim is to find compact and separated clusters (Rousseeuw, 1987). For any given clustering and distance measure, one value, the average silhouette coefficient, is obtained by first calculating the silhouette S of observation o which belongs to cluster A while B and C represent other clusters:

$$S(o) = \begin{cases} 0 & \text{if } \text{dist}(A, o) = 0 \\ \frac{\text{dist}(B, o) - \text{dist}(A, o)}{\max(\text{dist}(A, o), \text{dist}(B, o))} & \text{else} \end{cases}$$

using

$$\text{dist}(A, o) = \frac{1}{n_A} \sum_{a \in A} \text{dist}(a, o) \quad \text{and} \quad \text{dist}(B, o) = \min_{C \neq A} \left(\frac{1}{n_C} \sum_{c \in C} \text{dist}(c, o) \right).$$

The silhouette coefficient of cluster A is defined as the mean of all n_A silhouettes of cluster A . Taking the mean of these coefficients for all clusters determines the average silhouette coefficient. The optimal number of clusters k is the clustering with the highest average silhouette coefficient.

Distance function

The definition of a silhouette already uses the pairwise distances ($\text{dist}(a, o)$) between observations a and o . Three different distance functions were applied: Euclidean distance, a distance measure based on correlation and a distance measure for time-series data. The Euclidean distance between two observations $o_i \in \mathbb{R}^{1 \times p}$ and $o_j \in \mathbb{R}^{1 \times p}$ for $i, j \in 1, \dots, n$ is the distance according to the L^2 -norm:

$$\begin{aligned} d_{\text{Euclidean}}(o_i, o_j) &= \|o_i - o_j\|_2 = \sqrt{(o_i - o_j)^2} \\ &= \sqrt{(o_{i1} - o_{j1})^2 + (o_{i2} - o_{j2})^2 + \dots + (o_{ip} - o_{jp})^2} . \end{aligned}$$

The distance measure based on correlation is defined as

$$d_{\text{cor.-based}}(o_i, o_j) = \frac{1}{2} \cdot (1 - \text{cor}(o_i, o_j)) \text{ ,}$$

which fulfills the requirement that distances are bound by zero and one. For the value $\text{cor}(o_i, o_j)$ either Pearson's or Spearman's correlation coefficient was used depending on data types.

Distance function for time series data

For dissimilarities between time series data, there are several distance measures described in the literature and implemented in R, e.g. in the *TSclust* R-package (Montero et al., 2014). Measures like the Minkowski distance are invariant to time permutations since for every time point $t \in 1, \dots, p$ the distance between two observations $o_i \in \mathbb{R}^{1 \times p}$ and $o_j \in \mathbb{R}^{1 \times p}$ is evaluated separately and summed up to give overall distance between both observations. But the Minkowski distance is sensitive to shifting and scaling of data and does not cover patterns over time. Another measure is dynamic time warping distance (Kruskal, 1983; Berndt and Clifford, 1994) which aims a finding a mapping between o_i and o_j so that the distance between them is minimized. This approach can handle shifting and scaling but it ignores the temporal ordering of data points.

One measure which covers both distance of absolute values and proximity in temporal behavior was introduced by Chouakria and Nagabhushan (2007) and called d_{CORT} . In this thesis, the implementation of d_{CORT} in the *TSclust* R-package was used (Montero et al., 2014). To calculate the measure, first the proximity between the time course behavior of both observations is evaluated by the first order temporal correlation coefficient called *CORT*:

$$\text{CORT}(o_i, o_j) = \frac{\sum_{t=1}^{p-1} (o_i(t+1) - o_i(t))(o_j(t+1) - o_j(t))}{\sqrt{\sum_{t=1}^{p-1} (o_i(t+1) - o_i(t))^2} \sqrt{\sum_{t=1}^{p-1} (o_j(t+1) - o_j(t))^2}},$$

where $o_i(t)$ denotes the value of observation o_i on the time point t . If $\text{CORT}(o_i, o_j)$ is equal to one, both time series show the same pattern over time which means that the increases or declines are at all time points the same in direction and rate. The dissimilarity d_{CORT} introduced by (Chouakria and Nagabhushan, 2007) takes the first order temporal correlation coefficient *CORT* and multiplies it with the distance between the raw values of both observations, e.g. calculated using euclidean distance over all time points:

$$d_{CORT}(o_i, o_j) = \phi_k(CORT(o_i, o_j)) \cdot d(o_i, o_j),$$

where $\phi_k(\cdot)$ is an adaptive tuning function. Chouakria and Nagabhushan (2007) proposed to apply the exponential adaptive function:

$$\phi_k(u) = \frac{2}{1 + \exp(ku)} \quad \text{with} \quad k \geq 0.$$

Clustering algorithm

Which algorithm to choose for clustering is the third question which needs to be answered. There are many algorithms for clustering but in this thesis only those popular and widely used in the field of biomedical data analysis and those applicable for the data sets at hand were used. The algorithms fall into three categories: distribution-based clustering, centroid models and hierarchical clustering. In distribution-based clustering, observations are grouped into clusters if they are more likely to belong to the same distribution. One distribution-based clustering method is **Gaussian mixture models** where k Gaussians are fit to the observations using expectation maximization (Dempster et al., 1977). Expectation maximization is a generalized maximum likelihood approach for observations with some unobserved variables. Gaussian mixture models were estimated with *mclust* package in R (Scrucca et al., 2017) which optimizes shape, volume and orientation of the Gaussians.

Centroid models are the second group of clustering algorithms which were considered. Among them is **k-means** which was developed by Steinhaus (1956) and named by MacQueen et al. (1967). It is still very popular and widely used (Jain, 2010). The main idea is to minimize the squared Euclidean distance between all points in that cluster and the mean of each cluster, the cluster center or centroids. This is equivalent to minimizing the sum of variances within each cluster. The algorithm iteratively changes the cluster centers by adjusting the correspondence of each observation to the nearest cluster center. K-means is a greedy algorithm in the sense that it only finds a local minimum with each initialization. In this thesis, k-means was executed at least 10,000 times to overcome this limitation and the run with the smallest error was chosen which is the pre-implemented result in R's k-means version.

Another centroid based clustering algorithm is **partitioning around medoids** (pam, Kaufman and Rousseeuw (1987)) which represents a robust version of k-means. There are two main differences between pam and k-means. First, for pam the cluster centers are data

CHAPTER 2 METHODS

points and not arbitrary mean values like in k-means. Second, pam does not minimize the squared Euclidean distances but the sum of dissimilarities between cluster centers and observations within that cluster. The algorithm converges with any distance function. In this thesis, it was used with Euclidean distances.

The third group of algorithms are **hierarchical clustering** algorithms. Here, the aim is not to group the observations into a fixed number of clusters but rather to detect a hierarchical structure in the data. There are divisive (top-down) and agglomerative (bottom-up) hierarchical clustering methods (Friedman et al., 2001). In divisive clustering all observations are at the beginning in one cluster and successively split into smaller clusters based on similarity. In this thesis, no divisive clustering methods were applied. In agglomerative clustering every observation is at the beginning its own cluster and consecutively the two most similar clusters are merged. To perform the merging step, a measure for the distance of two clusters is needed. Two different approaches were applied: complete linkage and Ward’s minimum variance method. In **complete linkage** the distance between two clusters A and B is calculated as the maximum distance between any two observations from the two clusters (Johnson, 1967; Lance and Williams, 1967):

$$d_{\text{complete linkage}}(A, B) = \max_{\substack{i \in A \\ j \in B}} \text{dist}(o_i, o_j).$$

The definition implies that in complete linkage clustering two clusters are only considered close to each other and thus merged if all observations in both clusters are similar (Friedman et al., 2001). This leads to the detection of similar clusters. **Ward’s method**, on the other hand, finds compact, spherical clusters by merging clusters which lead to the minimal increase in total within-cluster variance (Joe and Ward, 1963). Different implementation of Ward’s minimum variance method exist. In the original publication dissimilarities are squared which is implemented in the method “ward.D2” in R which was applied in this thesis.

For most projects, the aim of clustering was to visualize the observations or predictors in a structured manner. To achieve this, hierarchical clustering based on Euclidean distances or correlations was used. Only if the grouping was of importance for further analytical steps, different algorithms were considered. When analyzing the results of different algorithms, **consensus clustering** was performed to determine a clustering independent of one specific clustering method but shared among several or all algorithms. Consensus clustering assesses whether groups of observations are always clustered together by several algorithms.

2.2 Gene expression measurements and processing

Gene expression describes the amount of mRNA present in a cell or in a collection of cells. There are three predominant ways to measure gene expression levels: quantitative polymerase chain reaction (qPCR, Higuchi et al. (1993)), gene expression microarrays (Schena et al., 1995) and RNA sequencing (Nagalakshmi et al., 2008). In this thesis, only data obtained from the first two methods was analyzed.

In short, qPCR is used to determine the levels of a small, hand picked number of genes. Microarrays, in contrast, cover the whole transcriptome according to annotations in databases. In RNA sequencing the mRNA is converted to coding DNA (cDNA) fragments. The sequence of those fragments is directly obtained and computational methods are employed to assess gene expression levels.

2.2.1 Quantitative polymerase chain reaction technology and data processing

Quantitative polymerase chain reaction (qPCR) is a targeted method to determine the levels of a single specific nucleotide sequence (Higuchi et al., 1993). It is a quantitative advancement of polymerase chain reaction (PCR, Mullis et al. (1986)). Through cycles of heating and cooling and the use of sequence specific primers the amount of nucleotide sequence as input is roughly doubled in every step. A fluorescent molecule which binds to double stranded nucleotide sequences is added to the mix at the beginning. As soon as a threshold of fluorescence is reached this specific cycle is termed the Ct-value (cycle threshold value). The Ct-value can be translated to the true amount of nucleotides at the beginning using an evenly expressed, house keeping gene as normalization. The process can be parallelized to determine expression levels of several nucleotide sequences. These nucleotide sequences can be protein-coding genes but also micro RNAs.

If a gene of interest X is measured in two conditions, e.g. lesional and noninvolved skin, and in parallel in a house keeping gene like 18S, the expression levels can be quantified using the $\Delta\Delta Ct$ method:

$$\begin{aligned}\Delta Ct &= Ct_{\text{geneX}} - Ct_{18S} \\ \Delta\Delta Ct &= \Delta Ct_{\text{noninvolved}} - \Delta Ct_{\text{lesional}} \\ \text{expression level}_{\text{geneX}} &= 2^{\Delta\Delta Ct}.\end{aligned}$$

CHAPTER 2 METHODS

The exponentiation in the last step is due to the fact that the amount is doubled in every cycle. Exponentiation introduces a shift of the values towards a log-normal distribution. So before analyzing the expression levels, a log10 transformation was performed.

If no house keeping genes is known or measured, a different normalization approach was performed. To make the levels comparable, all pairwise ratios of n Ct-values were calculated:

$$\Delta Ct_{ij} = Ct_{\text{gene } i} - Ct_{\text{gene } j} \quad \text{for } i \in \{1, 2, \dots, (n - 1)\}, j \in \{2, 3, \dots, n\}.$$

2.2.2 Microarray technology

If the target is not a specific set of genes a different approach than quantitative polymerase chain reaction for determining the gene expression levels in a sample is needed. Gene expression microarrays are a possible tool since they allow to assess the levels of all transcripts which are annotated in databases (first developed by Schena et al. (1995), explanation taken from Ewis et al. (2005)). Although, there are different microarray brands, in this thesis, only microarrays from Agilent Technologies were used due to their broad coverage. In principle, single stranded 60 nucleotide long probes are attached to a glass slide in clusters of identical probes. Fluorescently labeled single stranded RNA is added and binds to the probes if the nucleotides align to form a double strand. As a read-out, signals are measured with the “iScan microarray scanner” and further processed with “Agilent Feature Extraction Software” (Agilent Technologies).

A crucial step in measuring gene expression is the RNA isolation and quality control. The collaboration partners who acquired the data which was analyzed in this thesis followed the advice and protocols given by Agilent Technologies and the other manufacturers. Total RNA was isolated with a Qiagen kit. For quality control RNA yield and quality were determined with a NanoDrop from Thermo Fisher. As a final quality control step, RNA integrity numbers were assessed. Only samples with RNA integrity numbers larger than six were labeled with Cy3, amplified with oligo dT-promoter primers and hybridized on SurePrint G3 Human GE 8x60K BeadChips from Agilent Technologies. In all studies described in this thesis, the comparison of more than two conditions was the aim, so Cy3 labeling was used and measurements were taking using one-color microarrays, also called single-channel microarrays. In case of only two conditions of interest, one condition is labeled with Cy3 and the other one with Cy5. Both are measured together on the same two-color microarray giving the possibility of direct comparison between both conditions.

2.2.3 Computational processing of Agilent microarrays

Before microarrays were processed, their quality was evaluated. First, the quality control output of Agilent's Feature extraction software was checked for errors in the fluorescent detection and extraction by the collaboration partners. If mistakes occurred on the technical level the sample was measured again. Second, the thesis author used the *arrayQualityMetrics* package (Kauffmann et al., 2009) in R (R Core Team, 2019) to calculate several metrics evaluating the quality of each array relative to the others within one data set. Microarrays were excluded from the analysis if they were strikingly different in at least three metrics.

After quality control, all microarrays used in one study were processed together and analyzed in parallel. Processing was performed with several functions from the *limma* package in R (Ritchie et al., 2015). Microarray expression intensities were corrected for background fluorescence with the "backgroundCorrect" function applying the "normexp" method which is recommended by Ritchie et al. (2007) when local background estimates are available. After correcting for background intensities, the expression values were normalized. Since the data consisted of one-color microarrays between-array normalization was performed and not within-array normalization as would be done for two-color microarrays. Quantile normalization between arrays was applied which is recommended in Bolstad et al. (2003) and forces the empirical distribution of each sample to be the same. The function "normalizeBetweenArrays" performs quantile normalization and further log₂-transforms the expression values.

One possible, further processing step is exclusion of lowly expressed probes after background correction and normalization. To filter out lowly expressed probes, only probes which were 10% brighter in any of the arrays than the 95% quantile of all negative control probes are kept.

The Agilent gene expression microarray SurePrint G3 Human GE 8x60K consists of 62,976 probes which are spotted on glass slides. Of these 4,259 are negative and positive control probes. Within the remaining 58,717 probes, 34,085 have only one measurement on the microarray, 7,321 are measured twice and 999 are measured ten times. Taken together, there are 42,405 different probes spotted on the Agilent microarray. Within-array probe expressions were only averaged for identical probes with the "avereps" function. No average across probes which refer to the same gene but have different nucleotide sequences was performed since they possibly have different dynamical ranges or detect different variants of the same gene.

CHAPTER 2 METHODS

Oligonucleotide probes on microarrays are designed to correspond to transcripts. For the relation between probe identifier and genes, three mappings were investigated: gene expression omnibus' version under GPL14550, information from Agilent's internal website and ensembl's mapping accessed via *biomaRt* R-package (Durinck et al., 2009). Gene symbols provided by the Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) were used which are thought to be unique and meaningful. If the probe-gene symbol assignments did not agree among the first three, the mapping provided by ensembl was chosen as the most up to date version.

One possible, further processing step is a more stringent mapping of probes to exclude probes with low confidence. To exclude probes with low confidence, the thesis author applied "blastn" (Altschul et al., 1990) to map the 60 base pair long nucleotide sequences spotted on the microarray to the human transcriptome available from USCS via Genbank (February 2016) and encoded as RefSeq IDs. Out of all 42,405 probes, 13,430 did not match with total accuracy to one position of the human transcriptome and were removed prior to analysis. For probes mapping to more than one RefSeq ID, it was checked whether these IDs corresponded to the same gene (via gene symbol). In this step, 2,229 probes were excluded since they mapped to several genes. Mapping between RefSeq ID and gene symbol was performed with *org.Hs.eg.db* annotation package from Bioconductor's AnnotationDbi (Pagès et al., 2018). In total, 26,746 probes mapped with complete accuracy to one unique gene and were used in further analysis.

2.2.4 RNA sequencing and comparison to microarray technology

Microarrays were already developed in the 90s (Schena et al., 1995) but still used in daily research. They have three major limitations Bumgarner (2013). First, arrays only give an indirect measure of relative mRNA concentrations. Second, it is complicated to design arrays where each probe is only bound by one specific molecular sequence and not by several RNA/DNA sequences from the same gene family. Also determining alternative splicing is difficult if not all exons are uniquely detected with the array (Gardina et al., 2006). Finally, microarrays can only detect what is annotated in databases, so novel variants or genes cannot be determined.

Due to the limitations other methods were developed, especially RNA sequencing, first introduced by Nagalakshmi et al. (2008), is getting more and more popular. When performing RNA sequencing the mRNA is first converted to coding DNA (cDNA) fragments and ligated to an adapter sequence. All of those fragments are sequenced and computational methods are employed to map those fragments to the transcriptome. The sequence

2.2 GENE EXPRESSION MEASUREMENTS AND PROCESSING

fragments are typically between 50 and 500 base pairs long. Since every fragment is sequenced, absolute abundances are obtained and novel transcripts and splice variants can be detected. RNA-seq also enables the determination of polymorphisms in the RNA sequences (Mantione et al., 2014).

Even though RNA sequencing has many advantages, currently microarrays are still cheaper and reliable in model organisms (Mantione et al., 2014). Chen et al. (2017) showed that for most transcripts (89.8%) results obtained by RNA sequencing or microarrays can be reproduced in The Cancer Genome Atlas data. Both methods should compliment each other in scientific endeavors.

2.2.5 Gene set analysis

Analysis of gene sets is a useful tool to put lists of genes into a context and understand their relationships. For different studies different gene set databases were used depending on the study, clinical question and suitability. Three gene set and pathway databases were used: reactome (Fabregat et al., 2017), wikipathways (Slenter et al., 2018) and gene ontology (Gene Ontology Consortium, 2016).

For cases where the gene set of interest is known because of the clinical question at hand, the thesis author proposed an intuitive visualization. For comparing expression of genes within one gene set among two or more conditions, fold changes for all gene within the gene set were visualized using barplots. For this approach, fold changes should have been calculated relative to the same baseline for all groups. Fold changes of different conditions for the same gene were aligned horizontally while the vertical axis showed all genes in the gene set of interest (examples for two groups in Fig. 5.4 and 5.7, for three groups in Fig. 5.8 and 5.9). To quantify the visual similarity of expression for two conditions within one gene set correlation coefficients between fold changes were calculated.

If the gene set of interest is not known, enrichment analysis with two methods were performed: model-based gene set analysis (MGSA, Bauer et al. (2010)) and over-representation analysis. In over-representation analysis a statistical test is performed. The null hypothesis of no enrichment of differentially regulated genes in one pre-defined gene set is tested via e.g. Fisher's exact test. One crucial point is to define the background set, where all genes which were measured by the Agilent microarray which passed quality control were used. The enrichment test is calculated for all gene sets of interest and associated p-values are corrected for multiple testing. Another angle on gene set enrichment analysis, is a Bayesian model-based approach introduced by Bauer et al. (2010) where the hierarchical structure

CHAPTER 2 METHODS

of gene sets is taken into account in the model estimation. The results are posterior probabilities of each gene set being active or not, where a cut-off of posterior probability of 0.5 is recommended.

In cases where several gene lists obtained from different omics levels are jointly tested for gene set enrichment, the Multi-level ONtology Analysis (MONA) algorithm can be applied. It is model-based Bayesian method which infers probabilities for each GO term in a modular framework (Sass et al., 2013).

2.2.6 Mapping to protein-protein interaction networks

To gain deeper understanding of interactions between genes, corresponding proteins were mapped to protein-protein interaction networks. Several interaction network databases are available with different characteristics. For example, STRING is a large resource of over 5,000 organisms, 24.6 million proteins and more than 2,000 million interactions among them (Szklarczyk et al., 2016). It includes experimentally validated protein-protein interactions and also computationally predicted interactions or proteins pairs which are only functionally related (Snider et al., 2015). BioGRID is another protein-protein interaction database which provides only curated interactions (Chatr-Aryamontri et al., 2017). Currently, 1.6 million interactions are available in BioGRID in over 80 organisms which also include viruses.

In this thesis, mapping to protein-protein interaction networks was applied for results calculated for immune cells. So, proteins were mapped to the database InnateDB provided by Breuer et al. (2012) because the database is based the IMEx interactome (Orchard et al., 2012) and specifically curated for immune-related proteins (Snider et al., 2015). All protein-protein interactions in InnateDB have been experimentally validated.

During mapping of proteins of interest, only first-order interactions were of interest. First-order interactions include all proteins which are known from the database to directly interact with the queried protein.

2.2.7 Dimension reduction of gene expression data

In order to visualize high dimensional data, dimension reduction techniques were applied. Principal component analysis (PCA) is one popular method (Jolliffe, 2011) first introduced by Pearson (1901) and further developed by Hotelling (1933). It aims to project high

2.2 GENE EXPRESSION MEASUREMENTS AND PROCESSING

dimensional data consisting of n observations of p variables to a lower dimensional space (l dimensions) without losing much information. Data is represented in form of a matrix $X \in \mathbb{R}^{n \times p}$ where n is the number of observations and p the number of measured variables, e.g. genes. X is assumed to be centered by column, meaning the mean value of each variables is subtracted for all observations. PCA is a linear transformation where the first principal component gives the direction of highest variance, the second principal component the direction of second highest variance, and so on. If the data is normally distributed, then principal components are statistically independent of each other. If not, the components are decorrelated but statistically dependent. PCA is defined by weights $v_k \in \mathbb{R}^{p \times 1}$ which project each row i of X , called $x_i \in \mathbb{R}^{1 \times p}$, for $i \in 1, \dots, n$ to the principal component scores $t_i \in \mathbb{R}^{1 \times l}$ following this equation:

$$t_i(k) = x_i \cdot v_k \quad \text{for } i \in 1, \dots, n \text{ and } k \in 1, \dots, l.$$

Finding the first principal component in classical PCA solves the following equation for $v_1 \in \mathbb{R}^{p \times 1}$:

$$\max_v (v^\top X^\top X v) \quad \text{subject to } \|v\|_2^2 \leq 1.$$

Where $X^\top X$ represents to covariance matrix of the data and v_1 its eigenvector to the largest eigenvalue. Following the same principle, the k -th weight vector v_k for the k -th principal component is calculated by subtracting $k - 1$ components from X to form \hat{X}_k :

$$\hat{X}_k = X - \sum_{s=1}^{k-1} X v_s v_s^\top.$$

And then by solving the same equation:

$$v_k \text{ solves: } \max_v \left(v^\top \hat{X}_k^\top \hat{X}_k v \right) \quad \text{subject to } \|v\|_2^2 \leq 1.$$

Lin et al. (2016) extended the principal component framework by adding a regularization term. Including the regularization allows to adjust for confounding variation during dimension reduction (short AC-PCA) in the following way: let c be the number of confounders, then $Y \in \mathbb{R}^{n \times c}$ represents the confounding matrix, centered by column and $K = Y^\top Y$. Y is chosen so that $v^\top X^\top K X v$ describes the confounding variation in the projection. The PCA can be modified with:

$$\max_v (v^\top X^\top X v - \lambda v^\top X^\top K X v) \quad \text{subject to } \|v\|_2^2 \leq 1,$$

where $\lambda \geq 0$ gives the strength of regularization. The parameter λ is chosen so that confounding variation is relatively small in comparison to total variation. For $\lambda = 0$ it corresponds to classical PCA. If λ is large enough the subspace for the optimization problem is restricted to be orthogonal to the columns of Y Lin et al. (2016). If $Z = X^\top X - \lambda X^\top K X$, then the maximization problem of AC-PCA can be solved by finding an eigendecomposition of Z .

There are many more dimension reduction algorithms with new ones being regularly introduced, e.g. the Uniform Manifold Approximation and Projection for Dimension Reduction, short UMAP (McInnes et al., 2018). While PCA or multidimensional scaling (Kruskal, 1964) preserve distance structure within the data, other algorithms preserve local distances rather than global distances (McInnes et al., 2018). Algorithms like t-distributed stochastic neighbor embedding (Maaten and Hinton, 2008), diffusion maps (Coifman and Lafon, 2006) and also UMAP (McInnes et al., 2018) fall into the latter category and are getting more and more popular, especially in single cell RNA-seq analysis (Haghverdi et al., 2015; Becht et al., 2018).

2.2.8 Batch correction methods

Batch effects can arise in measured data due to technically different handling of the measurements. Examples are different reagents, technicians or varying laboratory conditions (Leek et al., 2010). Especially, in whole genome gene expression data batch effects are often present (Johnson et al., 2007). In this thesis, two different approaches for handling batch effects were applied. One for known batches and one for unknown sources of variability.

If the batch is known for the data, e.g. because data was measured on different days, ComBat was applied which was particularly introduced for microarray data by Johnson et al. (2007). ComBat uses an empirical Bayes framework to adjust for batches which is robust to outliers and applicable for small sample sizes. The core idea is to borrow information across measured variables, e.g. genes, since it is assumed that a batch effect has similar influence on several variables (Johnson et al., 2007). In this thesis, the implementation of ComBat in the R-package *sva* was used (Leek et al., 2012).

For the other case, if the batch is not known, surrogate variable analysis was applied (Leek and Storey, 2007, 2008) for cases where unknown sources of variation influencing the

2.2 GENE EXPRESSION MEASUREMENTS AND PROCESSING

measurements were suspected, e.g. because of grouping in principal component analysis which was not associated to any known covariate (see section 2.2.7). Surrogate variable analysis also borrows information across variables, e.g. genes, to estimate effects of all unmodeled factors. This estimation of heterogeneity is directly performed on the data in five steps (Leek and Storey, 2007). First, the signal of the primary outcome of interest, the effect which is investigated, is removed from the data and a residual matrix is obtained. Second, the residual matrix is decomposed using an orthogonal basis of singular vectors which identifies signatures of underlying heterogeneity. Third, each singular vector is tested for whether it represents more variation than expected by chance. Forth, for each singular vector, the subset of genes which drive the heterogeneity is identified. Fifth, for those subsets of genes a surrogate variable is built using the signatures of underlying heterogeneity on the original data. After the surrogate variables were constructed, they were included as covariates in the subsequent regression analysis. In this thesis, the iteratively re-weighted least squares implementation of surrogate variable analysis in the *sva* R-package was used Leek et al. (2012).

2.3 Measurement and processing of secretome expression

Tjalsma et al. (2000) first introduced the idea that all molecules secreted by a cell or organism are important to study and termed them the *secretome*. There are several methods to measure the secretome. In an *in vitro* design, cells often need an external stimulus to secrete the proteins into the surrounding media which is extracted and subjected to measurement. In humans, proteins are secreted by various cells and are also transported via the blood where they can be measured. In this thesis, two types of secretome measurements are used: enzyme-linked immunosorbent assays (ELISA) and multiplexed immunoassays. In short, both methods are targeted, antibody-based and fluorescence is the detected output which has to be translated to concentration using standard values. With ELISA only one protein at a time can be measured whereas the multiplexed technology provides a platform where tens of proteins are determined in parallel.

The collaboration partners who acquired the data analyzed in this thesis chose to apply targeted methods since cytokines which are known to be important for signaling within the immune system (Parkin and Cohen, 2001) are lowly expressed in human serum compared to functional plasma proteins like serum albumin and apolipoproteins (Geyer et al., 2017). They wanted to detect cytokines in human serum with targeted approaches since there are more than 100 serum markers which were approved or cleared by the FDA as clinical tests for various human diseases (Anderson, 2010).

2.3.1 Enzyme-linked immunosorbent assays (ELISA) technology

Enzyme-linked immunosorbent assays (ELISA) were introduced by Engvall and Perlmann (1971) and Van Weemen and Schuurs (1971) who showed first that enzymes can be used as reporter labels. There are different kinds of ELISA. The collaboration partners who acquired the data analyzed in this thesis applied so-called “Sandwich-ELISA” which relies on two antibodies specific for two different epitopes of the protein of interest (Bidwell et al., 1977). The first antibody, the capture antibody, is attached to a surface (Fig. 2.1). When the sample is added, the capture antibody binds to the protein of interest. After washing, another antibody specific to the protein is added and also binds the protein, thus forming a “sandwich” of two antibodies binding the same protein. To enable detection, an enzyme-linked secondary antibody is added that binds the constant region of the second antibody. When adding the enzyme’s substrate, the substrate is converted and a fluorescent signal is emitted which is detected.

2.3 MEASUREMENT AND PROCESSING OF SECRETOME EXPRESSION

Requirements for ELISA are two antibodies binding to different epitopes of the protein of interest. Only one protein can be determined in one reaction chamber. Standards with known concentrations are measured in parallel, to translate the emitted fluorescence into concentration values. If the protein concentration in the sample is far beyond the standard concentrations, the translation to concentration is not possible and the limit of detection is reached.

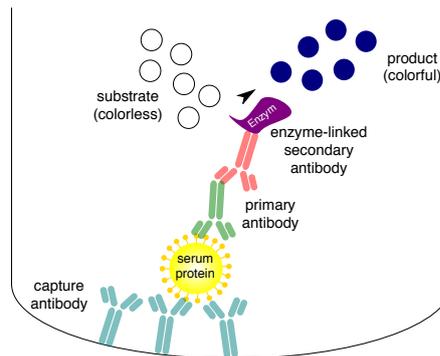


Fig. 2.1 – Enzyme-linked immunosorbent assays technology. Capture (light blue) and primary antibody (light green) are specific for the protein of interest (yellow). A secondary, enzyme-linked antibody (light red) binds the constant region of the primary antibody. When adding the enzyme’s substrate (white circle), the substrate is converted (dark blue) and a fluorescent signal is emitted which is detected. Figure was adapted from wikipedia.de/ELISA and was licensed under CC BY-SA 4.0.

2.3.2 Processing ELISA measurements

If secreted proteins were measured with ELISA and the aim of the study was to compare concentration values across proteins, the analysis was performed with translated concentrations. The main reason were the concentrations which lay outside of the detection limits and thus translation into concentration values was not possible. Since only one or a few proteins were measured simultaneously, imputation methods like multivariate imputation by chained equations (MICE, Raghunathan et al. (2001); Van Buuren (2007)) which use the whole data set for computing imputations were not applicable. Instead, a processing scheme was introduced which was oriented along the seven standard values measured in parallel with each protein. The protein concentrations were transformed to classes one to ten following the concept visualized in Table 2.2. Kotsiantis et al. (2006) advise to use discretization when performing machine learning tasks like clustering to improve effectiveness. Choosing the optimal internal border for discretization remains an open problem (Kotsiantis et al., 2006).

Tab. 2.2 – Preprocessing scheme for ELISA measurements making use of standard values.

ELISA measurement	transformation
below lower limit of detection	1
above lower limit of detection but below 1st standard	2
between i -th and $(i+1)$ -th standard for $i \in \{1, 2, \dots, 7\}$	$i + 2$
below upper limit of detection but above 7th standard	9
above upper limit of detection	10

2.3.3 Multiplex immunoassays

One caveat of ELISA is the constraint of determining only one protein at a time. The Luminex’s Multi-Analyte Profiling (xMAP) technology enables the measurement of several proteins at the same time (explanations taken from (Dunbar, 2006)). The key difference to ELISA is that the capture antibody is not attached to a surface but to a microscopic bead (Fig. 2.2). The bead itself is internally fluorescently labeled with a unique combination of two dyes and coated with specific antibodies. Proteins from the sample bind the antibody and a second fluorescent detection antibody is added. The detection is based on flow cytometry where two lasers with different wavelengths either excite the dye combination in the beads or the molecular tag attached to the detection antibody. In parallel, the dye in the bead tells which antibody is bound to it, so which protein is measured, and the molecular tags allows to measure the concentration of the protein. For translating fluorescent information into concentration values standards with known concentrations are used.

The collaboration partners who acquired the data analyzed in this thesis used the Bio-Plex Pro™ Human Cytokine 27-plex Assay which is build on Luminex xMAP technology. It uses magnetic beads and assesses the concentration of 27 proteins in human serum. Magnetic beads are shown to perform better than polystyrene beads since they reduce non-specific binding, are more robust and reproducible (Moncunill et al., 2013).

2.3.4 Processing of multiplex immunoassays

The output of multiplex immunoassays is fluorescence-intensity based and mostly not normally distributed. Another problem are out of range values. Concentrations of more than half of the samples within one study was outside the detection limits for several proteins. Depending on the clinical question different processing steps were chosen. If the interest was solely in the comparison of groups within the same protein, rank-based tests

2.3 MEASUREMENT AND PROCESSING OF SECRETOME EXPRESSION

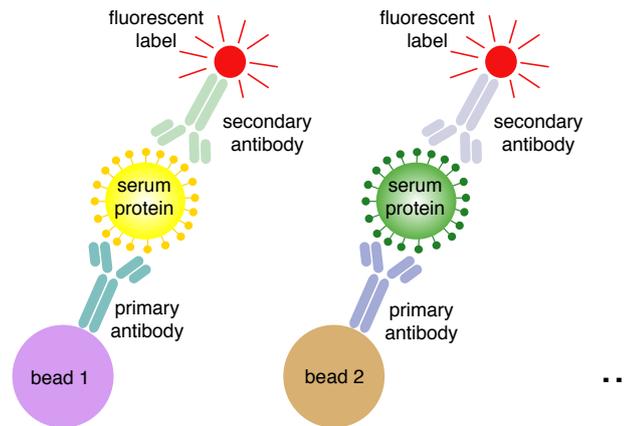


Fig. 2.2 – Multiplex immunoassay technology. The main difference to ELISA (Fig. 2.1) is that the primary antibody is not attached to a surface but to a fluorescently labeled bead. This enables the measurement of several different proteins simultaneously in one reaction chamber. Figure was adapted from wikipedia.de/ELISA and was licensed under CC BY-SA 4.0.

were applied. In these rank-based tests, the translated concentrations based on standard values were directly used and for all samples above the detection limit the highest rank was assigned and for those below the lower detection limit the smallest rank was assigned. For the other studies, where the interest was in the relationship among the proteins and statistical models were calculated, a two step processing scheme was applied.

The first step of the processing dealt with those proteins, where more than a specific percentage of samples in one data set at hand had values outside the detection limits. The thesis author, together with her supervisors and collaboration partners decided to set this percentage to 20%. Those proteins were discretized to the values zero and one for both directions. For proteins where more than 20% of the samples were above the upper limit of detection, the class zero represents samples where a concentration was measured below upper limit of detection and class one represents samples above the upper limit of detection. Vice versa, for proteins where more than 20% of the samples were below the lower limit of detection, class zero represents those sample below lower limit of detection and class one those where a concentration above the lower limit of detection was measured.

For all proteins which were not discretized first log₁₀-transformations were applied and samples beyond detection limits were imputed with GSimp (Wei et al., 2018). GSimp is an iterative algorithm based on a Gibbs sampler for left and right censored missing value imputation. The method can be summarized in four steps. First, the missing values are initialized with a method of choice. For left censored values quantile regression imputation of left censored data was chosen (Lazar et al., 2016) and for right censored data a method

CHAPTER 2 METHODS

which draws samples from the left tail of the distribution and symmetrically transforms them to the right tail was chosen as recommended by the authors. Second, a prediction model is build where the variable with missing values is the response and other variables are covariates. The prediction model is calculated with elastic net. Third, the prediction model is used to calculate an estimate of the response and an error between initialized value and predicted value. In a forth step, the model estimate and error is used to sample the missing element from a truncated normal distribution. Steps two to four are repeated and the values are updated for a set amount of iterations. The thesis author used default numbers of iterations.

The thesis author, together with her supervisors and collaboration partners chose to impute proteins measured with multiplex technologies since tens of proteins were measured simultaneously. The applied imputation approach borrows information from other proteins to stabilize the imputation result. When analyzing ELISA measurements (section 2.3.2) the number of proteins was much lower, so the imputation results were less stable and the transformation scheme described in Table 2.2 was applied.

As a comparison, Do et al. (2018) chose to not include metabolites in their correlation analysis if more than to 70% measurements were outside the detection limits. Sterne et al. (2009) discussed a study where a variable with 70% missingness was imputed and led to clinically wrong results. The cutoff of 20% is conservative compared to 70% also due to the fact that not hundreds of variables were measured in parallel like in Do et al. (2018) but only around 30.

2.3.5 Batch correction

Similarly to batch effects described for gene expression data (see section 2.2.8), these effects can also manifest in other kinds of measurements like in secretome data. The same procedure as described in section 2.2.8 was followed. In short, if the batch is known, e.g. because of different recruitment periods of patients, these known batches were adjusted for using ComBat (Johnson et al., 2007) which is implemented in the *sva* package by Leek et al. (2012). If the batch is not known, the thesis author tested for unknown latent sources of variation using surrogate variables analysis (Leek and Storey, 2007, 2008) which is also implemented in the *sva* package Leek et al. (2012). If the method detect unknown sources of variation they are represented as covariates which can be included in the regression modeling.

2.4 Overview of studied diseases

2.4.1 Atopic Eczema and psoriasis

Psoriasis and atopic eczema are both inflammatory skin diseases. They can arise if the balance of Th cell subsets is distorted. They are not only defined by the Th cell imbalance, but by several more factor, which makes them multifactorial, complex diseases. During the last years, modern technologies and computational analysis of whole genome expression data have changed insights in the pathogenesis of both diseases (Tsoi et al., 2012; Bieber, 2008; Perera et al., 2012). In parallel, novel therapies which target specific cytokines and other disease mediators have been developed for both diseases (Langley et al., 2014). These new therapies increasingly replace therapies which suppress the whole immune system (Schön, 2014; Beck et al., 2014). So far, the establishment of biomarkers which predict the success of these therapies is still challenging mainly because of the heterogeneity of both diseases. The heterogeneity is also the reason for difficulties of differentiating between psoriasis and eczema during diagnosis (Aydin et al., 2008). Wrong diagnosis of either disease and subsequently wrong therapy can lead to worsening of both diseases.

Atopic Eczema

Atopic eczema is common, chronic inflammatory skin disease. In patients with atopic eczema more Th2 cells were detected in affected skin in relation to Th1 cells (Bieber, 2008). Therapeutic effects in atopic eczema are measured by scores such as the “severity scoring of atopic dermatitis” (SCORing atopic dermatitis, short: SCORAD). Clinical variants and subtypes of eczema which were analyzed in this thesis are nummular eczema, palmoplantar and scalp variants and erythroderma.

Recently, multiple studies tried to identify disease biomarkers for atopic eczema in human serum (Zhang et al., 2013; Thijs et al., 2015a, 2017a). Atopic eczema is related to systemic inflammation and increased inflammatory marker proteins are measured in the serum of patients (Brunner et al., 2017).

Atopic eczema usually starts in infancy with a prevalence of 20% in children which reduces to 2%-10% in adults (Weidinger and Novak, 2016; Garmhausen et al., 2013). Risk factors for a persistence disease course have been described in the literature (Von Kobyletzki et al., 2015; Kiiski et al., 2015) but it is not possible to predict the disease course at an early age.

Psoriasis

Psoriasis is an inflammatory skin disease which is still underdiagnosed and undertreated (Lebwohl et al., 2016). One reason is the noteworthy heterogeneity in the clinical presentation of psoriasis (Boehncke and Schön, 2015). There are several clinical variants and various subtypes of psoriasis. In this thesis, data from patients with guttate psoriasis and inverse psoriasis was analyzed. In patients suffering from psoriasis an imbalance towards more Th1, Th17 and Th22 cells compared to Th2 cells was observed in the lesional skin (Nestle et al., 2009).

2.4.2 Interface dermatitis

Interface dermatitis is a type of skin inflammation which is present in several inflammatory and autoimmune skin diseases. It is also called lichenoid tissue reaction. Examples for diseases with interface dermatitis skin inflammation are lichen planus, lupus erythematosus, dermatomyositis, fixed drug eruption and many others. It is characterized by immune cell infiltration, enlarging of cells and death of keratinocytes (Sontheimer, 2009).

2.4.3 Allergic asthma

Asthma is characterized by a chronic inflammation of the airways. It is further associated with airway hyperresponsiveness and airflow obstruction. Diagnosis can be difficult, in particular for children in pre-school age and in elderly people. In addition, asthma is a heterogeneous disease with different phenotypes (Wenzel, 2012).

2.5 Clinical data and its processing

In this thesis, matched clinical data was utilized to analyze molecular cohort data of patients. Clinical data comes in a variety of formats, data types and value ranges. Information is stored in numerical and categorical variables with two or multiple levels. Thus, careful investigation and adjustment is needed to integrate clinical information with molecular data.

Other challenges in clinical data analysis are standardization and missing values. Standardization is of most importance if different cohorts are combined. Missing data, however, is almost always present in clinical information of cohort data. Here, a differentiation between random and systemic missingness is necessary. Missing at random means that the probability for a missing value does not depend on unobserved values but only on observed ones (Schafer and Graham, 2002). If this is not fulfilled then the data has systemic missingness and bias.

Processing of clinical data

Collection of clinical data is not standardized but depends on the collaboration partner and investigated disease. For this thesis, only analyses were performed where all available data was collected by one experimental group, so standardization across cohorts was not in issue. Yet, this point should be considered when comparing the results to similar studies. For the handling of categorical data, the thesis author, together with her supervisors and collaboration partners chose carefully whether each categorical level should be regarded as independent or whether the levels have an intrinsic ordering. Categorical variables with more than two levels often have an intrinsic ordering and can be transformed so that rank-based methods are applicable. If enough levels were available and those were approximately equidistant multi-level ordered categorical variables were treated as numerical predictors in regression analyses. Depending on the variable ranges numerical clinical variables were either log-transformed or used as is.

Since statistical modeling with missing data is difficult and not possible with many first-line approaches like principal component analysis and linear regression, the thesis author, together with her supervisors tried to overcome this challenge by following two ideas. First, going back to clinicians to determine whether there is a possibility to obtain the missing data. Second, imputing data with multivariate imputation by chained equations (MICE, Raghunathan et al. (2001); Van Buuren (2007)). MICE is an iterative algorithm where

CHAPTER 2 METHODS

each variable is imputed separately based on a regression model where the variable is the response and all remaining variables are predictors. The implementation in the R package *mice* from van Buuren and Groothuis-Oudshoorn (2011) was applied.

The decision how variables were processed, standardized and imputed was jointly made together with the respective clinical partners.

2.6 Overview of studies, cohorts and data sets

To answer the research questions introduced in section 1.4 one or more cohorts or data sets were analyzed. All measurements were performed in the labs of the respective collaboration partners.

First research question: disease classifiers

The first research question asked whether robust and interpretable molecular disease classifiers can be found. Together with Natalie Garzorz-Stark and Kilian Eyerich from Department of Dermatology and Allergy at the University Clinic of TU Munich (short: TUM Derma), a molecular disease classifier to differentiate between psoriasis and eczema was established by means of disease subtypes using 129 patient samples (section 3.1). In collaboration with Katrin Milger from the Institute of Lung Biology and Disease at Helmholtz Center Munich (short: Helmholtz Lung) and Susanne Krauss-Etschmann from Research Center Borstel (short: Borstel), a molecular disease classifier for allergic asthma was set up using micro RNA measurements in two independent patient cohorts (section 3.2).

affiliation	samples	molecular data	clinical data (# variables)
TUM Derma	129 patients with psoriasis or atopic eczema	expression of 2 genes	histology, anamnesis (64)
Helmholtz Lung, Borstel	46 allergic asthma patients, 21 healthy controls	expression of 13 miRNAs	anamnesis (29)

Second research question: serum proteins as surrogates

The second research question asked whether serum proteins can be used for disease monitoring and prognosis. Together with Stefanie Eyerich from the Center of Allergy and Environment of the Technical University of Munich and the Helmholtz Zentrum München (short: ZAUM), 30 serum proteins were associated to disease severity in 52 atopic eczema patients for disease monitoring (section 4.1). In collaboration with Veronika Baghin and Natalie Garzorz-Stark from TUM Derma, 33 serum proteins and IgE levels in 124 children were analyzed to determine markers for disease prognosis (section 4.2). Together with Karolin Dehlke and Katrin Hoffmann from Heidelberg University Hospital

CHAPTER 2 METHODS

(short: Heidelberg) and Ursula Klingmüller from the German Cancer Research Center (short: DKFZ), patient prognosis after liver resection in 30 patients were investigated using time-course data of eight serum proteins and twelve blood tests (section 4.3). Serum proteins were measured with the Luminex xMAP technology implemented in the Bio-Plex Pro™ Human Cytokine 27-plex Assay or with single plexes. Single secreted proteins very measured with ELISA.

affiliation	samples	molecular data	clinical data (# variables)
ZAUM	52 atopic eczema patients, 20 healthy controls	30 serum proteins	severity score (1)
TUM Derma	124 children with atopic eczema	33 serum proteins, IgE levels	anamnesis and questionnaire (59)
Heidelberg, DKFZ	30 patients after liver resection surgery	time-course of 8 serum proteins & 12 blood tests	hospital stay information, surgery course (46)

Third research question: adjusting patient-bias

The third research question asked whether inter-individual variability in patients can be adjusted for when analyzing whole genome gene expression data measured with microarrays. All gene expression microarrays were Agilent Technologies SurePrint G3 Human GE 8x60k BeadChip. Together with Felix Lauffer and Kilian Eyerich from TUM Derma, gene expression data in 32 patients with interface diseases were investigated and adjusted for inter-individual variability using linear mixed effects models. Whole biopsy gene expression data from patients was integrated with gene expression measurements of keratinocytes in five different stimulation conditions with each three replicates (section 5.1). In collaboration with Natalie Garzorz-Stark, Felix Lauffer and Kilian Eyerich from TUM Derma and Stefanie Eyerich from ZAUM, gene expression in imiquimod-induced skin reactions of 18 subjects was also analyzed using linear mixed effects models (section 5.2). For both projects, results were compared to gene expression data measured in a cohort of 49 patients with different inflammatory skin diseases published by the collaborators from TUM Derma and ZAUM in Quaranta et al. (2014b).

2.6 OVERVIEW OF STUDIES, COHORTS AND DATA SETS

affiliation	samples	molecular data	clinical data (# variables)
TUM Derma	32 interface dermatitis patients	gene expression microarray †	histology (24)
TUM Derma	3 keratinocyte samples each for 5 conditions	gene expression microarray †	—
TUM Derma, ZAUM	18 subjects with imiquimod-induced skin reactions	gene expression microarray †	diagnosis (1)
TUM Derma, ZAUM	49 inflammatory skin disease patients	gene expression microarray †	diagnosis (1)

† gene expression microarrays measured 42,405 probes, see section 2.2

Forth research question: characterizing T cells

The forth research question asked whether new marker genes for T helper cell subsets can be obtained. Together with Stefanie Eyerich from ZAUM, four secreted proteins of 79 T helper cell clones were analyzed to group cells into T helper cell subsets. Those results were integrated with levels of 27 secreted proteins measured by Luminex xMAP technology for a subset of 75 T helper cell clones. Whole genome gene expression data of 79 clones in both stimulated and unstimulated condition was analyzed to determine new marker genes for each of the subsets (section 6).

affiliation	samples	molecular data
ZAUM	79 T cell clones in stimulated condition	levels of 4 secreted proteins measured by ELISA
ZAUM	75 T cell clones in stimulated condition	levels of 27 secreted proteins measured by Luminex
ZAUM	79 T cell clones in unstimulated and stimulated condition, together 158 samples	gene expression microarray †

† gene expression microarrays measured 42,405 probes, see section 2.2

Chapter 3

Building and characterizing disease classifiers in inflammatory skin diseases and allergic asthma

Novel technologies and methodologies enabled a more detailed understanding of human diseases. Among other implications, this led to new, and more specifically targeted medications. However, the development of specific biomarkers, able to classify patients correctly or monitor disease progression, is an ongoing process. This chapter describes two studies where the primary aim was to differentiate between two groups of subjects using few molecular markers. The secondary aim was to make these markers interpretable by associating them with clinical features.

In the first study, the thesis author built a disease classifier which differentiates between psoriasis and atopic eczema, two inflammatory skin diseases, together with her supervisors and her collaborators at the Department of Dermatology and Allergy at the University Hospital of Technical University Munich. The classifier is based on gene expression levels of two genes: NOS2 and CCL27. Those two genes were described in earlier work published by Quaranta et al. (2014b). Here, the focus is not on the development of the classifier itself but on its applicability in subtypes of psoriasis and eczema and the association of the markers with clinical and histological attributes (section 3.1). Results were published in 2016 in the *Journal of Experimental Dermatology*. Recently, the framework was extended to paraffin-embedded skin samples opening this diagnostic tool to dermatological practices with no access to state-of-the-art experimental laboratories. This extension is still in the developmental stages but ethics is approved and collection of over 700 paraffin-embedded skin samples starts soon.

CHAPTER 3 DISEASE CLASSIFIERS

N. Garzorz*, **L. Krause***, F. Lauffer, A. Atenhan, J. Thomas, S.P. Stark, R. Franz, S. Weidinger, A. Balato, N.S. Mueller, F.J. Theis, J. Ring, C.B. Schmidt-Weber, T. Biedermann, S. Eyerich and K. Eyerich: *A novel molecular disease classifier for psoriasis and eczema*. *Experimental Dermatology* (2016).

* equal contribution

The second study described in this chapter (section 3.2) used a two step approach to identify a miRNA-based classifier for allergic asthma. First, miRNAs were screened for differential expression in two mouse models of asthma. The statistical identification of candidate miRNAs was performed by Nikola S. Müller. In the second step, those candidate miRNAs were measured in two human cohorts with patients suffering from allergic asthma and healthy controls. Regularized logistic regression was applied to determine the optimal combination of miRNAs in a training cohort which was then tested in an independent test cohort. Resulting miRNAs were associated to clinical attributes to assess whether they differentiated asthma subphenotypes. The study was conducted with the collaborators at Research Center Borstel and the Institute of Lung Biology and Disease at Helmholtz Center Munich.

K. Milger, J. Götschke, **L. Krause**, P. Nathan, F. Alessandrini, A. Tufman, R. Fischer, S. Bartel, F.J. Theis, J. Behr, S. Dehmel, N.S. Mueller, N. Kneidinger and S. Krauss-Etschmann: *Identification of a plasma miRNA biomarker-signature for allergic asthma: a translational approach*. *Allergy* (2017).

3.1 Molecular classifier for psoriasis and eczema

The study was conducted in collaboration with Natalie Garzorz-Stark and Kilian Eyerich from the Department of Dermatology and Allergy at the Technical University Munich. Molecular data consisted of gene expression levels of two previously described genes which separate psoriasis from eczema patients. The classifier was validated in three cohorts and the gene expression levels were associated to histologic attributes, anamnestic, clinical and laboratory markers. Content, text and data of this section are based on Garzorz-Stark et al. (2016). Copied text passages are indicated as quotations.

In this study the thesis author was responsible for data management, data processing and all statistical analyses were performed by her including modeling and association analyses. All figures presented in this section, except the photographs of patients' skin, were created and designed by the thesis author. The paper Garzorz-Stark et al. (2016) was written jointly together with Natalie Garzorz-Stark. Here, the part on how the molecular classifier was built is described in more detail than in the published paper.

3.1.1 Motivation

Psoriasis and atopic eczema are inflammatory skin diseases which are treated with diverse specific therapies (section 2.4.1). Predicting clinical response to treatment is still challenging because both diseases are heterogeneous. Already diagnosing both diseases is difficult (Aydin et al., 2008). Misdiagnosis of either disease can lead to severe impairment of the disease.

There has been previous work trying to build molecular classification systems for psoriasis and eczema. Guttman-Yassky et al. (2009) published a prediction model which uses the expression 13 different genes. Researchers also tried to build multi-disease classifiers in the field of inflammatory skin diseases. For example, Inkeles et al. (2015) used microarray data from more than 300 samples from people suffering from 16 skin disease to establish such a classifier. But, all these classifiers have a rather large size in terms of needed genetic markers and reducing them remains a challenge.

“We recently defined characteristic pathways and key players for psoriasis and eczema by analyzing a unique patient group which - suffering concomitantly from psoriasis and eczema - represent an excellent model to study inflammatory responses independent of genetic background and environmental influences prior to tissue sampling (Eyerich et al., 2011; Quaranta et al., 2014b). Based on these findings, several marker combinations to diagnose

CHAPTER 3 DISEASE CLASSIFIERS

psoriasis or eczema were proposed, among them the combination of NOS2, the inducible nitric oxidase synthase which produces NO upon stimulation by proinflammatory cytokines (Kanwar et al., 2009) and CCL27, the cutaneous T-cell-attracting chemokine eliciting a crucial role in T-cell-mediated inflammation (Homey et al., 2002). Here, we validated this molecular classifier [...] for practical clinical use.” (Garzorz-Stark et al., 2016)

The classifier was applied to a large cohort of patients ($n = 129$) suffering from psoriasis or eczema or both, and also several clinical subtypes of both diseases and clinically and histologically unclear patients. Finally, the markers were set in context with known histological and clinical attributes to characterize them further.

3.1.2 Methods

A detailed description of the patient cohort and laboratory methods can be found in Garzorz-Stark et al. (2016). In this thesis section, the focus is on the statistical basis for defining the molecular classifier.

Patients and material sampling

This study presents data of in total 129 patients in three different cohorts. The first cohort contained patients with plaque psoriasis ($n = 45$) and eczema ($n = 43$). The second cohort contained 31 patients with different variants and subtypes of both diseases. The third cohort consisted of patients with unclear diagnosis. Some patients have been previously published in Quaranta et al. (2014b). For all patients anamnestic, clinical, histological and laboratory variables were gathered. Biopsies were taken from lesional and in 119/129 cases also in noninvolved skin and divided in half. One part was analyzed using histology and from the other part RNA was isolated for gene expression analysis.

Molecular classifier

RNA from lesional and noninvolved skin was isolated and transcripts of 18S (house keeping gene), NOS2 and CCL27 from lesional and noninvolved skin were measured by quantitative polymerase chain reaction (section 2.2.1). Data were expressed as mRNA fold change, relative to house keeper and noninvolved skin as calibrator, which will be termed “normalized” expression (section 2.2.1 for formula). The molecular classifier was built using a logistic regression (section 2.1.6) with normalized and log10-transformed levels of NOS2

3.1 MOLECULAR CLASSIFIER FOR PSORIASIS AND ECZEMA

and CCL27 as covariates $x_j \in \mathbb{R}^{n \times 1}$ for $j \in 1, 2$ and disease state (psoriasis or eczema) as response variable $y \in \mathbb{R}^{n \times 1}$. The model was trained on the first cohort, $n = 88$ patients with clear diagnosis, and tested on patients from the second and third cohort, suffering from subphenotypes of psoriasis and eczema or with unclear disease state. To infer the robustness of the model, a 10-fold cross-validation (section 2.1.8) was performed in the first cohort.

For patients where there is no information about gene expression in noninvolved skin a pool of mean expression of CCL27 and NOS2 in noninvolved skin was established and these values were used as calibrators for normalization of NOS2 and CCL27 transcripts in lesional skin. A second classifier was trained on the same patients but all values were normalized to the pooled values. This classifier can be used on patients without information from noninvolved skin.

Statistical analysis

“To correlate expression of NOS2 and CCL27 with clinical and histological features, data type appropriate statistical tests were used [(see section 2.1.3)]: statistical significance for categorical features with two levels was determined using Welch’s t-test, and for those with more than two levels, analysis of variance was applied. For features on the interval scale, significance was determined using Pearson’s correlation coefficient. The term association refers to categorical features; the term correlation refers to features on the interval scale. Only associations with a controlled false discovery rate of less than 10% were selected. All listed p-values were adjusted using the Benjamini-Hochberg procedure [(section 2.1.2)] unless indicated otherwise.” (Garzorz-Stark et al., 2016)

3.1.3 Results

The molecular classifier precisely separates classical cases of eczema and plaque psoriasis

The first cohort was used to build the classifier consisting of the expression of NOS2 and CCL27 using a logistic regression (Fig. 3.1). Evaluating the model on the same data leads to disease probabilities for each patient for both diseases. The cut-off probability for clear prediction was set to 55%. The classifier predicted correct disease status for 87/89 samples in the first cohort. Correct disease status was defined with histology as gold standard. “[...] Test specificity for psoriasis (eczema) was 100% (97.7%), sensitivity was 97.7% (100%)

and the AUC [...] was 0.9929 [on training data]. We tested the molecular classifier for robustness using a 10-fold cross-validation yielding a specificity for psoriasis (eczema) of $100\% \pm 0\%$ ($96\% \pm 8.4\%$), a sensitivity of $96\% \pm 8.4\%$ ($100\% \pm 0\%$) and an AUC of $99\% \pm 3.2\%$." (Garzorz-Stark et al., 2016)

The classifier was extended to a version where autologous healthy skin was not necessary for predicting disease status. In this version, mean expression values in healthy skin for all needed genes was calculated and used in modeling. The performance of the extended classifier was comparable (sensitivity 97.8%, specificity 97.8%).

The molecular classifier correctly identifies subtypes of psoriasis and eczema

Next, the classifier was applied to 31 patients in the second cohort (subtypes of psoriasis and eczema). Of these, 29 patients were classified correctly (Fig. 3.2).

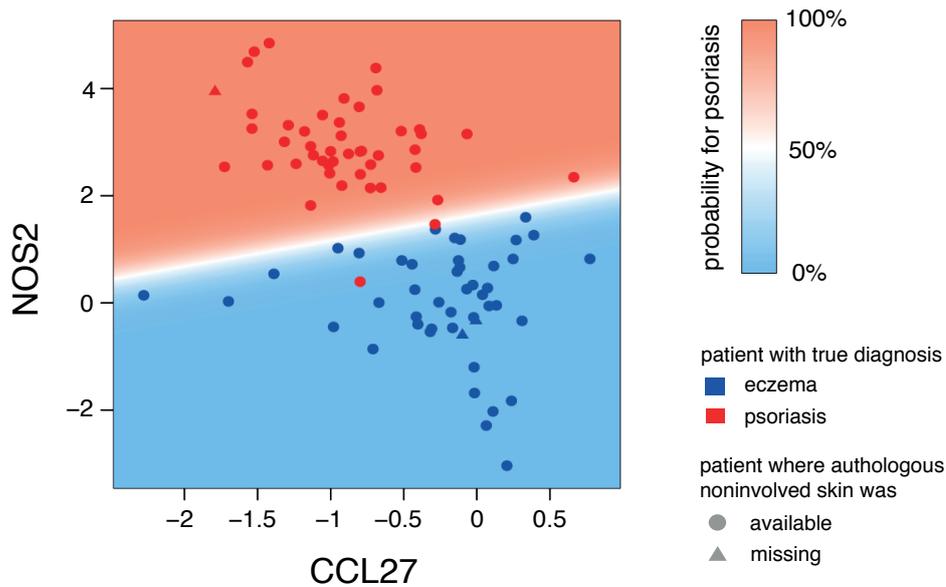


Fig. 3.1 – Gene expression of NOS2 and CCL27 in first cohort. Scatterplot showing the gene expression of NOS2 and CCL27 in all patients from the first cohort (first cohort = training cohort). Dots represent patients, color coded according to disease state (blue = eczema, red = psoriasis). Background colors represent the predicted probabilities of the model. Samples where autologous healthy skin was measured are represented as a circle, while samples with missing corresponding autologous skin are depicted as triangles. Figure adapted from Fig. S1 in Garzorz-Stark et al. (2016).

3.1 MOLECULAR CLASSIFIER FOR PSORIASIS AND ECZEMA

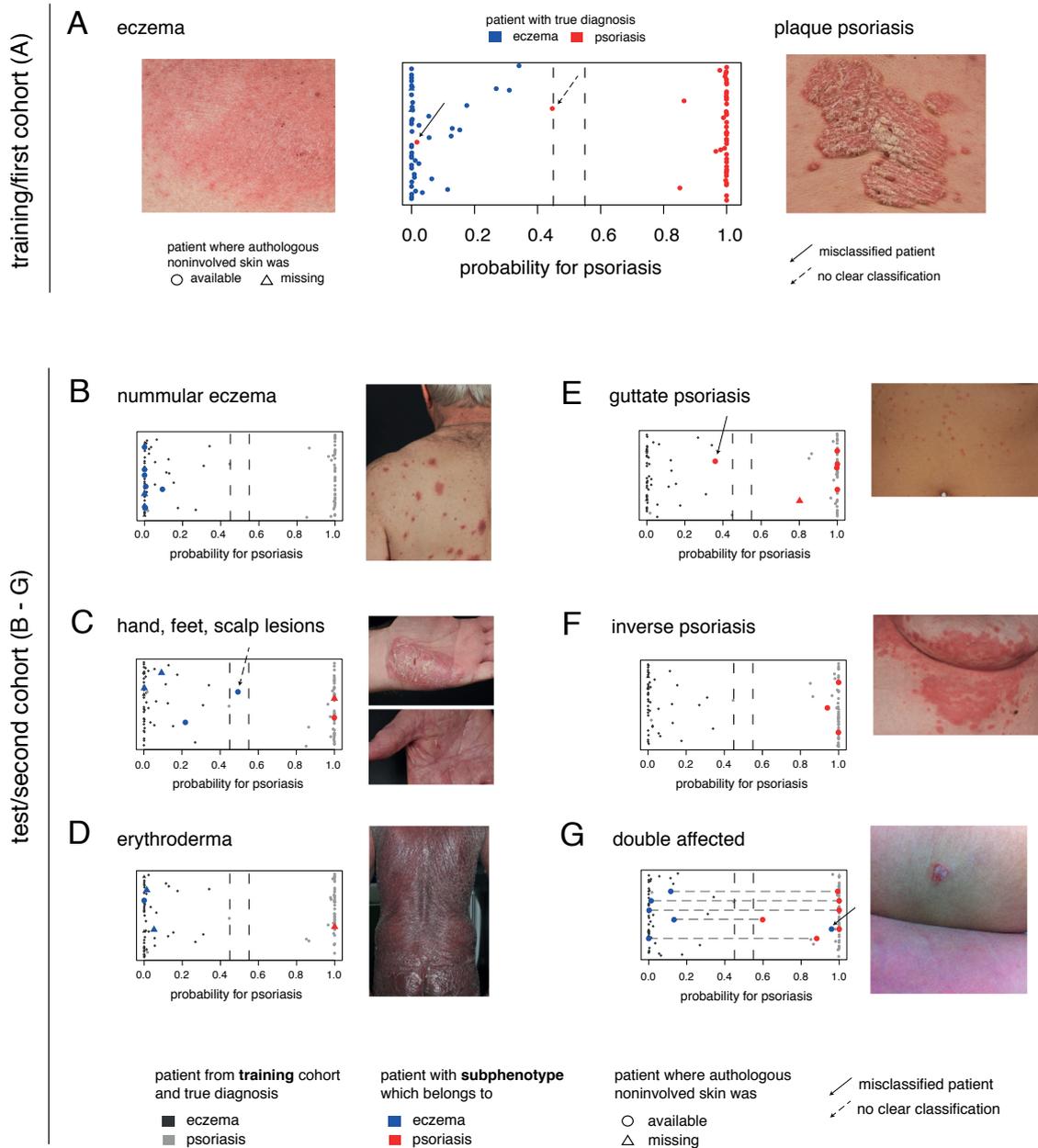


Fig. 3.2 – Molecular classifier for psoriasis and eczema tested on histologically and clinically clear cases of psoriasis and eczema as well as on subphenotypes of both diseases. (A) Classical cases of plaque psoriasis (indicated in red) and eczema (indicated in blue) were analyzed with the classifier (n=89, testing on training data). (B-G) Classifier was applied to subtypes of both diseases (n=31), including nummular eczema (B), hand/feet and scalp lesions (C), erythroderma (D), guttate psoriasis (E), inverse psoriasis (F) and patients suffering from coexisting psoriasis and eczema (G). In (B-G), samples are represented as coloured circles, grey dots in the background of (B-G) represent the clear cases of eczema and psoriasis from (A) for comparison. Samples with missing corresponding autologous skin are depicted as triangles, others as circles. Arrows point to misclassified patients and those with no clear classification (those who lie between 45% and 55%, indicated with vertical dashed lines). Figure adapted from Fig. 1 in Garzorz-Stark et al. (2016).

NOS2 and CCL27 correlate with well-established characteristics of psoriasis and eczema

As the classifier is highly specific and sensitive, both predictors in the model (CCL27 and NOS2) were analyzed for their association with anamnestic, clinical, histological and laboratory attributes which are known for being specific for either psoriasis or eczema (Fig. 3.3 A). Of note, NOS2 expression levels were strongly associated to histologic hallmarks of psoriasis (hypogranulosis, microabscess and dilated dermal capillaries, Fig. 3.3 B). CCL27 showed negative correlation with BMI. Further, CCL27 was associated negatively with dilated dermal capillaries and positively with allergic rhinoconjunctivitis (Fig. 3.3 C).

Molecular classifier proves to be a reliable tool for diagnostic purposes

As a final step, the classifier was applied to the third cohort (10 patients with unclear diagnosis). Here, the ground truth was defined based on positive clinical outcome after disease-specific medication was administered. Psoriasis or eczema assignment of the molecular classifier agreed in all ten cases with the corresponding diagnostic response of the patients.

3.1.4 Discussion

Here, a molecular classifier for diagnosing psoriasis and eczema was established, validated in different cohorts and revealed high specificity and sensitivity. The classifier is based on the expression of only two genes and thus small in number of features compared to similar models. Associating both predictors of the classifier to known hallmarks of both diseases corroborated their validity. Only few patients suffer from a wrong diagnosis but both diseases are very prevalent (2%–4% for psoriasis (Parisi et al., 2013), 2%–10% for eczema in adults (Bieber, 2008)), so the total number of patients this classifier is able to help might be large.

3.1 MOLECULAR CLASSIFIER FOR PSORIASIS AND ECZEMA

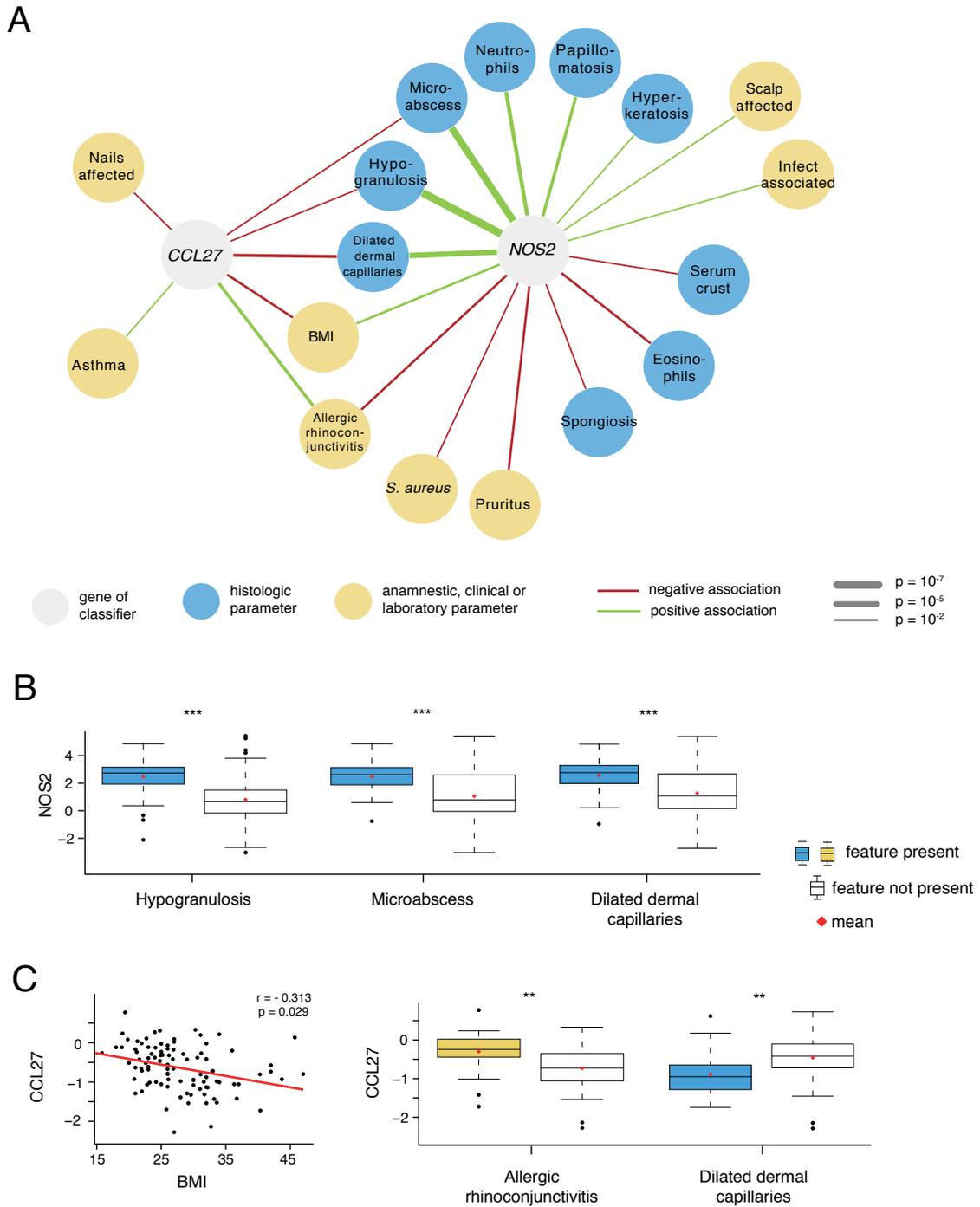


Fig. 3.3 – Correlation of NOS2 and CCL27 with hallmarks of psoriasis and eczema. (A) NOS2 and CCL27 expression levels were associated to anamnestic, clinical and laboratory attributes (indicated in dark blue) as well as histological attributes (indicated in light blue). Positive associations are indicated as green lines, negative associations as red lines. Line thickness is inverse proportional to p-value (thicker lines indicate higher significance). (B) The most significant associations for NOS2. (C) The most significant associations for CCL27. Figure adapted from Fig. 3 in Garzorz-Stark et al. (2016).

3.2 Plasma miRNA classifier for allergic asthma

Similarly to the previous section, this study also aimed to utilize gene expression levels measured by quantitative polymerase chain reaction to classify patients. However, in this study expression of micro RNAs was detected and exploratory analysis was first conducted in mouse models before the classifier was applied to humans. The micro RNAs were also associated to clinical attributes which comprised of routine clinical measurements and medication use. The study was conducted in collaboration with Katrin Milger from the Institute of Lung Biology and Disease at Helmholtz Center Munich and from the Comprehensive Pneumology Center at Ludwig Maximilian University Munich, and Susanne Krauss-Etschmann from the Research Center Borstel. Content, text and data of this section are based on Milger et al. (2017). Copied text passages are indicated as quotations.

In this study, the thesis author was responsible for the processing and analysis of all data gathered in humans, starting with data management and quality control of molecular and clinical data. Part of the analysis was building of the human miRNA based asthma classifier using regularized regression modeling. Further associations to clinical attributes were calculated by her. The thesis author was further responsible for interpreting the computational results and designing and creating all figures presented here. The mouse data was analyzed by Nikola S. Mueller and results are summarized here in order to make the study more understandable. All text in the publication Milger et al. (2017) concerning statistical analysis, interpretation and description of results in the human data were originally written by the thesis author. The statistical methods are described here in more detail than in the publication.

3.2.1 Motivation

Asthma is characterized by chronic inflammation of the airways (section 2.4). Diagnosis can be challenging due to its heterogeneity and different phenotypes (Wenzel, 2012). So a molecular classifier might be beneficial for correct diagnosis and in developing therapies.

“Micro RNAs (miRNAs) are evolutionary conserved, short (20-22 nucleotides long), noncoding RNAs that regulate gene expression by promoting messenger RNA degradation or inhibiting protein translation.” (Milger et al., 2017) MiRNAs are present for example in plasma so they are good candidates for biomarkers. They are stable (Patton et al., 2015) and can be measured with quantitative polymerase chain reaction (see section 2.2.1). MiRNAs have already been used as biomarkers in complex chronic diseases like cancer (Mo et al., 2012).

3.2 PLASMA MIRNA CLASSIFIER FOR ALLERGIC ASTHMA

Also in asthma, deregulated miRNAs have been described in airways in humans (Solberg et al., 2012) and mouse models (Garbacki et al., 2011).

Changes in miRNA expression levels may be small in asthma compared to diseases like cancer (Brase et al., 2010). To overcome this limitation, a translational approach to determine miRNAs which classify between patients with asthma and healthy subjects was applied. It was shown that miRNA targets are conserved between human and mouse (Friedman et al., 2009), so the translational approach contained a first screening step in mouse models of asthma followed by a second step of data collection and analysis in humans.

3.2.2 Methods

This method section focuses on the statistical analysis in the human part of the project, for further information please see Milger et al. (2017) and corresponding supplementary methods.

Mouse study

Two different mouse models of asthma were used: house dust mite (HDM) induced asthma and ovalbumin (OVA) sensitization. Treatment with phosphate buffered saline (PBS) was used as control. The OVA-model consists of a sensitization period and a challenge. Three different approaches were used. First, both sensitization and challenge was done with ovalbumin (OVA/OVA) leading to asthmatic mice. Second, to gain an atopic, only sensitized, but not asthmatic phenotype, the mice were sensitized with ovalbumin but challenged with PBS (OVA/PBS). Third and as a control, mice underwent the sensitization procedure but were only treated with PBS, later they were challenged with OVA (PBS/OVA). Blood was collected for six mice per group and 179 miRNAs were measured with a system based on polymerase chain reactions.

Plasma miRNA expression was compared between asthmatic (OVA/ OVA) and control mice (PBS/OVA). Further comparison were between asthmatic (OVA/ OVA) and sensitized mice (OVA/PBS) and between HDM-treated mice and controls in the HDM model.

Student's t-test was used to identify miRNA ratios that were significantly regulated in both the OVA and the HDM model. To account for multiple testing, the Benjamini-Hochberg

CHAPTER 3 DISEASE CLASSIFIERS

critical value was calculated for each sample and compared to the p-value obtained by t-test.

Patient cohorts

For building and testing the biomarker signature in patients, two independent adult cohorts of patients with allergic asthma were recruited. The training cohort consisted of 20 patients from a local respiratory medicine practice. The test cohort consisted of 26 patients from the asthma clinic at the University Hospital Munich. In both locations also healthy controls were recruited for both cohorts (9 for training cohort, 12 for test cohort). As important covariates, age, sex and smoking status were collected. In all subjects plasma was gathered from peripheral blood. In asthma patients further asthma-specific clinical measurements were obtained. RNA was extracted and levels of miRNA were assessed via qPCR (section 2.2.1) leading to threshold cycle values (Ct value) per miRNA and patient. Only those miRNAs were measured in patient cohorts which were deregulated in both mouse models.

RNA was extracted and levels of miRNA were assessed via qPCR (section 2.2.1) leading to threshold cycle values (Ct value) per miRNA and patient. Only those miRNAs were measured in patient cohorts which were deregulated in both mouse models.

Statistical analysis of human data

For normalization, pairwise differences between Ct values were calculated to obtain miRNA ratios ($\Delta\text{Ct} = \text{Ct}_1 - \text{Ct}_2$). In the training cohort, disease status (asthma versus control) was modeled with a regularized logistic regression (section 2.1.4) using all miRNA ratios as covariates $x_j \in \mathbb{R}^{n \times 1}$ for $j \in 1, \dots, 13$. Variable selection was performed with a generalized linear model via penalized maximum likelihood with lasso regularization (Friedman et al., 2010). The optimal lambda value for lasso regression was calculated using leave-one-out cross validation. The model accuracy was calculated on the test cohort and assessed via AUC (section 2.1.8).

For the five ratios that were selected in the final model, associations with clinical characteristics were analyzed using the data from both cohorts combined (training and test cohort) to overcome limitations in numbers of participants per characteristic. Analysis of variables available in patients with asthma and controls, i.e. age, sex, smoking status, sensitization and allergic rhinitis were performed using data from both groups; the remaining variables were tested among patients with asthma only. Categorical variables were tested for

3.2 PLASMA MIRNA CLASSIFIER FOR ALLERGIC ASTHMA

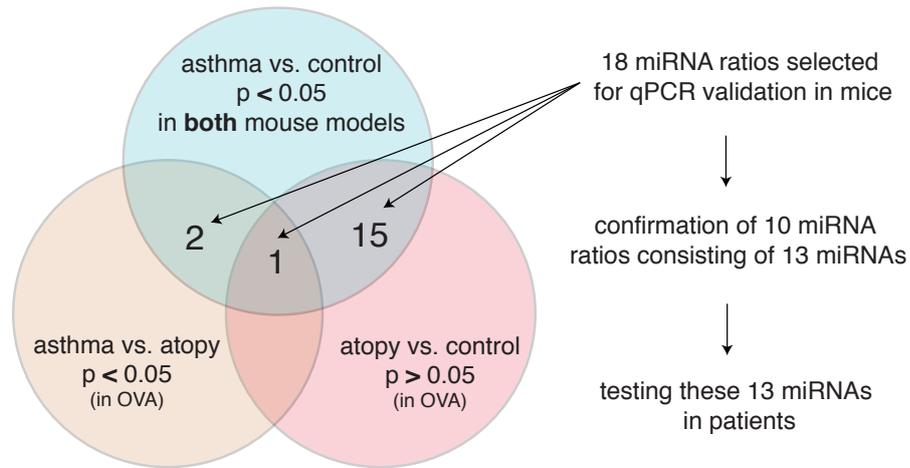


Fig. 3.4 – Selection of candidate miRNA ratios using data from both animal models. Venn diagram illustrating sets of miRNA ratios. Blue circle represents miRNA ratios which are differentially regulated ($p < 0.05$) in both asthma mouse models compared to control. Orange circle represents miRNA ratios which are differentially regulated ($p < 0.05$) between the asthma and the atopy ovalbumin mouse model. Red circle indicates miRNA ratios which are not differentially regulated ($p > 0.05$) between the atopy mouse model and control. Adapted from Fig. 2A in Milger et al. (2017).

differences in miRNA ratios using one-way analysis of variance (section 2.1.3). For interval scale variables, significance was determined using Pearson's correlation coefficient. All p-values were adjusted using the Benjamini Hochberg procedure (section 2.1.2). Only associations with a controlled false discovery rate of less than 10% are shown.

3.2.3 Results

Summary of mouse study

In the ovalbumin mouse model, 32 miRNAs were significantly altered in the asthmatic versus control mice (OVA/OVA vs PBS/OVA). Of these, three miRNAs were also significantly different between asthmatic and atopic mice (OVA/OVA vs OVA/PBS). For the house dust mite model, ten miRNAs were significantly regulated between house dust mite and PBS treated mice. The expression levels of miRNAs had a high between-animal variability. To obtain similar distributions and increase sensitivity, miRNA ratios were calculated by dividing expression levels of each miRNA one-by-one by all the other levels.

Of particular interest were those miRNA ratios which showed significant regulation in the same direction in both asthmatic mouse models when compared to the respective controls,

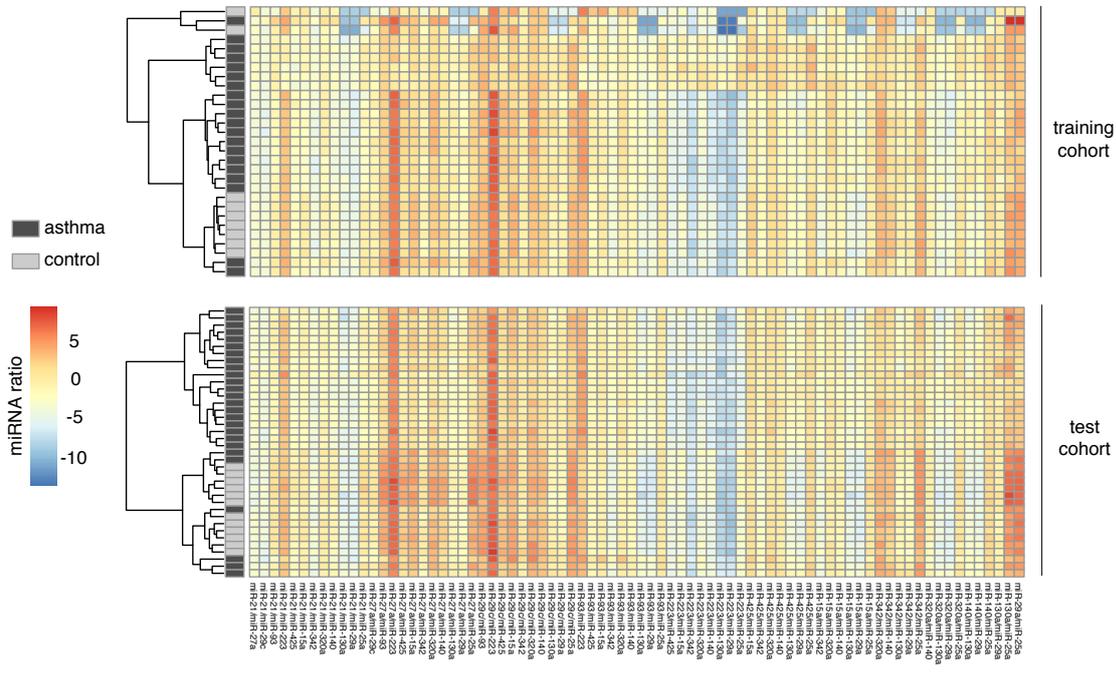


Fig. 3.5 – Heatmap of miRNA ratios in both human cohorts. Clustering was calculated using “ward.D2” method on Euclidean distances separately for each cohort and thus the plot is split in two parts (training (upper panel) and test cohort (lower panel)). In each plot, rows represent subjects and columns miRNA ratios. Figure adapted from Fig. S4 A and Fig. S4 B combined with Fig. 3 A and B in Milger et al. (2017).

but which were not regulated in atopic animals (Fig. 3.4, intersection of blue and red circle). Sixteen miRNA ratios fulfilled this criterion and were picked as first candidates. Two further candidates were identified where miRNA ratios were significantly regulated between asthmatic and atopic mice and between both asthma models and controls (Fig. 3.4, intersection of blue and orange circle). The 18 ratios were validated by quantitative polymerase chain reaction in an independent set of samples. The validation detected the regulation of ten miRNA ratios which included 13 miRNAs. These 13 miRNAs were picked for the next step, further confirmation in human samples.

Building and testing a human miRNA asthma classifier

The 13 miRNAs, identified through the study in mouse models, were measured in both human cohorts and all possible ratios were calculated. Clustering based on all miRNA ratios revealed similarities within patients and controls, respectively (Fig. 3.5).

3.2 PLASMA MIRNA CLASSIFIER FOR ALLERGIC ASTHMA

Next, the question was whether a model comprising a subset of the miRNA ratios successfully predicts asthma disease status from miRNA ratios. Using all data from the training cohort, a regularized logistic regression was calculated. The final model contained five miRNA ratios (Fig. 3.6 B). This model was then tested for its generalizability in the independent test cohort. The resulting receiver operator characteristics curve had an area under the curve of 0.9167 (Fig. 3.6 A).

Univariate differences in the identified biomarker miRNA ratios were identified in both cohorts (Fig. 3.6 C). For all ratios the direction of difference between patients and controls was the same in both cohorts, except for miR-223/miR-425 which was downregulated in controls compared to patients in the training cohort but reversely associated in the test cohort. However, the downregulation of miR-223/miR-42 was not significant in the training cohort ($p = 0.084$, Fig. 3.6 D) but the upregulation was significant in the test cohort ($p = 7.71 \cdot 10^{-4}$).

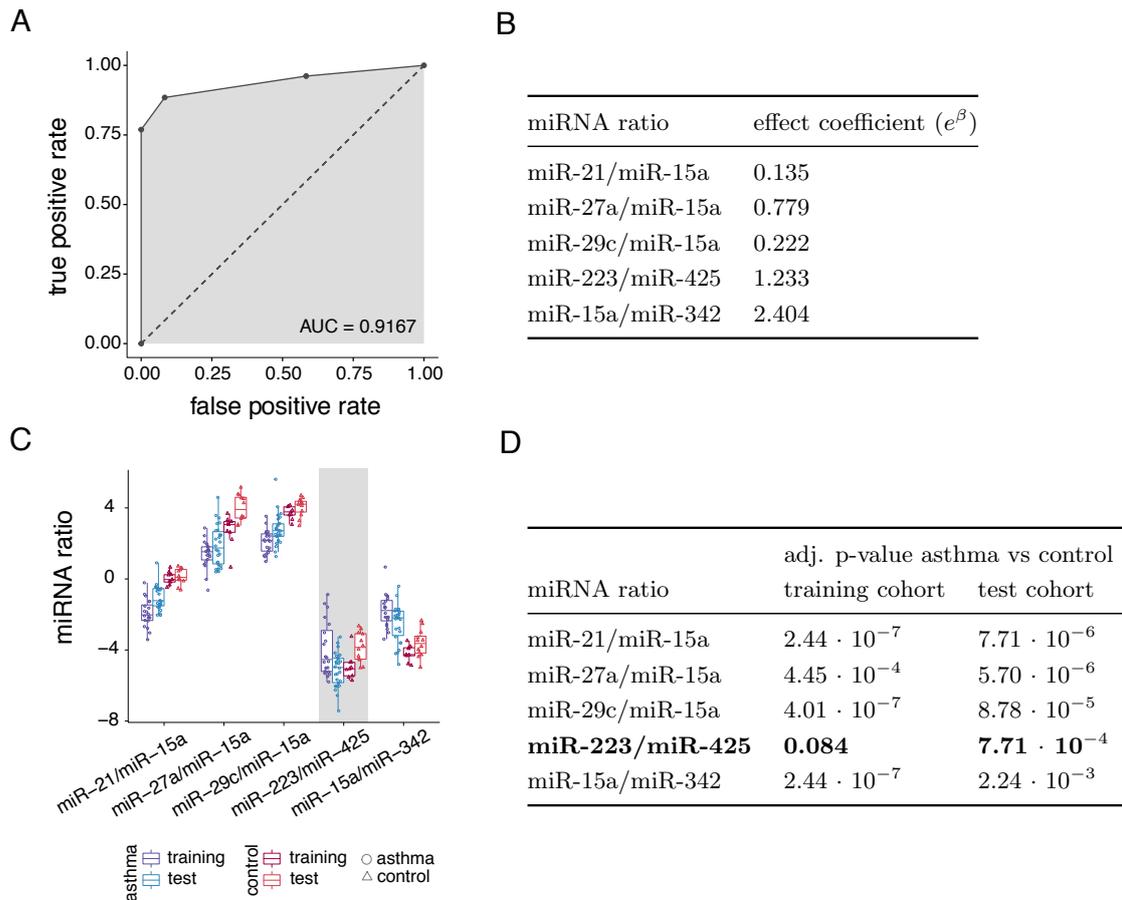


Fig. 3.6 – The miRNA asthma classifier separating asthma patients from controls.

(A) Receiver operator characteristics (ROC) curve for the test cohort depicting false positive rate versus true positive rate as a measure for the accuracy of the model shown in (B). (B) Effect coefficients of classifier (regularized logistic regression). (C) Boxplots comparing miRNA ratios between asthma patients (shades of blue, circles) and controls (shades of red, triangles) in both cohorts (darker color for training and lighter color for test cohort). (D) Adjusted p-values for tests of asthma patients versus controls in (C) for both cohorts separately. Four of the five identified ratios are regulated the same in training and test cohort, except for miR-223/miR-425 which is highlighted in C and D. Figure adapted from Fig. 3 C and D and Table 2 in Milger et al. (2017).

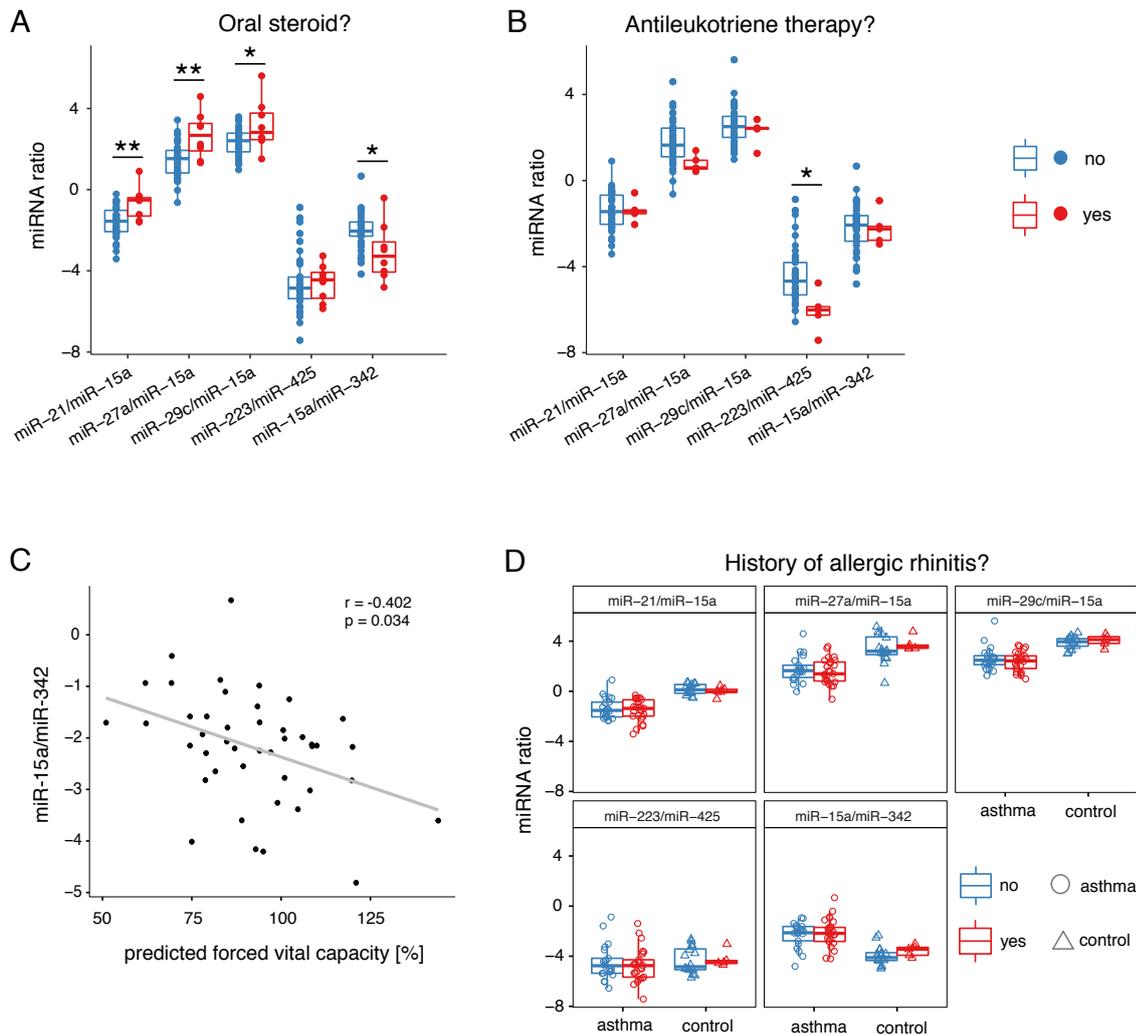


Fig. 3.7 – Associations of biomarker miRNA ratios with phenotypical data. (A) Association of miRNA ratios and current oral corticosteroid therapy in asthma patients of both cohorts. (B) Association of miRNA ratios and current antileukotriene therapy in asthma patients of both cohorts. (C) Correlation of miR-15a/miR-342 with forced vital capacity % predicted in asthma patients of both cohorts. (D) History of allergic rhinitis is not associated to miRNA biomarker ratios in neither asthma patients nor controls in both cohorts. * $p < 0.05$, ** $p < 0.01$. Figure adapted from Fig. 4 in Milger et al. (2017).

Associations of biomarker miRNA ratios with phenotypical data

To determine whether the predictors (five miRNA ratios) in the model (Fig. 3.6 B) associated to known phenotypical data, both training and test cohort were combined to increase power. “[...] Patients taking oral steroids had significantly lower values for miR-15a/miR-342 and significantly higher values for miR-21/miR-15a, miR-27a/miR-15a and miR-29c/miR-15a compared to those not on oral steroids [(Fig. 3.7 A)]. Patients receiving antileukotriene therapy showed decreased values for miR-223/miR-425 [(Fig. 3.7 B)]. [Furthermore,] miR-15a/ miR-342 was inversely correlated to [...] [predicted forced vital capacity (Fig. 3.7 C)].” (Milger et al., 2017) For none of the other variables an association was observed. Since no association was detected for patients with and without a history of allergic rhinitis, specificity of the classifier for asthma is indicated (Fig. 3.7 D).

3.2.4 Discussion

In this study, first miRNAs that are regulated in two mouse models of experimental asthma were identified in plasma. Then, these miRNAs were analyzed in human samples and a regularized regression model was built which differentiates between patients with allergic asthma and healthy individuals. The model was tested in an independent cohort and showed a good predictive power with an area under receiver operator characteristics curve of 0.92. The identified miRNA ratios correlated with clinical characteristics.

A limitation of the presented approach is that miRNAs which are specific for human allergic asthma and not present in the used mouse models might have been overlooked. However, the aim of the study was not completeness, but to determine whether the approach leads to candidate biomarkers linked to human disease.

Chapter 4

Serum proteins as easily accessible surrogates for monitoring human diseases

One of the basic clinical examinations is drawing blood and sending it to the laboratory for analysis of disease-relevant markers. Relevant markers are often proteins in the serum and represent biomarkers (Zhang et al., 2013). Drawing blood is minimally invasive and gives the possibility to gain deeper understanding of pathologic processes within the body (Zhang et al., 2013). Functions of all organs can be traced back to serum which is used in clinics on a daily basis, especially in diagnosis of cancer (Petricoin et al., 2002).

This chapter describes three studies which try to answer the question whether serum proteins can be used for disease monitoring and prognosis to standardize clinical characterization. In all described studies, protein levels in patient serum were measured by collaborators who also collected clinical attributes of the patients. The data was analyzed and integrated by the thesis author in close communication with the collaborators to answer the specific clinical question. Serum proteins were evaluated for their potential use as clinical biomarkers in three applications: disease severity (section 4.1), disease persistence (section 4.2) and post-surgery regeneration (section 4.3).

Content of the first study described in this chapter (section 4.1) was published in 2016 in the *Journal of Allergy and Clinical Immunology* (Krause et al., 2016).

L. Krause, V. Mourantchianian, K. Brockow, F.J. Theis, C.B. Schmidt-Weber, B. Knapp, N.S. Mueller and S. Eyerich: *A computational model to predict severity of atopic eczema from 30 serum proteins*. *Journal of Allergy and Clinical Immunology* (2016).

Together with the collaborators at the Department of Dermatology and Allergy at the University Hospital of Technical University Munich (TUM) and the Center of Allergy and

CHAPTER 4 SERUM PROTEINS AS SURROGATES

Environment (ZAUM) at TUM and Helmholtz Center Munich, the thesis author together with her supervisors aimed at determining a computational model to predict severity of atopic eczema from serum protein levels. Atopic eczema is, among others, characterized by different levels of severity. Measurements of 30 proteins were taken in the serum of atopic eczema patients to evaluate their potential as disease severity biomarkers. The publication led to a response letter from colleagues at the University Medical Center Utrecht where they indicated inconsistencies with previous studies (Thijs et al., 2017b). In fact, some misassignments of patients to measurements were identified and corrected. The corrected analysis was published alongside a reply (Krause et al., 2017). This thesis presents results of the updated, corrected analysis.

L. Krause, N.S. Mueller and S. Eyerich: *Reply to: Multiplex platform technology and bioinformatics are essential for development of biomarkers in atopic dermatitis*. Journal of Allergy and Clinical Immunology (2017).

The second part of the chapter focuses on another aspect of atopic eczema. Atopic eczema is also characterized by the age of onset during infancy and a decreases in prevalence with age. In collaborations with the partners from the Department of Dermatology and Allergy at the University Hospital of TUM, the thesis author together with her supervisors investigated the question whether diseases persistence or remission in children can be explained from protein serum levels and clinical attributes (section 4.2).

V. Baghin*, **L. Krause***, S. Eyerich, K. Eyerich, F.J. Theis, N.S. Mueller, F. Lauffer and N. Garzorz-Stark: *Predicting persistence of atopic eczema in children using serum proteins and clinical data*. - in preparation -

In the third study investigating serum proteins as markers for disease monitoring and prognosis the questions was whether serum proteins can be used to predict the prognosis of patients who undergo major liver resection. In collaboration with the Heidelberg University Hospital and the German Cancer Research Center (DKFZ) who collected the data, the thesis author applied computational methods which were comparable to the first two studies described in this chapter. Serum of patients who underwent major liver resection at Heidelberg University Hospital was analyzed for serum proteins before and on three time points after surgery. Together with clinical attributes and further blood makers, the thesis author investigated the potential of these markers for prognosis prediction. A validation cohort is currently measured and a manuscript is prepared.

K. Dehlke*, **L. Krause***, F.J. Theis, U. Klingmüller, N.S. Mueller and K. Hoffmann: *Prediction of individualized liver regeneration capacity after liver resection based on cytokine and growth factor profiling*. - in preparation -

4.1 Predicting severity of atopic eczema from 30 serum proteins

The study was conducted in collaboration with the Department of Dermatology and Allergy at the Technical University Munich and the laboratory of Stefanie Eyerich at the Center of Allergy and Environment (ZAUM) at the Technical University Munich and Helmholtz Center Munich. The aim of this study was to investigate the potential of serum proteins for predicting severity of atopic eczema patients. Except for age and gender the collaboration partners did not collect any further clinical attributes. Content, text and data of this section are based on Krause et al. (2016) and Krause et al. (2017). Copied text passages are indicated as quotations. All analyses and figures shown in this section result from the correct data which was presented in the online correction and the reply Krause et al. (2017).

In this study, the thesis author was responsible for and performed all data processing and all statistical analyses including handling of missing data, correlation analyses, differential testing and regression modeling. All figures were designed and created by her. The thesis author wrote all text concerning statistical analyses and interpretations for the original manuscript Krause et al. (2016), its correction and the reply Krause et al. (2017).

4.1.1 Motivation

In atopic eczema (AE) disease severity and thus treatment effects are measured using scores like SCORAD (SCORing atopic dermatitis, see section 2.4.1). These scores have subjective components so an objective measure based on a computational model would be desirable. As predictors for these models, serum proteins seem useful because serum is easily accessible.

4.1.2 Methods

Patient cohort

This study includes data collected from 52 patients with atopic eczema and 20 healthy control subjects. In the serum of each subject the concentration levels of 32 serum proteins were measured. Twenty-seven proteins were measured using a cytokine multiplex assay. The remaining five, namely CCL17, CCL22, IL-22, lactatedehydrogenase (LDH) and total IgE, were measured using single plexes and ELISA. The severity of atopic eczema was

CHAPTER 4 SERUM PROTEINS AS SURROGATES

determined in all patients using SCORAD, which evaluates the intensity, extent and subjective signs of the disease.

Statistical analysis

Analysis of log10-transformed measurement values and SCORAD prediction was conducted in R. “Two proteins (IL-2 and IL-15) were not detectable in serum of more than 25% of the patients and controls and were therefore excluded from subsequent analysis.” (Krause et al., 2016) For two patients CCL-17 was above the detection limit and set to the doubled highest standard value. All other measurements were within detection limits. Differences in single variables were detected with a Welch’s 2-sample t-test. Multiple testing correction was performed according to the Benjamini Hochberg procedure also known as false-discovery rate (FDR, see section 2.1.2). Correlation between probands and among serum proteins were calculated using Pearson’s correlation coefficient r on log10 transformed values in a pairwise manner. Subsequently, hierarchical clustering was performed based on the distance measure $(1 - r)/2$ which was introduced in section 2.1.10.

Predictive model for severity of atopic eczema

The predictive model for severity of atopic eczema was learned using a linear regression model. Model outcome $y \in \mathbb{R}^{52 \times 1}$ was SCORAD and log10-transformed protein concentrations were the measured variables $x_j \in \mathbb{R}^{52 \times 1}$ for $j \in 1, \dots, 32$. All (variable) subset regression analysis was performed with the “regsubsets” function from the *leaps* package (Lumley, 2017) in R which optimizes the adjusted residual sum of squares (R_{adj}^2) to select important variables. The R_{adj}^2 was used as a first criterion for the quality of the model (see section 2.1.8). Next to R_{adj}^2 , the generalizability of the model was investigated by leave-one-out cross-validation (see section 2.1.8) and by calculating the mean squared prediction error.

4.1.3 Results

The concentration distribution of all analyzed proteins is depicted in Fig. 4.1. “Significant differences between patients and controls were detected for the levels of CCL17, CCL22, CXCL10, IgE and LDH (FDR < 0.05).” (Krause et al., 2017).

4.1 PREDICTING SEVERITY OF ATOPIC ECZEMA

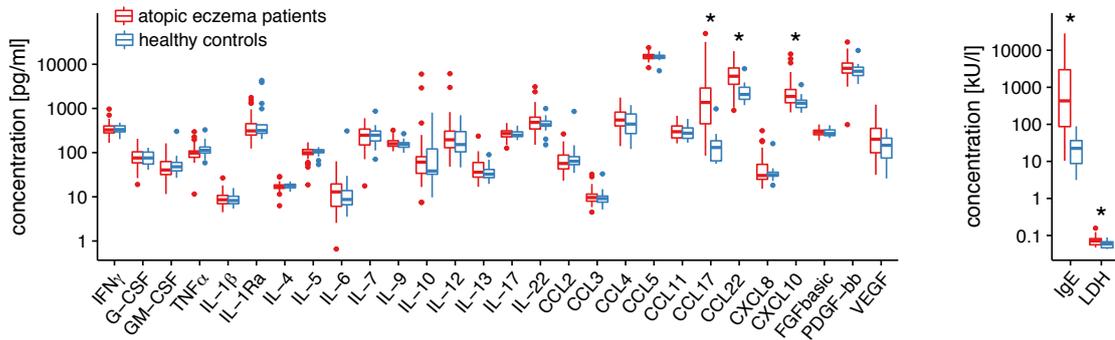


Fig. 4.1 – Boxplots of protein concentrations. Concentrations of 30 serum proteins of patients with atopic eczema ($n = 52$) and controls ($n = 20$) are shown in log scale. Stars indicate $FDR < 0.05$. Figure adapted from Fig. 1 A in Krause et al. (2016) but updated with correct values.

When performing hierarchical clustering using a distance measure based on correlations between probands, “two clusters were identified with one containing patients and controls (cluster 1) and one with patients only (cluster 2) [(Fig. 4.2, A)]. Significant differences between the two clusters were detected for levels of IgE, [...] [CCL17, CCL22 and CCL5, but not for any of the other serum markers or clinical attributes like disease severity or age (5% FDR) (Fig. 4.2, B)]. Here, IgE, CCL17 and CCL22 concentrations are higher in cluster 2.” (Krause et al., 2016)

“To get a first glimpse on potential inter-protein relations, a pairwise correlation analysis and subsequent hierarchical clustering was performed. We used [...] [Ward’s minimal variance method] (Joe and Ward, 1963) to identify [compact, spherical] clusters [(Fig. 4.3)]. In total, [seven clusters] of proteins, containing at least two proteins, were detected in the patient cohort [via cutting the hierarchical clustering tree (tree not shown)].” (Krause et al., 2016)

“Significant differences exist between patients and controls for [five] serum proteins (Fig. 4.1) [and thus] single serum proteins might correlate with SCORAD. [...] Based on Pearson correlation, twelve proteins, namely G-CSF, IL-5, IL-13, IL-22, CCL22, IL-1Ra, CXCL8, IFN- γ , CCL3, IL-1 β , CCL17, and IL-6, significantly correlate with SCORAD (based on $FDR < 5\%$, r range: 0.3 - 0.45). Interestingly, [the main Th2-associated cytokine IL-4 did not show significant correlation with SCORAD.] [...] A reason for this might be the biphasic course of atopic eczema, being dominated by Th2 cytokines in the acute phase and [a combination of Th1 and Th2] cytokines in the chronic phase (Eyerich et al., 2008). Hence, these cytokines may have functional relevance in disease pathology, but they are not suitable as biomarkers for atopic eczema.” (Krause et al., 2016)

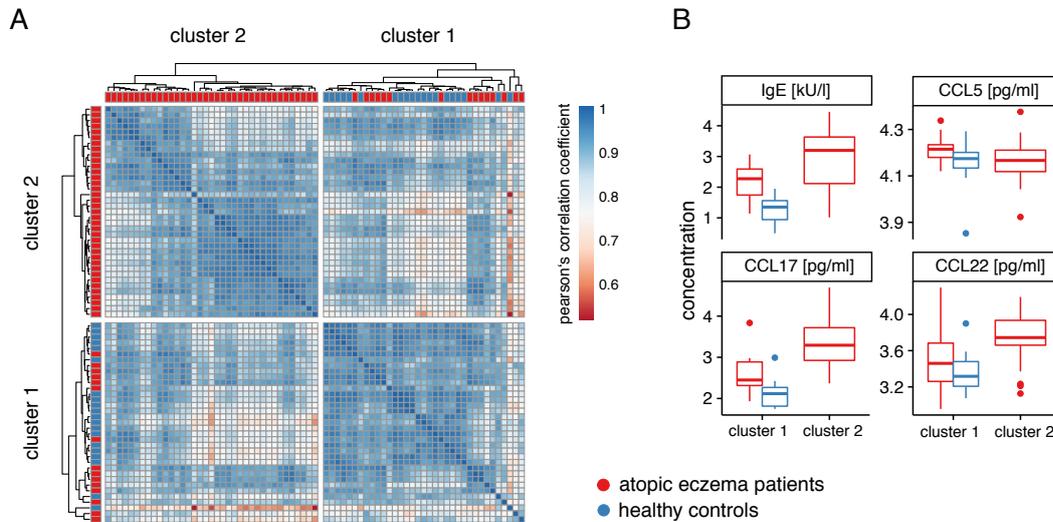


Fig. 4.2 – Clustering of all studied subjects. (A) Hierarchical clustering using a distance measure based on correlation between patients with atopic eczema (red) and healthy controls (blue) indicating the presence of two main clusters. (B) The clusters stratify patients with atopic eczema based on four markers. Figure adapted from Fig. 1 B in Krause et al. (2016) but updated with correct values.

Previously, “CCL17, CCL22 and LDH have been postulated as single protein biomarkers for atopic eczema severity [but were not tested in independent cohorts] (Thijs et al., 2015a).” (Krause et al., 2016) In this cohort, only CCL17 and CCL22 significantly correlated with SCORAD, LDH did not. “CCL17 levels in the healthy control group range from 56 pg/mL to 979 pg/mL, with an average of 169 ± 202 pg/mL (mean \pm SD) in our cohort.” (Krause et al., 2017) “This is in line with observations from other groups that reported high inter-individual differences in serum concentrations of these proteins (Thijs et al., 2015a).” (Krause et al., 2016) “Despite high inter-individual variations in the CCL17 expression levels in patients with atopic eczema (4629 ± 9559 pg/mL), we confirmed an overall correlation of CCL17 with SCORAD (Pearson correlation coefficient of log₁₀-transformed CCL17 is 0.450; adjusted p-value = 0.00084; 95% CI, 0.201-0.643) in our data set. This correlation was in the lower range when compared with the cross-sectional studies of the meta-analysis performed by Thijs et al. (2015a) (meta-analysis of four longitudinal studies, correlation coefficient 0.60; 95% CI, 0.48-0.7).” (Krause et al., 2017).

In order to derive a biomarker signature consisting of more than one serum protein, “a statistical model was used that selects protein combinations to predict SCORAD. The SCORAD outcome was learned using a [...] linear regression model [...] [and variables were selected using] all (parameter) subset regression [which] optimized the adjusted residual

4.1 PREDICTING SEVERITY OF ATOPIC ECZEMA

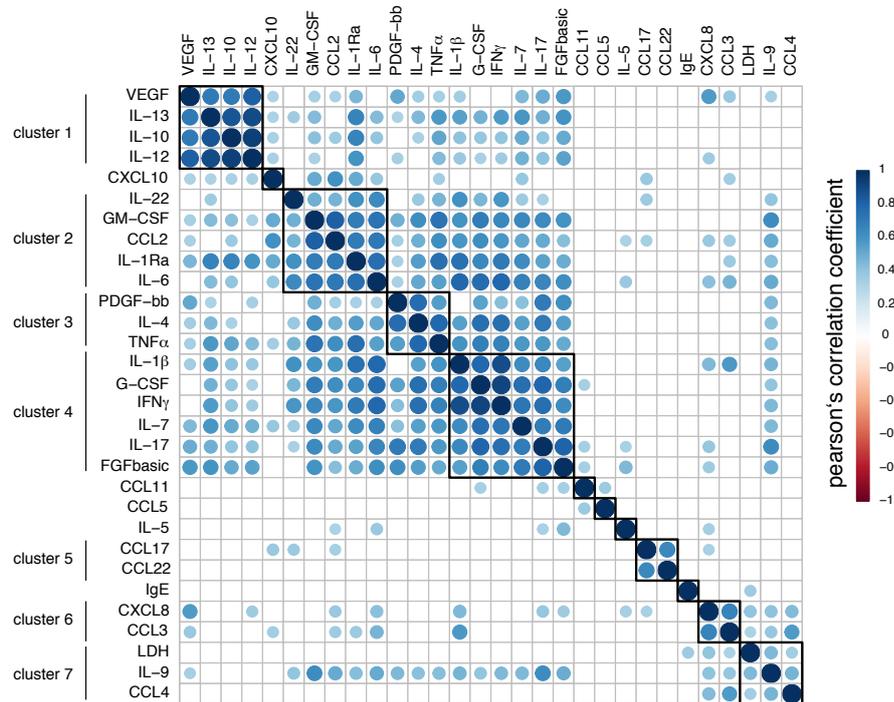


Fig. 4.3 – Hierarchical clustering of serum proteins in the atopic eczema cohort. In total, seven protein clusters were defined, indicated on the left. Only correlations with $FDR < 5\%$ are shown as colored circles, non-significant correlations are left blank. Figure adapted from Fig. 1 C in Krause et al. (2016) but updated with correct values.

sum of squares (R^2). We found that the identified optimal model included [ten] serum proteins [(Tab. 4.1)] [...] [and had an R^2_{adj} of 0.47].” (Krause et al., 2016) The correlation between original and predicted SCORAD was high ($r^2 = 0.759$; 95% CI = 0.612-0.855, Fig. 4.4). “Despite the good overall correlation, the root mean squared prediction error estimated by leave-one-out cross-validation pinpointed a prediction error of 18.9 SCORAD points at patient level.” (Krause et al., 2017)

4.1.4 Discussion

The correlation between original and predicted SCORAD was high and in the same range as in comparable publications (multimarker disease severity model developed by Thijs et al. (2015b) gives correlation of 0.856 (no 95% CI given)). “However, computing the correlation coefficient between predicted and true values cannot assess the goodness of fit of any predictive model. [...] [In this case,] the [...] prediction error was 18.9 SCORAD points at patient level. Thus, even this optimized best-fit prediction model underestimated or overestimated the severity of atopic eczema by 20%.” (Krause et al., 2017) With this

kind of error, the clinical description, even if inter-rater variability is considered, is still superior to the computational model. One reason for non-optimal modeling results might be that those serum proteins which predict severity best are still unknown and need to be determined in further, larger scale studies investigating more than 32 proteins.

covariate	coefficient	p-value
Intercept	216.86	0.0421 *
FGFbasic	-186.25	0.0004 ***
CCL-17	13.11	0.0045 **
CCL-2	-42.96	0.0065 **
IL-6	40.33	0.0070 **
LDH	52.50	0.0324 *
IL-13	32.47	0.0340 *
IL-5	43.41	0.0376 *
IL-4	56.99	0.1123
IL-22	-19.39	0.2070
IL-1 β	32.85	0.2397

Tab. 4.1 – Prediction model for SCORAD. A regression model to predict SCORAD from serum protein concentrations was learned using all parameter subset regression. Optimizing the adjusted residual sum of squares led to this model. Table adapted from Fig. 2 B in Krause et al. (2016).

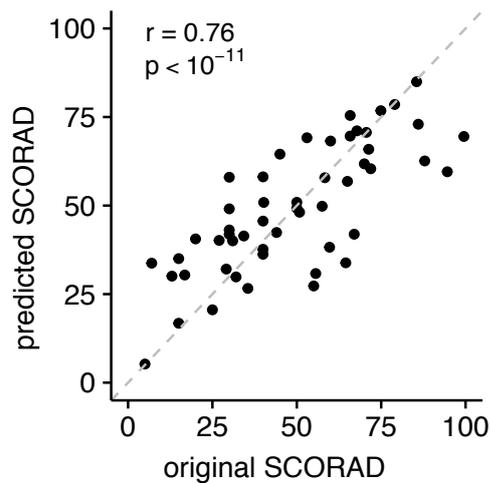


Fig. 4.4 – Comparing predicted and original SCORAD. Correlation is high in our cohort ($r = 0.76$, prediction on training data) but cross-validation prediction error is large (18.9 SCORAD points). Figure adapted from Fig. 2 B in Krause et al. (2016).

4.2 Predicting persistence of atopic eczema in children using serum proteins and clinical attributes

Similar to the study described in the previous section, protein concentrations for this study were measured in serum of patients diagnosed with atopic eczema. In addition, clinical attributes about family history, comorbidities and disease course were collected and jointly analyzed with serum protein concentrations. For this study, the thesis author did not only apply one modeling approach but compared the results of twelve statistical models. The study was conducted in collaboration with Veronika Baghin, Felix Lauffer and Natalie Garzorz-Stark from the Department of Dermatology and Allergy at the University Hospital of Technical University Munich.

In this study, the thesis author was responsible for data processing including imputation of missing values and normalization. Further, she came up with the idea to use several modeling approaches and compare the results. The thesis author performed all statistical analyses herself, including all modeling. All figures were designed and created by her. The results were interpreted together with supervisors and collaborators.

4.2.1 Motivation

Atopic eczema (see section 2.4.1) is usually starting in infancy where a prevalence of 20% in children is described. The prevalence reduces to 2%-10% in adults (Weidinger and Novak, 2016; Garmhausen et al., 2013) which indicates that only a proportion of children suffers from a persistent disease course while others go into remission. Known risk factors for a persistence disease course are high disease score at first appearance, positive family history and elevated total IgE (Von Kobyletzki et al., 2015; Kiiski et al., 2015). Even though risk factors are proposed, it is not possible to predict the disease course at an early age. This prediction would offer patients the possibility to benefit from individualized therapy concepts and combining close medical supervision with drug treatments and allergy prevention. The aim of this study was to investigate the reasons for different disease courses of atopic eczema and to identify predictive markers. The thesis author, together with her supervisors and collaborators investigated whether serum proteins alone or in combination with clinical attributes stratify children for their individual risk of disease persistence.

4.2.2 Methods

Patient cohort

The in-house cohort consisted of 124 children suffering from atopic eczema with age of diagnosis and blood draw between birth and three years of age. Grouping of children into persistence and remission group was based on clinical signs for atopic eczema at the age of seven. If children did not show any signs of atopic eczema for twelve months at the age of seven, they were classified as the remission group. Otherwise, the children had a persistent disease course. Seventy-eight of the children had a persistent disease course (63%), 46 went into remission (37%).

Concentration of total and specific IgE levels and 33 serum proteins were measured with the Luminex platform. Clinical attributes about children's family history, comorbidities, trigger factors for atopic eczema and disease course were collected through questionnaires. Nineteen numerical and 40 categorical clinical attributes were available, so 59 clinical attributes altogether.

Data processing

Total IgE concentrations were log₁₀-transformed. Cytokine measurements were discretized into factors with two levels (cytokine absent or present) if more than 20% were beyond the detection limits. Altogether concentrations for six serum proteins were discretized. For the remaining 27 cytokines, values which were beyond the detection range were imputed using GSimp, a Gibbs sampler based left and right-censored missing value imputation method (Wei et al., 2018).

Clinical data was processed as described in section 2.5. Clinical attributes were carefully analyzed for correlation among them and to the outcome of interest. Highly correlated features were not included in the imputation and modeling procedures. Thus 15 attributes were omitted. Missing clinical data was assumed to be missing at random and for the remaining 44 attributes missing data was imputed with MICE (Raghunathan et al., 2001).

Statistical analysis

Differences between dichotomous categorical variables with regard to numerical variables were tested with Welch's t-test. Fisher's test was used to test association between two

4.2 PREDICTING PERSISTENCE OF ATOPIC ECZEMA

categorical variables. Correlations between two numerical variables were assessed with Pearson’s correlation coefficient. Correlation between mixed variables was calculated with polyserial correlation (Drasgow, 2004; Olsson et al., 1982) which assumes that for the categorical variables there exists an underlying normal variable which was categorized while being observed. The polyserial correlation is inferred based on maximum likelihood estimation. P-values were adjusted for multiple testing with the Benjamini-Hochberg procedure, known as false-discovery rate (FDR).

For visualization purposes and to assess the importance of interconnections between clinical attributes, the thesis author adjusted one variable for another using linear regression. Only measured, numerical variables were adjusted by calculating a linear regression and the adjusted variables were not used for further modeling approaches. The variable which was supposed to be adjusted was set as the outcome y of the regression model. The variable for which that variable should be adjusted for was included as a predictor x in the linear regression. The residuals ϵ of the linear regression are the corrected, adjusted versions of the initial measured variable ($y_{adjusted}$). For example, the concentration of total IgE, which is a measured, numerical variable, was corrected for age of blood draw, which is also numerical. Total IgE levels were set as the outcome y while age of blood draw was represented by $x_{age\ of\ blood\ draw}$. Calculating $\beta_{age\ of\ blood\ draw}$ via least squares approach (see section 2.1.4) allowed to calculate $y_{adjusted}$ by subtracting from the original total IgE levels y the intercept β_0 of the regression model and the product $\beta_{age\ of\ blood\ draw} \cdot x_{age\ of\ blood\ draw}$:

$$y = \beta_0 + \beta \cdot x + \epsilon \quad \rightarrow \quad y_{adjusted} = \epsilon = y - \beta_0 - \beta \cdot x.$$

Modeling of disease persistence

Three approaches to model disease persistence were applied. In all approaches, disease persistence was the response variable and serum proteins as well as clinical information were the covariates. For the first approach, AIC-based stepwise optimization for logistic regression was used, starting either from the intercept-only model (called “start null”) or from the full model including all covariates (called “start full”). The second modeling approach relied on regularized logistic regression. Lasso and elastic net ($\alpha = 0.5$) regularization were applied and deviance as well as AUC were used as loss functions. Third, random forest was utilized and variables were ranked based on variable importance, permutation p-values (Altmann et al., 2010) and Boruta approach (Kursa et al., 2010). Boruta approach is based on permutations and results differed slightly in each run so results of three different runs

CHAPTER 4 SERUM PROTEINS AS SURROGATES

were included in the feature selection results. Models with and without interactions were assessed.

There was no access to an independent test cohort so 10-fold cross validation was performed to estimate model performance. The data set was split into ten subsets, each having an equal number of remitted and persistent atopic eczema cases. Models were trained with the different approaches on nine of ten subsets (=training data) and evaluated using AUC on the training data and on the left-out subset (=test data).

4.2.3 Results

Adjusting for age removed association between total IgE and disease persistence

Atopic eczema is known to show both extrinsic and intrinsic phenotypes (Tokura, 2010). These endotypes are mostly defined on levels of total and specific IgEs where high total IgE and presence of specific IgEs is associated to an extrinsic phenotype (Tokura, 2010). To test the hypothesis whether extrinsic and intrinsic endotypes are associated to persistence and remission of atopic eczema, total IgE levels were measured in serum of the patients. Indeed, children with a persistent disease course showed a trend towards higher levels of total IgE ($p = 0.09$, Fig. 4.5). Notably, total IgE levels are known to increase with age which was confirmed in this cohort (Ott et al. (2010), Fig. 4.5, $r = 0.39$, $p = 10^{-5}$). Moreover, age was associated to disease persistence ($p = 0.00038$, Fig. 4.5). Adjusting total IgE for age abolished the significant association between total IgE and disease persistence ($p = 0.4594$, Fig. 4.5).

The applied concept is similar to partial correlations. Partial correlations describe correlation between two random numerical variables when both are linearly adjusted for all remaining variables (Krumsiek et al., 2011). Representing all pairwise partial correlations which are significantly different from zero in an undirected graph gives a Gaussian graphical model (Schäfer and Strimmer, 2005). Application of partial correlations and Gaussian graphical models to biological data is used for network inference in biological systems (Krumsiek et al., 2011; Benedetti et al., 2017). In this study, no partial correlation following its original definition was calculated since the two variables which were investigated were not both numerical but disease persistence is categorical. Adjusting for the effect of age for the variable total IgE abolished significance of association between total IgE and disease persistence which implies that there exists no partial correlation between total IgE and disease persistence.

4.2 PREDICTING PERSISTENCE OF ATOPIC ECZEMA

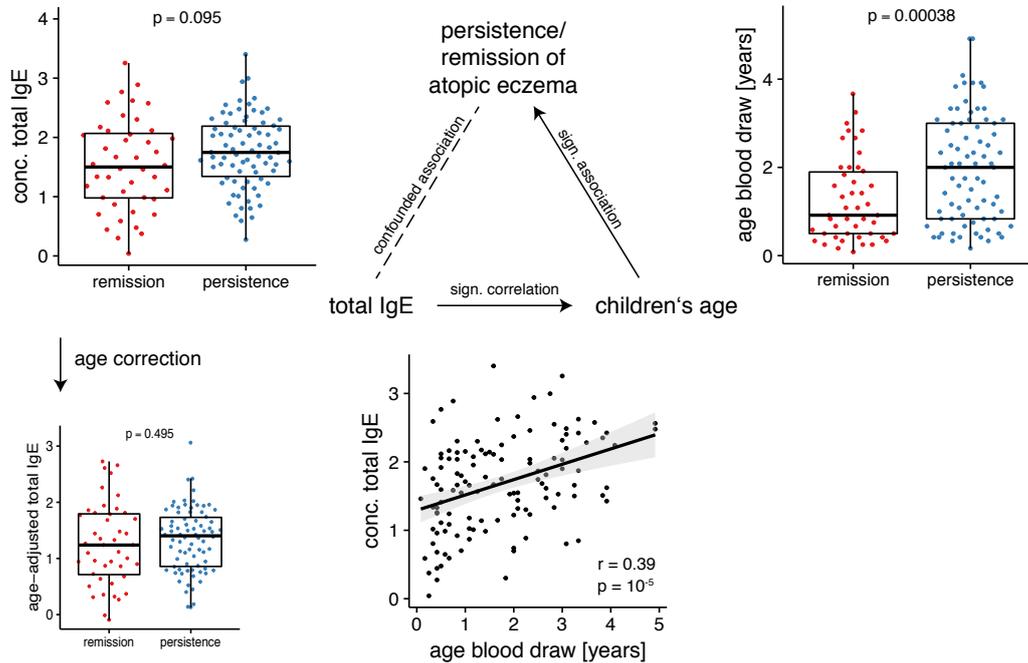


Fig. 4.5 – Age is a confounder for the association between total IgE and disease persistence. Children with persistent atopic eczema showed increased total IgE (upper left) and were older (upper right). Further, total IgE correlated with age of children (lower middle). Adjusting total IgE concentration for age, removed association between IgE and persistent atopic eczema (lower left).

Associations with cytokines and clinical data

None of the 33 measured cytokines showed a significant difference between persistence and remission of atopic eczema, also not after adjusting for age of blood draw. To investigate the relationship between total IgE, age of blood draw and atopic eczema disease course, correlation coefficients between these three variables and all measured cytokines were calculated (Fig. 4.6 A). No significant polyserial correlation was found between persistent atopic eczema and any of the cytokines, but with age of blood draw (first row in Fig. 4.6 A). Age of blood draw was associated to total IgE but also negatively correlated with CCL4, IL-13, CCL2, CCL17, TIMP1 and CCL22 (FDR corrected p-value < 0.1, second row in Fig. 4.6 A). IgE showed positive correlation to FGF-basic, IL-17, IL-4, G-CSF and CCL3 (FDR corrected p-value < 0.1, third row in Fig. 4.6 A).

Factor analysis for mixed data (FAMD) of cytokine measurements did not reveal clusters (Fig. 4.6 B). Within the numerical clinical information only age of asthma start was significantly associated to remission/persistent status of a child (Fig. 4.6 C). Four

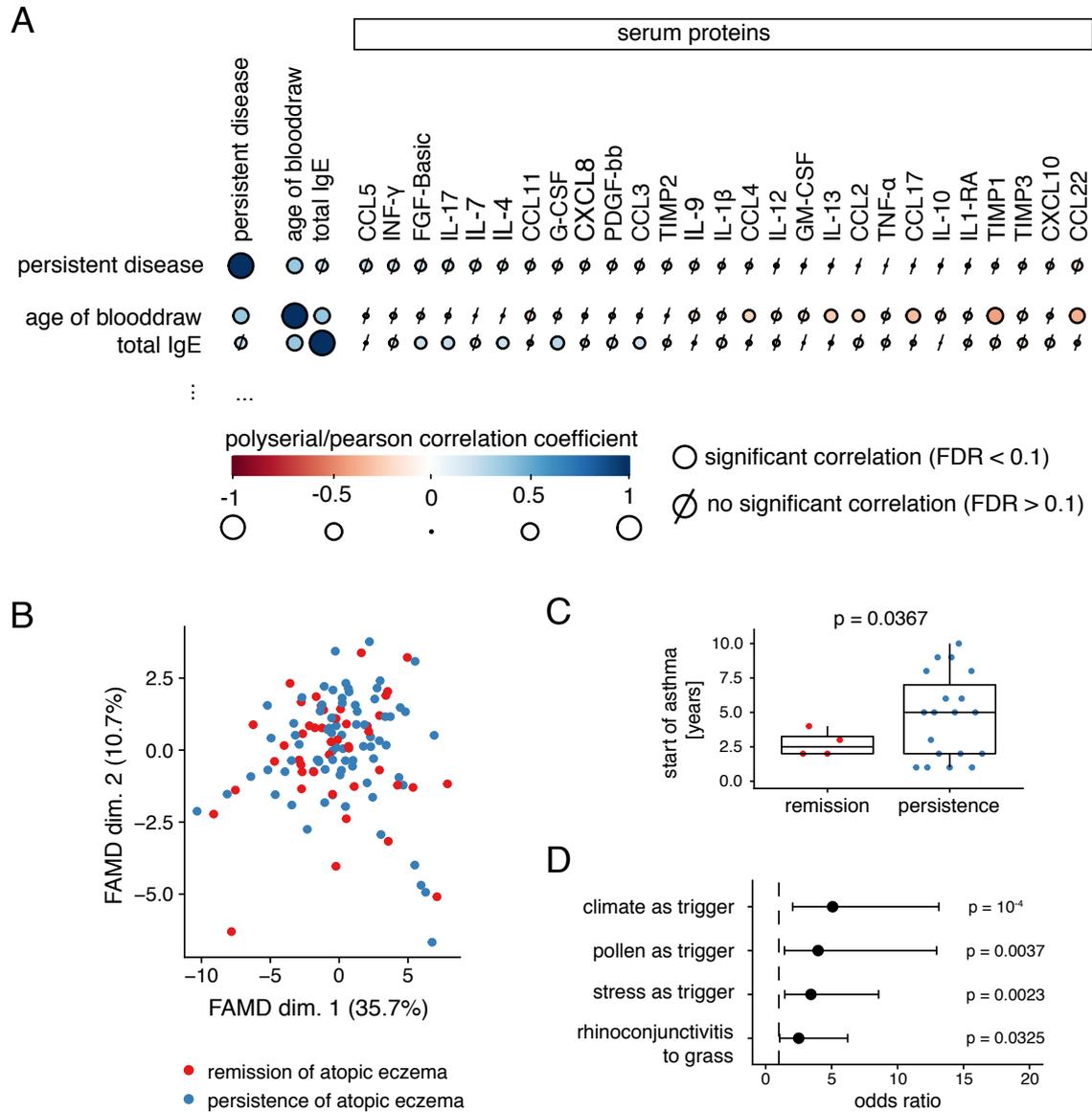


Fig. 4.6 – Association of disease persistence and remission to serum cytokine levels and clinical information. (A) Polyserial (categorical variables) or pearson (numerical variables) correlation coefficients between disease persistence, age of blood draw, total IgE and 33 serum cytokines. Color and circle diameter indicate strength of correlation. Non significant correlations are struck through (FDR cut-off 10%). (B) Factor analysis of mixed data on cytokine data. Colors indicate atopic eczema disease course. (C) Start of asthma was the only significant numerical clinical variables associated to disease persistence (Welch’s 2-sample t-test). (D) Significant hits in Fisher’s test. Odds ratio and 90% confidence intervals are given.

categorical clinical variables were significantly associated to atopic eczema persistence: rhinoconjunctivitis to grass and whether atopic eczema is triggered by a change in climate, pollen or stress (Fig. 4.6 D, p-value cut-off 0.05).

Modeling disease persistence

Since single serum proteins did not correlate with disease persistence (Fig. 4.6 A) a predictive model for disease persistence based on serum proteins as well as clinical information was built. Three modeling approaches were applied to compare results and find shared important variables. Random forest and AIC-based method performed almost perfectly on training data but poorer on test data (Fig. 4.7 A). Regularized-regression models performed equally well on training and test data (Fig. 4.7 A). On average, test AUCs show large variability among the 10-folds reflecting the heterogeneity in the patients (AUC ranges for most methods between 0.5 and 0.8). Regardless of non-perfect model performance, identification of variables commonly selected among methods might hint at variables important for prediction. The methods overlapped in four variables, most of them represent trigger factors for atopic eczema (Fig. 4.7 B).

4.2.4 Discussion

The study aimed at determining predictive markers for persistence of atopic eczema in children. Age was identified as a confounder for the association between total IgE and disease persistence in this data set. No strong association between serum proteins, clinical attributes and disease course was detected in univariate analysis. Since the analysis should not be biased by model choice, results of twelve predictive modeling approaches were compared to investigate the variability and determine common selected features. Combining the commonly selected features, a core set of possibly predictive markers was identified. These markers were solely clinical attributes and not serum proteins. Especially individual trigger factors for disease exacerbation were detected as consistent predictive variables for disease persistence. These findings highlight the importance of environmental and social factors influencing the natural course of atopic eczema in infancy and warrant future considerations of individualized disease management concepts.

The natural course of atopic eczema is diverse and so far unpredictable (Von Kobyletzki et al., 2015). In this study cohort, tendencies for certain attributes were observed, but it was not possible to establish a robust predictive model. This might be based on the heterogeneity of atopic eczema on the one hand and the multimodal disease pathogenesis. Yet, put together in a simple clinical questionnaire, the commonly selected variables present a possibility for testing the models in clinical practice.

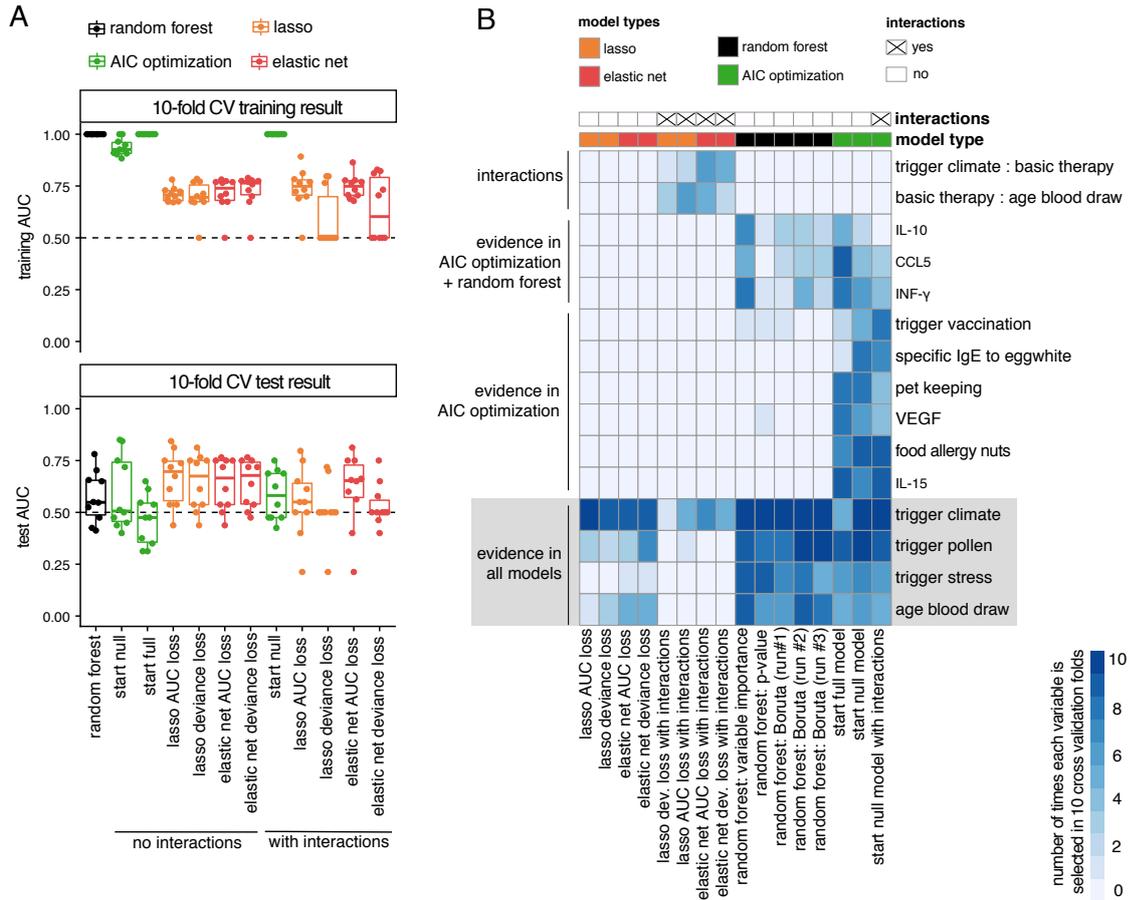


Fig. 4.7 – Result of 10-fold cross validation (CV) of disease persistence modeling. (A) Training and test AUC for twelve applied modeling approaches for ten cross validation folds each. (B) Comparison of important or selected variables for each modeling concept. The darker the color, the more often a variable (name on the right) was selected in ten folds of cross validation by a method (name on the bottom).

4.3 Prediction of individualized liver regeneration capacity after liver resection based on cytokine and growth factor profiling

Comparable to the two other studies described in this chapter, this study also investigated serum proteins for their predictive potential. However, proteins were measured at four time-points over a time span of eight days and not only once. Similar to the study in section 4.2 the collaborators collected clinical attributes and additionally provided the results of blood tests routinely done in their clinic. The study was performed in collaboration with Katrin Hoffmann, Karolin Dehlke and Ursula Klingmüller from the Heidelberg University Hospital and the German Cancer Research Center (DKFZ).

In this study, the thesis author was responsible for all data processing including data normalization, imputation of missing values and batch correction. All statistical analyses including time series clustering, pairwise correlation analyses and regularized regression modeling were designed and performed by the thesis author. Possible networks visualizations were discussed with Nikola S. Mueller and implemented by the thesis author. All figures presented here were designed and created by the thesis author. Results were discussed and interpreted jointly with supervisors and collaborators.

4.3.1 Motivation

If the liver is damaged due to cirrhosis or cancer, up to three-quarters of it can be surgically resected since the liver regenerates (Delis and Dervenis, 2008). Liver regeneration happens in three steps: priming, proliferation and termination (Mohammed and Khokha, 2005). During priming, the liver injury triggers an inflammatory response including the release of cytokines like IL-6 and growth factors (Tao et al., 2017). In proliferation, hepatocytes are induced to divide by mitogens, like the hepatocyte growth factor (HGF) (Furchtgott et al., 2009). The process is stopped by anti-proliferative factors like transforming growth factor beta (TGF- β) to guarantee normal liver function (Tao et al., 2017). Even though the whole liver regeneration process is well studied, the precise timing is not known in humans (Mangnall et al., 2003).

Posthepatectomy liver failure occurs in 15% of patients after a major liver resection and is associated to an increased mortality risk (Rahbari et al., 2011). It is defined on day five post surgery by a functional deterioration which is accompanied by an increase in international normalized ratio and hyperbilirubinemia (Rahbari et al., 2011). The international normalized ratio is calculated by standardizing the prothrombin time which

CHAPTER 4 SERUM PROTEINS AS SURROGATES

describes the clotting tendency of blood. Hyperbilirubinemia describes the presence of too much bilirubin in the blood. Bilirubin is a metabolite in the catabolic pathway, formed in liver and spleen during the degradation of red blood cells and is excreted in bile.

The aim of the study was to use serum samples of patients before and shortly after surgery to predict the risk for liver failure and the regenerative capacity of the remaining liver tissue. Apart from variables measured in serum also clinical attributes were taken into account.

4.3.2 Methods

Patient cohort

Serum from several time points (day -1, day 1, day 3 and day 7 relative to liver resection) for 30 patients who underwent major liver resection were analyzed using Luminex assays. Individual profiles of hepatocyte growth factor (HGF), Interleukin 6 (IL-6), Interleukin 8 (CXCL8), vascular endothelial growth factor (VEGF), transforming growth factor beta (TGF- β), Angiopoietin-2 (APO2), placental growth factor (PLGF) and epidermal growth factor (EGF) were measured (Fig. 4.8). Further, clinical information, blood test results and surgery outcome of patients is available. Clinical information contains data about patients' characteristics like sex, age, BMI and ASA categories which describe the general physical status of patients according to the American Society of Anesthesiologists (ASA). Blood test results were measured during routine hospital examinations in preparation and after liver surgery. They include blood components (e.g. hemoglobin, thrombocytes) and proteins and metabolites known to be associated to processes in kidney (e.g. creatinine), bile (e.g. bilirubin) or liver (e.g. albumin). Bilirubin is a metabolite in the catabolic pathway, formed in liver and spleen during the degradation of red blood cells and excreted in bile. Aspartate transaminase (AST) and alanine transaminase (ALT) are enzymes important in amino acid synthesis and increased in patients with liver disease (Desmet et al., 1994). Quick describes the clotting tendency of blood in the prothrombin time. Standardized prothrombin time is called international normalized ratio (INR). Alpha-fetoprotein (AFP) is a transport protein in fetuses and can be elevated in people with liver cancer.

Outcomes of interest for the analysis are individual surgery outcomes, in particular posthepatectomy liver failure, exitus and Clavien Dindo. Liver failure is once binary coded containing only the information whether liver failure occurred or not (called "categorical"). Once it is coded with a finer grained linear scale describing the necessary interventions once

4.3 PREDICTION OF INDIVIDUALIZED LIVER REGENERATION CAPACITY

the liver has failed (0-3, called “numeric”). Exitus is the information whether the patient died during the hospital stay. Clavien Dindo is a classification system for complications after surgery.

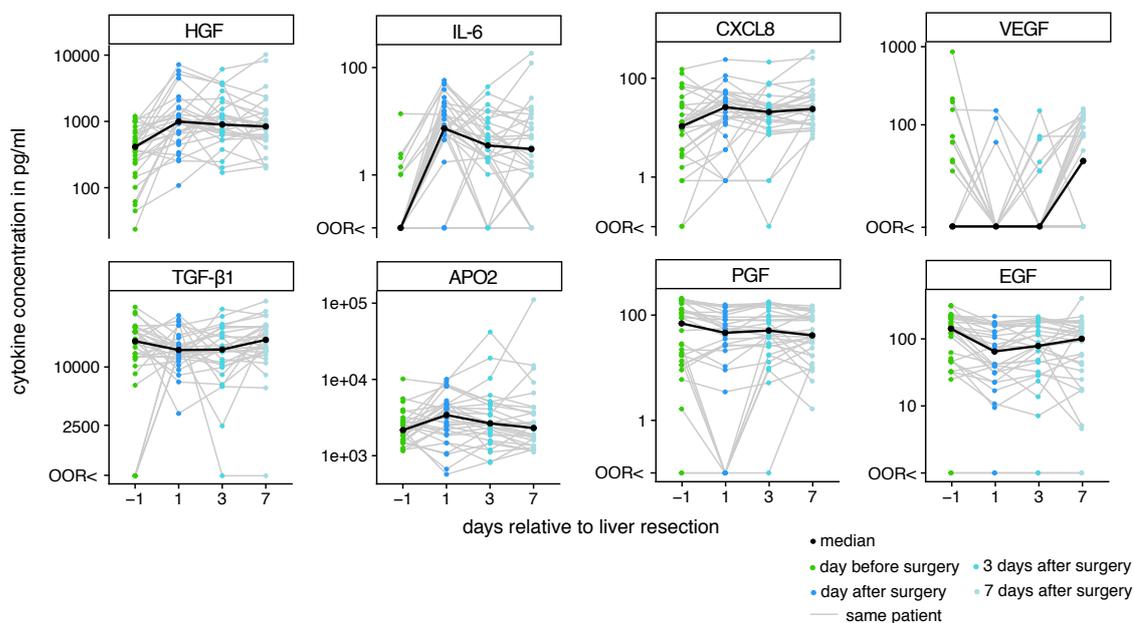


Fig. 4.8 – Time course of cytokine concentrations in serum of patients before and after liver resection. Lines connect measurements of a single patient. OOB< represents values below the detection limit.

Data processing

Patients were collected in three batches and serum protein data included values beyond the detection limit. Serum proteins with more than 20% of values beyond detection limit were binarized as described in section 2.3.4 for modeling approaches (IL-6 and VEGF). For the remaining proteins, first missing serum protein values were imputed with GSimp (Wei et al., 2018) and then log₁₀-transformed. Next, the three batches, which corresponded to rounds of patient recruitment, were corrected for in serum protein data using ComBat (details in section 2.2.8 and 2.3.5, (Johnson et al., 2007; Leek et al., 2012)). The batch corrected and imputed data was inspected with factor analysis for mixed data. All clinical attributes were complete for all patients.

Statistical analysis

Dissimilarities between time-series were calculated using the *TSchust* R-package (Montero et al., 2014) with “CORT” method (Chouakria and Nagabhushan, 2007). The “CORT” method combines pairwise temporal correlation between two observations and raw value behaviors through overall proximity of observations (details in section 2.1.10). Complete hierarchical clustering was performed on these dissimilarities (see also section 2.1.10) and grouping in two to six clusters was assessed for the association to clinical outcome of patients. Associations to clinical outcome were tested with Kruskal Wallis test and χ^2 -test.

Pairwise analysis between all gathered variables was performed (see section 2.1.3). Variables which were assessed at particular time points relative to liver resection were only compared to variables from the same time point. Numerical variables were compared with Spearman’s rank correlation (r_s), numerical and factor variables were compared with Kruskal Wallis test and two factor variables with Fisher’s test. All comparisons were commonly adjusted for multiple comparisons using the Benjamini-Hochberg procedure. An association network (see also section 2.1.3) was drawn for measured variables where nodes represent variables which are connected by an edge if the pairwise association is significant.

Outcomes of interests were modeled using a generalized linear model via penalized maximum likelihood with the lasso penalty and leave-one-out cross-validation for the hyperparameter λ (Friedman et al., 2010). The response family is binomial for categorical liver failure and exitus and Gaussian for Clavien Dindo and degree of liver failure. Calculations were performed in R with the *glmnet* package.

4.3.3 Results

Time-series clustering

To find subpopulations with similar temporal behavior in the data set, dissimilarities between time-series were computed and cluster analysis was performed. A significant association between EGF time series clustering into six clusters and exitus outcome was detected (χ^2 -test, $p = 0.027$, Fig. 4.9). Cluster four is particularly overrepresented and shows a characteristic time course of EGF with a substantial drop on day one after surgery, an increase on day three and the lowest values of EGF on day seven. Only one of four patients clustered into that group did survive the liver resection in the long term. Further significant associations are listed in Table 4.2.

4.3 PREDICTION OF INDIVIDUALIZED LIVER REGENERATION CAPACITY

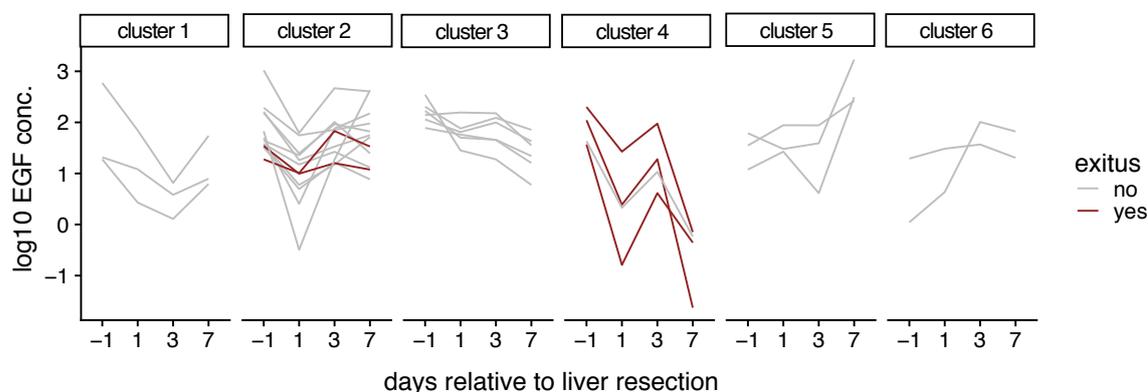


Fig. 4.9 – Time course of EGF grouped into six clusters and association to exitus. Lines connect measurements of a single patient. Red lines highlight patients who did not survive the liver resection in the long term.

Tab. 4.2 – Significant hits in time series clustering. Clusters were associated to clinical outcomes of interest (liver failure and its degree, exitus, Clavien Dindo).

cytokine	# of clusters	clinical outcome	p-value	test
HGF	3, 5, 6	exitus	0.031, 0.029, 0.002	χ^2 -test
HGF	6	degree of liver failure	0.023	Kruskal Wallis test
PLGF	5, 6	exitus	0.011, 0.012	χ^2 -test
PLGF	5	degree of liver failure	0.038	Kruskal Wallis test
PLGF	6	Clavien Dindo	0.045	Kruskal Wallis test
EGF	6	exitus	0.027	χ^2 -test

Association between all variables revealed connection between all levels

To further understand the interconnection between measurements, pairwise analysis between all gathered variables was performed. All associations with $FDR < 10\%$ were visualized as a network (Fig. 4.10). Major clinical outcomes like liver failure, exitus, stay in intensive plus intermediate care and Clavien Dindo are tightly connected. They are linked to different cytokines and blood test results: IL-6 (day +7) is associated to both duration of intensive plus intermediate care and values of the complication classification system Clavien Dindo. IL-6 itself is strongly connected with CXCL8 and HGF on several time points. Total bilirubin is a major hub in the network, connecting blood test results like creatinine, AFP and Quick with clinical outcomes (exitus, liver failure and intensive plus intermediate care).

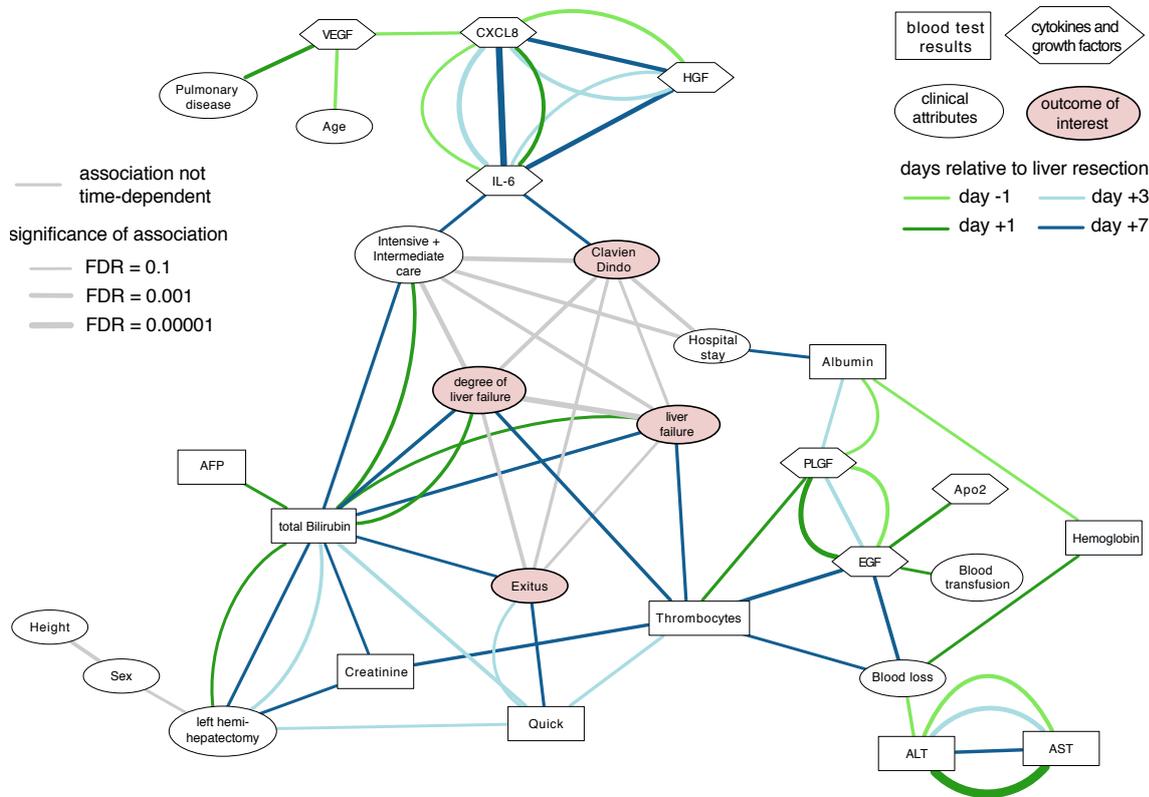


Fig. 4.10 – Association network revealed interconnections between variables. Nodes represent measured variables: cytokines and growth factors (hexagonal shape), clinical attributes (round shape) and standard blood test results (square shape). Two nodes are connected if the appropriate statistical test shows significant association (FDR < 10%). Edge width is inversely proportional to level of significance. Colors indicate whether association is not time-dependent (gray), detected for day -1 (green), day +1 to +7 (shades of blue) relative to liver resection.

Clinical outcomes can be predicted

Some variables already showed univariate association to clinical outcomes of interest. Next, a combination of variables measured before or shortly after surgery was investigated for whether it was able to successfully model clinical outcome. First, only individual profiles of HGF, IGF, VEGF, TGF β 1, IL-6, and CXCL8 were used to predict clinical outcomes. Only measurements of day seven post liver resection predicted clinical outcomes degree of liver failure and Clavien Dindo (Fig. 4.11 A, B). In both models, APO2 was positively associated with the outcome whereas PLGF was negatively associated. Meaning higher APO2 values lead to higher degrees of liver failure and more complications for the patient. However, higher PLGF lead to lower values meaning less complications for the patient.

4.3 PREDICTION OF INDIVIDUALIZED LIVER REGENERATION CAPACITY

Since the aim was to predict clinical outcome before or shortly after surgery next all available clinical information and blood test results were included as covariates in the model. For degree of liver failure, measurements on day three post surgery of total bilirubin and INR together (Fig. 4.11 C, $r_s = 0.526$) model degree of liver failure equally well as four cytokines measured on day seven (Fig. 4.11 A, $r_s = 0.522$). Surprisingly, for Clavien Dindo being modeled by just ASA (Fig. 4.11 D), which is a physical status classification system before surgery, a similar association between original and predicted values was observed ($r_s = 0.526$) than a model including two cytokine concentrations on day seven (Fig. 4.11 B, $r_s = 0.697$).

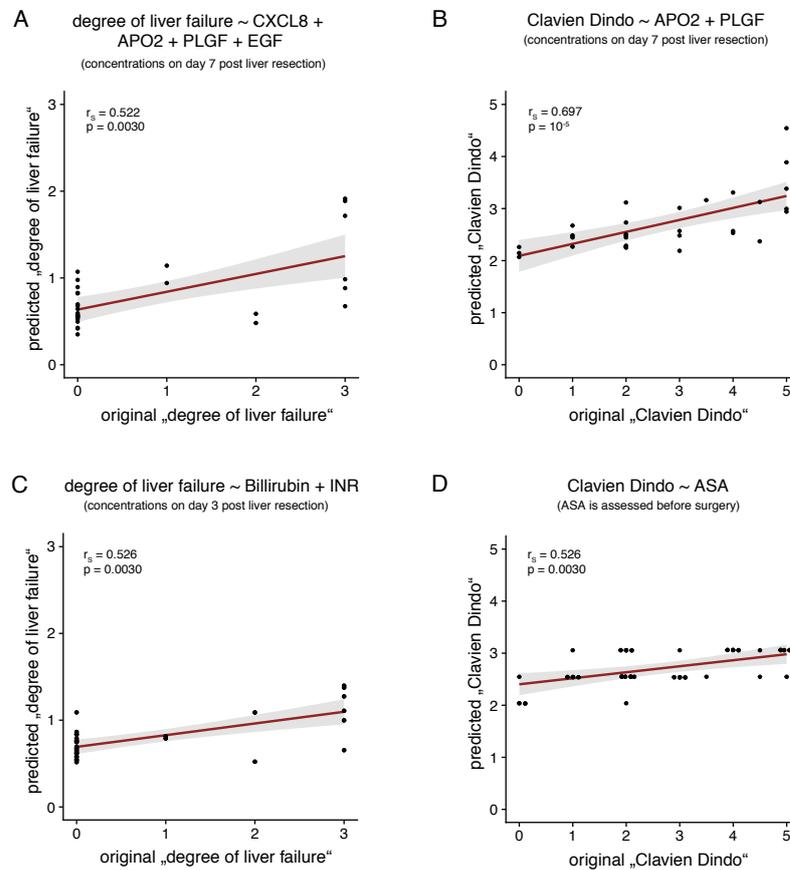


Fig. 4.11 – Regularized regression models predicting clinical outcomes. (A) Degree of liver failure was modeled with concentrations of CXCL8, APO2, PLGF and EGF from day seven post liver resection. (B) Clavien Dindo was modeled with APO2 and PLGF also from day seven. Including also clinical variables led to new models for degree of liver failure (C) and Clavien Dindo (D). Original and predicted values were compared and Spearman's correlation coefficients r_s with associated p-values are given.

4.3.4 Discussion

Time course analysis showed an increase in inflammatory cytokine IL-6 on day one after surgery (Fig. 4.8) as was expected during priming phase after liver resection (Tao et al., 2017). Also the level of hepatocyte growth factor (HGF) in the serum increased post surgery. It stayed constant until day seven, probably inducing hepatocytes to proliferate (Furchtgott et al., 2009) and thus regenerate the liver. Patterns in cytokine and growth factor profiles were detected which were associated to clinical outcomes after surgery (Fig. 4.9). Especially, one time course pattern of epidermal growth factor (EGF) was associated to in-hospital death of the patient.

Pairwise associations between cytokines, growth factors, clinical attributes and blood test results revealed a strong interconnection within and between these layers of information (Fig. 4.10). The level of bilirubin in the blood was strongly associated to degree of liver failure. This association was mainly due to the definition of posthepatectomy liver failure which is based on factors including bilirubin measured on day five post surgery (Rahbari et al., 2011).

Modeling clinical outcomes using cytokines and growth factors did not reveal a strong predictive power of markers measured before or shortly after surgery but only of markers measured on day seven post surgery. Combining cytokine and growth factor profiles with blood test results and clinical information with regularized models did not lead to combined models but revealed different angles on predictive models for individual regeneration capacity of the liver tissue.

Potentially novel connections between clinical attributes and blood markers detected in the network (Fig. 4.10) are currently tested in an independent cohort. Models predicting clinical outcomes are also currently tested and evaluated for their clinical potential.

Chapter 5

Adjusting patient-bias in microarray data using linear mixed effects models

Full skin biopsies are a perfect tool to study the current disease state in inflammatory skin diseases. Routinely, they are taken from patients to fully characterize their disease by histological analysis. Skin biopsies consist of several cell types, mainly skin epithelial cells, the so-called keratinocytes, and immune cells. When performing whole genome gene expression analysis of full skin biopsies, total mRNA of all cell types is combined and measured. The study of Quaranta et al. (2014b), which laid the foundation for the psoriasis/eczema classifier from section 3.1, already showed that gene expression rather clusters by patient than by disease in principal component analysis. In this thesis this effect, which is based on inter-individual variability due to everybody's uniqueness with regard to his or her gene expression patterns, is called patient-bias.

Patient-bias is related to batch effects because both cloud the underlying, biological effect and potentially lead to incorrect conclusions drawn from data if not treated appropriately in the analysis. Batch effects arise due to technically different handling of measurements, e.g. through varying laboratory conditions, different reagents or technicians (Leek et al., 2010). In statistical analyses known batch effects can be corrected for and also unknown sources of variation can be identified and corrected using approaches like those described in Leek et al. (2012).

To overcome the problem of patient-bias in the initial study, lesional and healthy expression values were subtracted per individual and gene to form \log_2 fold changes (Quaranta et al., 2014b). Here, a different approach to handling patient-bias in microarray data of skin biopsies is presented: linear mixed effects models. This method was already successfully applied in two studies which were both published in 2018 and are described in this chapter. Both studies were in collaboration with the Department of Dermatology and Allergy at

CHAPTER 5 ADJUSTING PATIENT-BIAS

the University Hospital of Technical University Munich. This chapter focuses on the gene expression analysis in both studies and briefly summarizes the remaining results which are discussed in detail in the published papers (Lauffer et al., 2018; Garzorz-Stark et al., 2018).

In the first study (section 5.1) gene expression data from patients suffering from either lichen planus or lupus erythematosus, two interface dermatitis diseases, was analyzed. Using the mixed effects modeling a shared type I immune response signature in both diseases was detected which was validated experimentally (Lauffer et al., 2018).

F. Lauffer, M. Jargosch, **L. Krause**, N. Garzorz-Stark, R. Franz, S. Roenneberg, A. Böhner, N.S. Mueller, F.J. Theis, C.B. Schmidt-Weber, T. Biedermann, S. Eyerich and K. Eyerich: *Type I immune response induces keratinocyte necroptosis and is associated with interface dermatitis*. The Journal of Investigative Dermatology (2018).

In the second study (section 5.2) the participants were subjected to topical application of a toll-like receptor 7/8 agonist, called imiquimod, and histologically monitored. Two to five whole genome gene expression microarrays per study participant were available. The thesis author was able to show that the transcriptional profile was closest to the one of acute contact dermatitis compared to other inflammatory skin diseases. Further, the collaborators showed that plasmacytoid dendritic cells act as primary sensors and IL-23 plays a major role in imiquimod-induced inflammation (Garzorz-Stark et al., 2018).

N. Garzorz-Stark*, F. Lauffer*, **L. Krause**, J. Thomas, A. Atenhan, R. Franz, S. Roenneberg, A. Böhner, M. Jargosch, R. Batra, N.S. Mueller, S. Haak, C. Groß, O. Groß, C. Traidl-Hoffmann, F.J. Theis, C.B. Schmidt-Weber, T. Biedermann, S. Eyerich and K. Eyerich: *Toll-like receptor 7/8 agonists stimulate plasmacytoid dendritic cells to initiate TH17- deviated acute contact dermatitis in human subjects*. Journal of Allergy and Clinical Immunology (2018).

5.1 Application in interface dermatitis

First, linear mixed effects models were applied to two interface dermatitis diseases to adjust for patient-bias in the microarray data. The thesis author calculated a shared interface dermatitis gene expression signature from patient data which she compared to *in vitro* stimulated cell line gene expression. The study was conducted in collaboration with Felix Lauffer and Kilian Eyerich from the Department of Dermatology and Allergy at the University Hospital of Technical University Munich who acquired all data analyzed for this section. Content, text and data of this section are based on Lauffer et al. (2018). Copied text passages are indicated as quotations.

In this study the thesis author was responsible for all data analysis of clinical and whole-genome gene expression data. In detail, she processed clinical data and performed correlation analysis of histological attributes of mixed data types to determine objective criteria mostly representing subjective criteria. For the whole-genome gene expression analysis, she analyzed microarray data by performing the whole workflow: data processing, differential expression analysis, pathways analysis using model based gene set analysis and visualization of the results. For differential gene expression analysis the thesis author came up with the idea of using and implemented the usage of linear mixed effect models for microarray data. In particular, she designed the analysis so that the specific medical question could be answered by implementing two models and comparing results. She further integrated *in vitro* gene expression measurements. All figures presented here were designed and created by the thesis author in discussion with her supervisors. The thesis author wrote the original text about analysis and interpretation of computation results for the publication Lauffer et al. (2018).

5.1.1 Motivation

“An increased understanding of the underlying immune mechanisms in inflammatory skin diseases led to the development of specific therapeutic compounds [over the last years] (Noda et al., 2015).” (Lauffer et al., 2018) Interestingly, most of the compounds can be effectively used in several skin diseases that share a similar pathogenesis. Interface dermatitis is a type of skin inflammation present in several inflammatory and autoimmune skin diseases, for example in lichen planus and lupus erythematosus (see also overview of studied diseases in section 2.4). The underlying mechanism of interface dermatitis is insufficiently understood and there is a clinical need for new therapies to treat these kinds

CHAPTER 5 ADJUSTING PATIENT-BIAS

of skin diseases. Deeper understanding of the pathogenesis could open all interface diseases to treatment with new specific therapeutic compounds.

“Aim of this study was to investigate the underlying molecular mechanism of interface dermatitis in a disease independent manner focusing on the overlap of lichen planus and lupus erythematosus in terms of histological architecture, genetic regulations and cellular immune response.” (Lauffer et al., 2018)

5.1.2 Methods

Patient cohort and in vitro experiments

For this study, 25 patients suffering from an interface disease were recruited (lichen planus $n = 14$, lupus erythematosus $n = 11$). Skin biopsies were taken and analyzed using histology. In histology 24 criteria, subjective and objective ones, were assessed to clarify the diagnosis of interface dermatitis. Computational clustering of all histological criteria identified the objective criteria “number of dyskeratotic epidermal cells” to best correlate with the subjective criteria “rating of interface dermatitis” (Fig. 5.1). As a conclusion, in the further analysis only those patients were analyzed which had a least one dyskeratotic epidermal cell in the histological analysis (true for $n = 11$ lichen planus and $n = 5$ lupus erythematosus samples). For comparisons, gene expression of psoriasis patients was included ($n = 16$) which were already published in (Quaranta et al., 2014b). Gene expression of two skin samples per patient were measured, namely lesional and autologous noninvolved skin.

For further understanding genetic regulation in keratinocytes, primary human keratinocytes were stimulated *in vitro* with different cytokines, mimicking different types of immune responses and gene expression was measured (five conditions each $n = 3$).

For all samples, whole-genome gene expression was determined with Agilent microarrays.

Modeling of gene expression to determine differently regulated genes

Gene expression microarrays of human skin biopsies and keratinocytes were preprocessed together and analyzed in parallel following the procedure explained in detail in section 2.2.3. In short, after quality control microarrays were background corrected and normalized. 108 probes were removed because the detected fluorescence was not 10% brighter than the 95% quantile of negative control probes. Within-array replicated probes were averaged.

5.1 APPLICATION IN INTERFACE DERMATITIS

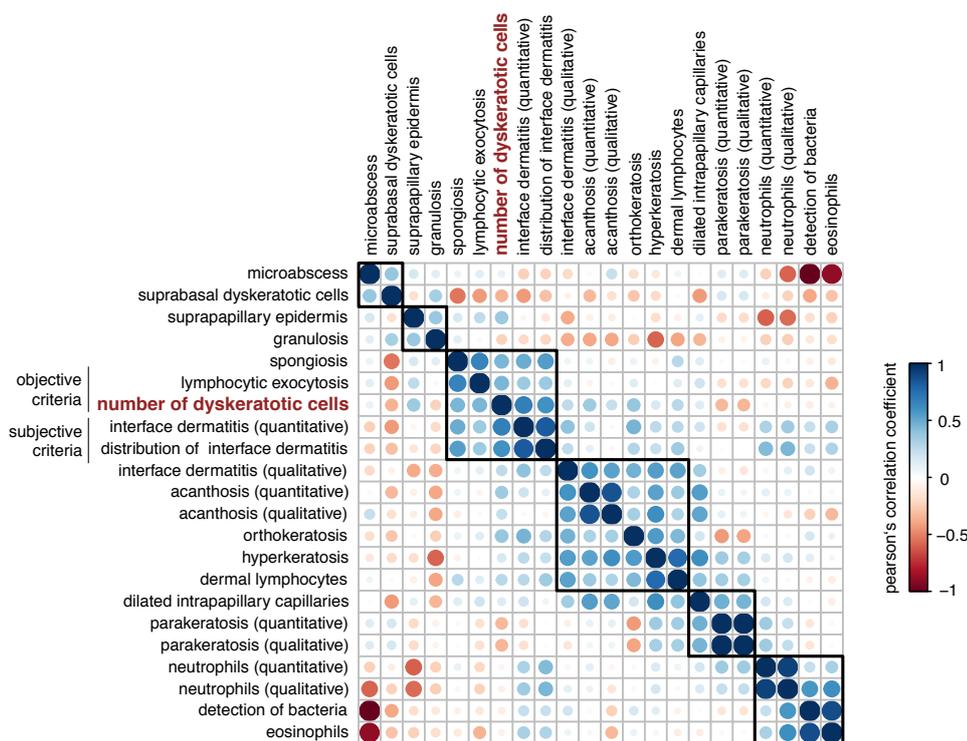


Fig. 5.1 – Computational correlation plot of histological attributes. The objective criteria “number of dyskeratotic cells” is the one which correlates highest with subjective interface dermatitis attributes. Pearson correlation coefficient were calculated and rows and columns are ordered using hierarchical clustering. Figure adapted from Fig. S1 in Lauffer et al. (2018).

Quality controlled number of probes that went into the analysis was 42,297. No averaging over probes which were specific for the same gene was performed because they potentially measure different splice variants or have different dynamic ranges. Issues with microarray measurements, also in comparison with RNA sequencing techniques, are discussed in section 2.2.4.

The aim of the study was to account for inter-individual variability and find shared molecular signatures for interface dermatitis. Therefore, linear mixed effects models were applied to adjust for patient-bias (details see section 2.1.5) and two model designs were used to analyze the gene expression data. First, gene expression of lichen planus, lupus erythematosus and psoriasis skin biopsies was analyzed regarding all three diseases as separate predictor variables in comparison to the autologous healthy measurement (= **model 1**). Second, both interface diseases were combined in one predictor and compared to autologous healthy skin and psoriasis gene expression (= **model 2**). The results of model 2 are referred to as the shared interface dermatitis signature. In both models a linear

mixed effects model was fitted using the restricted maximum likelihood criterion where individual patients were included as random effects in the model (= random intercepts). “This results in an intercept calculated for each individual patient (= random effect) and an overall adjusted fold change (= fixed effect) for each predictor (lichen planus, lupus erythematosus, psoriasis) compared to healthy.” (Lauffer et al., 2018) Each gene was modeled separately and with both model designs. Corresponding p-values were calculated using the Kenward-Roger approximation for the denominator degrees of freedom (Kenward and Roger, 1997) and adjusted for multiple testing using Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

Expression of a particular gene y was modeled by fixed effects β and random effects b . X and Z are design matrices relating y to β and b . The model can be represented as:

$$y = X\beta + Zb + \epsilon \quad \epsilon \sim N(0, \sigma^2 I) \quad b \sim N(0, \tau^2 I)$$

with $y \in \mathbb{R}^{n \times 1}$, $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^{p \times 1}$, $Z \in \mathbb{R}^{n \times s}$, $b \in \mathbb{R}^{s \times 1}$, $\epsilon \in \mathbb{R}^{n \times 1}$,

with s being the number of random effects (= number of patients), n being the number of measurements taken ($n > s$ since at least two measurements per patient) and p being the number of fixed effects. The described modeling designs differ in the number of fixed effects:

model 1: $p = 4$ (non-lesional, lichen planus, lupus erythematosus, psoriasis)

model 2: $p = 3$ (non-lesional, interface diseases combined, psoriasis).

As described in section 2.1.5 linear mixed effects models were estimated with the *lme4* package (Bates et al., 2015) in R (R Core Team, 2019). P-values for each coefficient were obtained using “mixed” function from the *afex* R-package (Singmann et al., 2018) which applies the Kenward-Roger approximation for the degrees of freedom (Kenward and Roger, 1997).

Gene expression of *in vitro* stimulated keratinocytes was modeled using linear regression (see section 2.1.4). To reveal differences and similarities between keratinocyte and interface dermatitis gene expression a correlation analysis between the fold changes of the genes of interest was performed using Pearson correlation. Top hits were defined as p-value < 0.05 and then sorted by absolute fold change.

Gene set analysis

Gene set analysis was applied to understand the context of differentially regulated genes (details see section 2.2.5). Gene sets were investigated using model-based gene set analysis (MGSA) which takes the hierarchical structure of gene sets into account (Bauer et al., 2010). Active gene sets are those whose posterior probability for being active, also called MGSA estimates, is larger than 0.5 (Bauer et al., 2010). Wikipathways were used as a database for known gene sets (Slenter et al., 2018).

5.1.3 Results

Interface dermatitis is dominated by a type I immune response

Evaluation of model 1 on whole-genome expression data revealed “5,675 genes [which] were regulated exclusively in lichen planus; 4,354 genes were regulated exclusively in lupus erythematosus. 3,888 genes showed differential regulation in both [interface diseases] [...] when compared to healthy skin [(Fig. 5.2 A)]. Only these genes were regarded as independent of the specific diseases [...] but shared among interface dermatitis diseases and were included in [...] [pathway] analysis. Pathway [...] [analysis using model based gene set enrichment analysis (Bauer et al., 2010) of shared genes (model 1, $n = 3,888$)] showed an activation of interferon, chemokine and T cell related pathways [(Fig. 5.2 B)].” (Lauffer et al., 2018) Interferon signaling pathways indicative of a type I immune response are overrepresented (indicated in red, Fig. 5.2 B).

The computational result was validated *in vitro*: T cells isolated from lesional skin of interface dermatitis patients showed high frequencies of IFN- γ and TNF- α positive cells in intracellular cytokine staining. Further, immunohistological stainings for TBX21, the major transcription factor for type I immunity, revealed higher number of TBX21 positive cells in interface dermatitis compared to psoriasis tissue (see Fig. 2D in Lauffer et al. (2018)).

Molecular signature of interface dermatitis resembles keratinocytes stimulated with IFN- γ and TNF- α

The next question was which T cell stimulus was able to change the gene expression in keratinocytes so that it is similar to the expression patterns in interface dermatitis. To answer the question, primary human keratinocytes were stimulated *in vitro* with cytokines

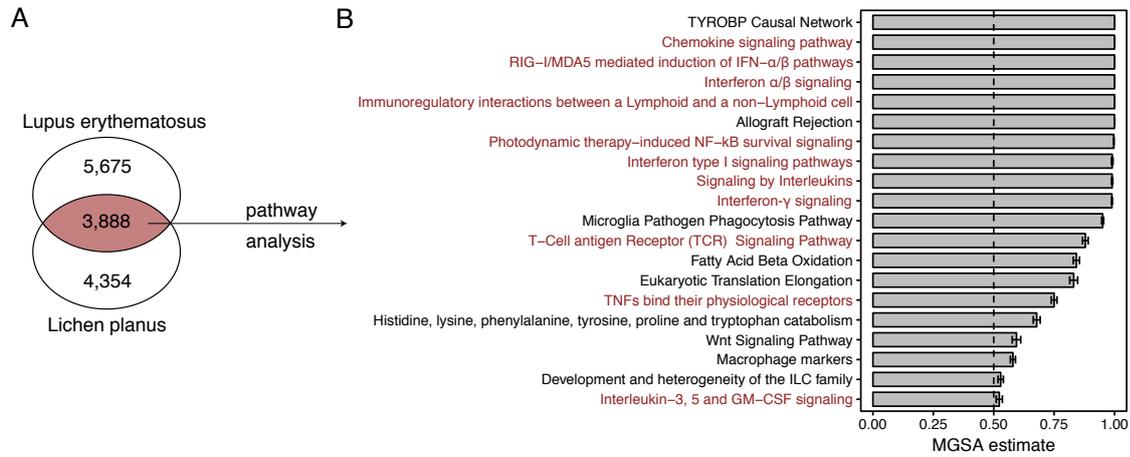


Fig. 5.2 – The molecular signature of interface dermatitis. (A) Transcriptome analysis of lichen planus (n=11) and lupus erythematosus (n=5). Depicted is the number of regulated genes shared and specific for the diseases. Only the shared genes were taken into account for pathway analysis. (B) Active pathways (estimate larger than 0.5) of shared interface dermatitis genes. Pathways related to type I immune responses are marked in red. Wikipathways were used as a pathway database (Slenter et al., 2018). Figure adapted from Fig. 1 B and C in Lauffer et al. (2018).

specific for certain T helper (Th) cell subsets: with IFN- γ and TNF- α for Th1 subset, with IL-4 and IL-13 for Th2 subset, with IL-17 and IL-22 for Th17 subset and with IL-22 for Th22 subset. The shared expression pattern of interface dermatitis, which was calculated using model 2, was compared to the top 100 most differentially regulated genes of keratinocytes stimulated with above mentioned cytokines. The highest correlation was detected for keratinocytes stimulated with Th1 cytokines IFN- γ and TNF- α (Fig. 5.3 for Th1 and Th2, not shown: Th17 condition with $r = 0.53$, and Th22 condition with $r = 0.31$). “[This indicates] that keratinocytes in interface dermatitis are exposed to a type I immune response microenvironment.” (Lauffer et al., 2018)

In an additional experiment using three-dimensional skin models, it was shown that keratinocytes which are stimulated with Th1 cyotkines IFN- γ and TNF- α show signs of cell death.

Necroptosis and apoptosis pathways are activated in interface dermatitis

To understand the underlying mechanism of the observed gene expression patterns, pathways were analyzed. The apoptosis and the necroptosis pathways were regulated in the interface dermatitis specific gene expression (Fig. 5.4).

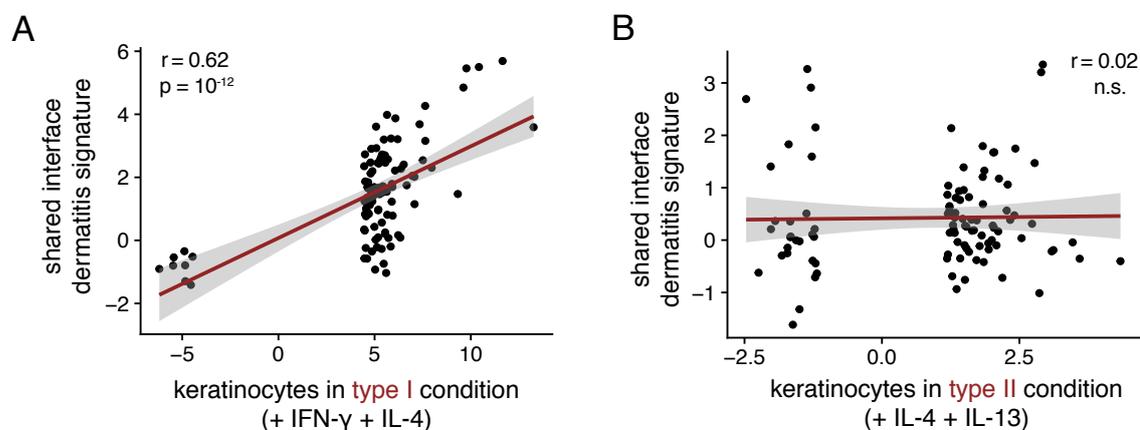


Fig. 5.3 – The gene expression pattern of interface dermatitis is similar to the expression pattern of keratinocytes stimulated with IFN- γ and TNF- α . Top 100 differentially regulated genes in stimulated keratinocytes were compared to the shared gene expression signature of interface dermatitis (model 2) based on fold changes. Shown are results for keratinocytes stimulated with cytokine milieu produced by Th1 (A) and Th2 (B) cells. Figure adapted from Fig. 3 B in Lauffer et al. (2018).

For further support this finding, markers for apoptosis (cleaved caspase 3) and necroptosis (receptor-interacting- protein-kinase 3 (RIP3)) were tested using immunohistochemistry. For RIP3 an enhanced expression in the epidermis was measured but not for cleaved caspase3. Further, depending on IFN- γ or TNF- α presence, phosphorylation of RIP3 was induced in keratinocytes when they were stimulated with supernatant of lesional T cells. After knock-down of RIP3, keratinocytes did not show any more cell death after stimulation with IFN- γ and TNF- α .

5.1.4 Discussion

Applying linear mixed models for determining differentially expressed genes while adjusting for inter-individual variability proved to be a useful method for this data set since the results were experimentally validated. “So far, it has been assumed that apoptosis was the mechanism leading to characteristic epidermal changes of interface dermatitis (Bascones-Ilundain et al., 2006; Skiljevic et al., 2017; Yoneda et al., 2008). [...] Beyond apoptosis, we here define two additional mechanisms associated to interface dermatitis: a type I dominant cellular immune response with the key cytokine IFN- γ based on gene expression analysis and an activation of the necroptosis pathway mediated by the phosphorylation of RIP3 in keratinocytes [using immunohistochemistry and small hairpin RNA knock-down].” (Lauffer et al., 2018) Recently, necroptosis inhibitors have been

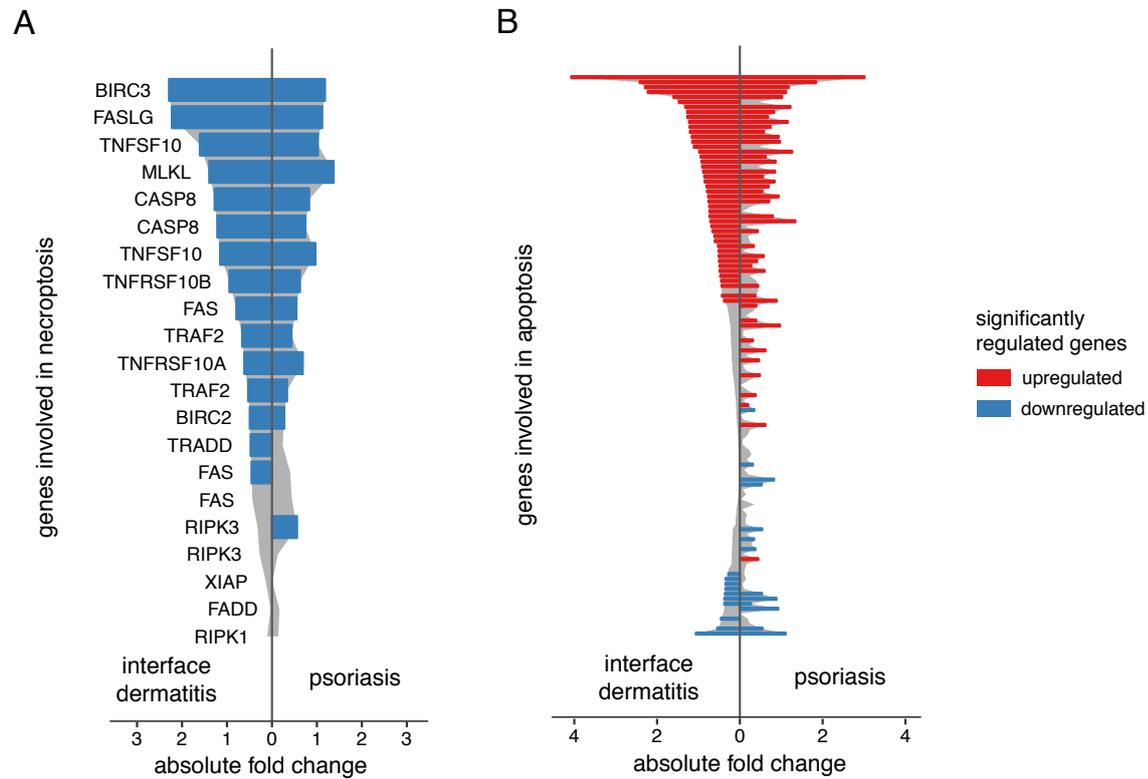


Fig. 5.4 – Regulation of genes belonging to necroptosis (A) and apoptosis (B) pathways in interface dermatitis compared to psoriasis. All genes which are annotated to the respective pathway are ordered by fold change in interface dermatitis. Both fold changes were obtained from results of model 2. Significantly regulated genes are highlighted in colors (red for upregulation, blue for downregulation). Figure adapted from Fig. S2 in Lauffer et al. (2018).

reported (Fauster et al., 2015; Yan et al., 2017) and together with the findings described here, these present assuring therapeutic approaches for interface dermatitis positive skin reactions.

5.2 Application in imiquimod-induced skin reactions

Linear mixed effects models were also applied to gene expression microarray measurements in a second study. In that study, the effect of topical application of a toll-like receptor 7/8 agonist, called imiquimod (brand name “Aldara”), to human skin was investigated. The study was conducted in collaboration with Natalie Garzorz-Stark, Felix Lauffer and Kilian Eyerich from the Department of Dermatology and Allergy at the University Hospital of Technical University Munich. Content, text and data of this section are based on Garzorz-Stark et al. (2018). Copied text passages are indicated as quotations.

In this study, the thesis author was responsible for gene expression analysis. Gene expression was measured with microarrays. Gene expression analysis for this study contained extensive data preprocessing, dimension reduction in a non-standard way and differential gene expression analysis using linear mixed effects models. There are two main differences in the analysis of gene expression data in this study compared to the study described in section 5.1 about interface dermatitis. First, the microarray preprocessing was done more stringently and included adjustment for unknown sources of variability. Second, the linear mixed effects model setup was more complex since more groups were compared. The thesis author further designed and created all figures shown in this section with a special focus on visualizing pathway analysis results comparing several conditions. The text passages concerning computational methods and results were originally written by the thesis author for the publication Garzorz-Stark et al. (2018).

5.2.1 Motivation

Psoriasis is an inflammatory skin disease (details section 2.4.1) where identification of early triggers remains difficult (Tian et al., 2012). Mouse models can only reflect human psoriasis to some degree (Swindell et al., 2011). So a human model to perform standardized analyses would be beneficial. The most often used mouse model of psoriasis is topical application of imiquimod which is a toll-like receptor 7/8 (TLR7/8) activator (Hawkes et al., 2017). This mouse model can reflect histologic hallmarks of psoriasis on mice skin (van der Fits et al., 2009) and it can also reflect molecular pathways (van der Fits et al., 2009; Grine et al., 2015). So far, it is not clear what the driver for the psoriasis-like skin inflammation induced by imiquimod in mice is but both dendritic cells and $\gamma\delta$ T cells are important for the process (Singh et al., 2016; Yoshiki et al., 2014).

“The aim of this study was to comprehensively characterize imiquimod-induced inflammation in human subjects and to evaluate its potential use as a standardized human psoriasis model.” (Garzorz-Stark et al., 2018)

5.2.2 Methods

Only study design and methods applied on gene expression data are given here. For further clinical and laboratory methods please see Garzorz-Stark et al. (2018).

Patient cohort

“Eighteen patients with or without a known history of psoriasis, atopic eczema or both were included in the study. [...] Imiquimod (Aldara 5% cream), [a TLR7/8 agonist,] [...] was applied at days zero, two, and four and subsequently twice per week for four weeks in an occlusive manner on the backs of the patients. [...] Development of skin lesions was monitored over a period of four weeks.” (Garzorz-Stark et al., 2018) Six-millimeter punch biopsies were obtained from lesional and noninvolved skin and cut into three pieces: one part for histologic assessment, one for gene expression analysis and primary cells were obtained from the last piece.

Preprocessing of whole-genome expression data

Total RNA was isolated from skin punch biopsy specimens and measured with SurePrint G3 Human GE 8x60K BeadChips (Agilent Technologies). Basic preprocessing and quality control was performed as for the other microarray studies described in section 2.2. For this study some additional steps were performed to exclude probes with low expression and low confidence. First, only probes were kept which were 10% brighter in any of the arrays than the 95% quantile of all negative control probes which filtered out 82 low expressed probes. Within-array replicated probes were averaged, different probes for the same gene were not averaged. Second, the algorithm *blastn* (Altschul et al., 1990) was used to map the 60-bp nucleotide sequences spotted on the array (probes) to the human transcriptome available from UCSC via Genbank (February 2016) and encoded as RefSeq IDs. Of all probes, 13,430 did not match with 100% accuracy to a position of the human transcriptome and were removed prior to analysis. For probes mapping to more than one RefSeq ID it was checked if these IDs corresponded to the same gene (via gene symbol). 2,229 probes mapped to several genes and were excluded from the analysis. Mapping between RefSeq ID

5.2 APPLICATION IN IMIQUIMOD-INDUCED SKIN REACTIONS

and gene symbol was performed with *org.Hs.eg.db* annotation package from Bioconductor's AnnotationDbi (Pagès et al., 2018). In total, 26,664 probes mapped with 100% accuracy to a unique gene and were used in further analysis.

Dimension reduction of gene expression data

To visualize whole-genome gene expression data, dimension reduction was employed (details in section 2.2.7). Since, the data set contained repeated measurements of individuals, no standard principal component analysis (PCA) was used but a new method introduced by Lin et al. (2016) called AC-PCA (adjustment for confounding principal component analysis). The method in parallel performs dimension reduction and adjusts for confounding factors (Lin et al., 2016). Patient heterogeneity was adjusted for by using patient identifier as confounding factors. The method was applied on all patient samples which each contained 26,664 normalized probes per sample. For better between-group comparison, 95% confidence intervals were calculated and added in two dimensional space.

Analysis of gene expression data

To estimate unknown sources of variability in the data set, surrogate variable analysis was performed (Leek and Storey, 2007) which is related to batch correction. One surrogate variable was calculated (Leek et al., 2012) and included as a covariate in the regression model. The surrogate variable was associated to chip IDs. Chip ID was not explicitly used as a confounder or for direct batch correction since the chips, which contain up to eight microarray measurements, also included measurements from other studies.

For each of the 26,664 probes one linear mixed effects model was fitted following the general description in section 2.1.5. The analysis was comparable to section 5.1.2 with the difference that five fixed effects were used in this model design: one surrogate variable, psoriasis, eczema, acute contact dermatitis (ACD) and imiquimod-induced contact dermatitis (ICD). Model estimation and p-value calculation was performed as described in section 5.1.2. P-values were adjusted for multiple testing by the Bonferroni correction (Dunn, 1961). Genes were defined as significantly differentially regulated when the adjusted p-value was less than 0.05 and top hits were defined when the absolute fold change was greater than 2.5.

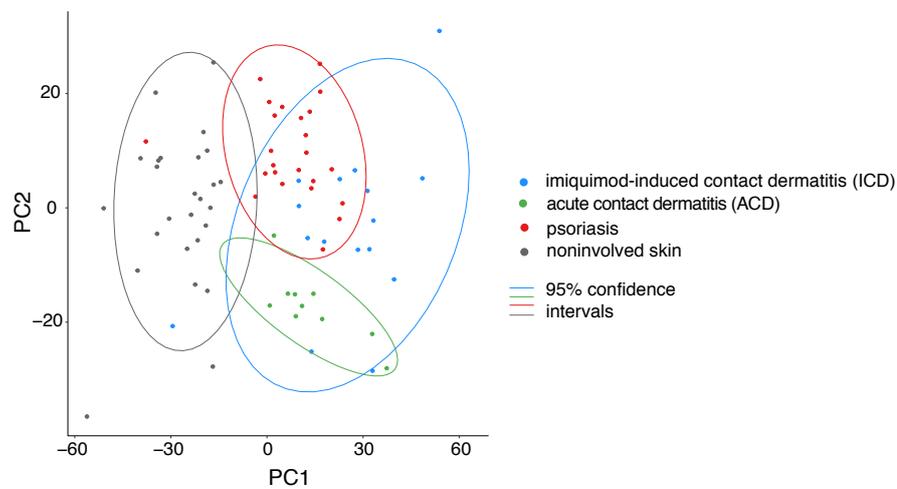


Fig. 5.5 – Principal component (PC) analysis which simultaneously adjusts for confounding variation on whole-genome expression data. Patients were taken as confounding variables. 95% confidence intervals are drawn per group of patients with ICD ($n = 15$), patients with ACD ($n = 10$), patients with psoriasis ($n = 24$) and compared to noninvolved skin ($n = 26$). Figure adapted from Fig. 2 A in Garzorz-Stark et al. (2018).

5.2.3 Results

“Imiquimod [induced] a monomorphic and self-limited inflammatory response in healthy subjects, as well as patients with psoriasis or eczema.” (Garzorz-Stark et al., 2018) The reaction was heterogeneous, looked contact dermatitis-like in histology and missed hallmarks of psoriasis. Also, the clinical phenotype of imiquimod-induced inflammation (ICD) in human skin resembled acute contact dermatitis (ACD) rather than psoriasis (see Fig. 1 in Garzorz-Stark et al. (2018)).

Transcriptome of human imiquimod-induced dermatitis closely overlaps with contact dermatitis

“Whole-genome expression analysis of [imiquimod-induced] lesional skin (ICD, $n = 16$) was compared with skin of patients with psoriasis ($n = 24$), patients with acute contact dermatitis to nickel (ACD, $n = 10$), patients with eczema ($n = 15$) and noninvolved skin ($n = 26$) to investigate ICD in a heuristic global approach. As a first step, dimension reduction simultaneously adjusting for the confounding variation from patient heterogeneity was performed (Lin et al. (2016)). A close overlap of ICD reactions with both ACD and psoriasis reactions compared with noninvolved skin was observed when examining the 95% confidence intervals per group [(Fig 5.5)].” (Garzorz-Stark et al., 2018)

5.2 APPLICATION IN IMIQUIMOD-INDUCED SKIN REACTIONS

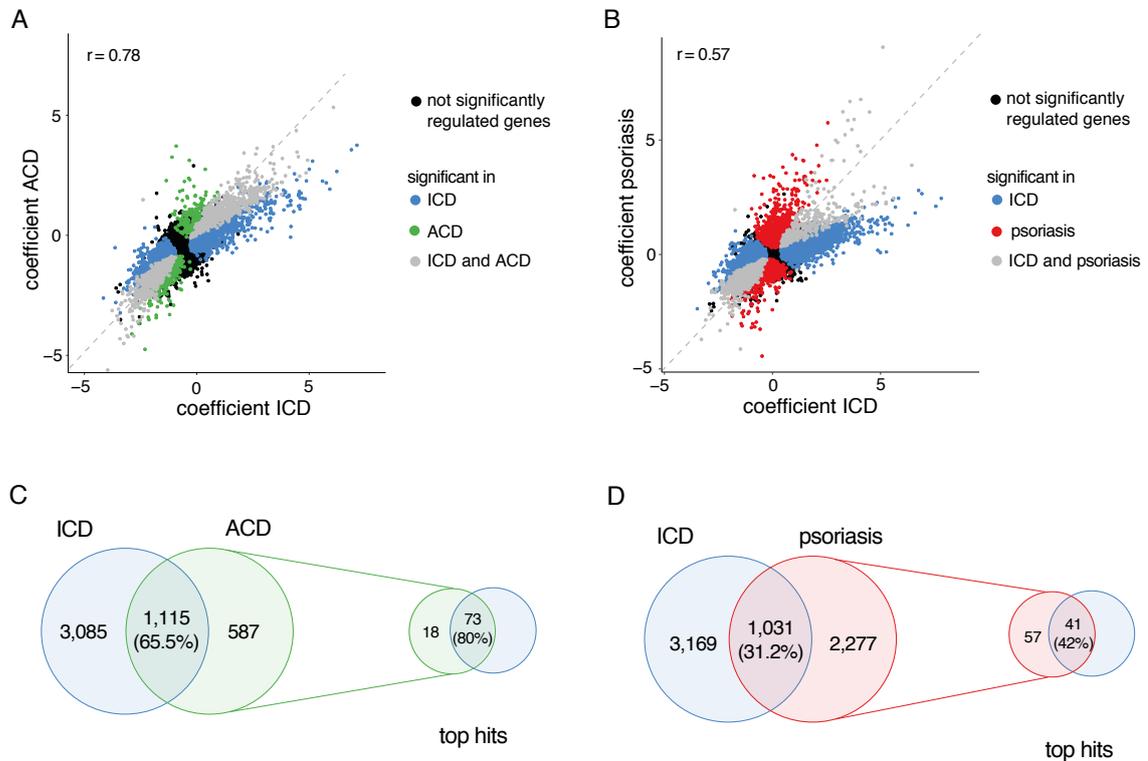


Fig. 5.6 – ICD transcriptome closely overlapped with ACD and, to a lesser extent, with psoriasis. (A, B) Correlation of fold changes of all genes of ICD and ACD (A) and of ICD and psoriasis (B), respectively, compared to intrinsic noninvolved skin. Colors indicate genes which are uniquely significantly regulated in one of the groups (ICD in blue, ACD in green and psoriasis in red), grey genes are shared between two groups and black ones are not significantly regulated. (C, D) Overlap of significantly regulated genes shown by Venn plots. Percentages indicate relative number of genes regulated significantly in either patients with ACD (C) or those with psoriasis (D) that are also regulated significantly in patients with ICD. Smaller Venn plots indicate overlap of top hit genes with an absolute log fold change of greater than 2.5. Pearson’s correlation coefficient r is given for comparison of fold changes in (A) and (B). Figure adapted from Fig. 2 B and C in Garzorz-Stark et al. (2018).

“For a more detailed insight into the similarity of ICD [compared to] ACD, eczema and psoriasis, significantly regulated genes were compared. The ICD and ACD transcriptomes showed a strong correlation of all significantly regulated genes ([Pearson’s correlation coefficient] $r = 0.78$, [Fig. 5.6 A]). In comparison, ICD and psoriasis ($r = 0.57$, [Fig. 5.6 B]), as well as ICD and eczema ($r = 0.56$, see supplementary Fig. E2 A in Garzorz-Stark et al. (2018)), correlated less. Furthermore, 65% of all significantly regulated genes in patients with ACD were also regulated in patients with ICD [(Fig. 5.6 C)], whereas 31.2% of the psoriasis [(Fig. 5.6 D)] and 37.1% of the eczema genes were also regulated in patients with ICD [(see supplementary Fig. E2 B in Garzorz-Stark et al. (2018))]. Among the top

hit genes with a log₂ fold induction of greater than 2.5, the overlap was higher, with 80.2%, 41.8% and 64.3% for the ACD, psoriasis and eczema transcriptomes, respectively [(Fig. 5.6 C, D)].” (Garzorz-Stark et al., 2018)

Apoptosis pathway is shared among ICD and ACD

One of the major pathogenic hallmarks of acute contact dermatitis (ACD) is apoptosis (Traidl et al., 2000). To assess whether ICD resembles ACD in this regard, gene expression fold changes of apoptosis-associated genes were directly compared (Fig. 5.7 A). ICD and ACD showed a strong correlation ($r = 0.89$, Fig. 5.7 B), whereas the correlation of ICD and psoriasis was less pronounced but still present ($r = 0.61$, Fig. 5.7 B).

To get a further understanding of immune cells involved in the process, immunohistological staining was performed which showed that ICD is infiltrated by more CD8⁺ T cells (89.3 ± 12.58) compared to psoriasis (42.3 ± 7.97 , $p = 0.0008$). However, the number of CD8⁺ T cell infiltrating ACD (64.8 ± 12.05) was similar to ICD. Differences between ICD and ACD were found for the levels of CD4/CD8 T cell ratio with more CD8⁺ cells in ICD compared to ACD and psoriasis. Both CD4⁺ and CD8⁺ T cells secreted IFN- γ in ICD and upon T cell receptor stimulation more IFN- γ was secreted by T cells isolated from ICD patients compared to ACD and psoriasis patients.

Regulation of Interferon- α/β signaling pathway is ICD-specific and induced by plasmacytoid dendritic cells

Analysis of gene expression data showed a specific, exclusive upregulation of the Interferon- α/β signaling pathway in ICD. The pathway consists of 73 genes and 53 are significantly upregulated in ICD, whereas only 14 and 13 are upregulated in patients with psoriasis and ACD, respectively (Fig. 5.8). ICD was induced by imiquimod which is a TLR7/8 agonist. Since TLR7/8 is highly expressed on plasmacytoid dendritic cells (pDCs) and they are the main producers of IFN- α , they were further investigated in patients with ICD reactions. The number of pDCs correlated with severity of clinical reaction as well as with TLR7 density, indicating a functional role for pDCs in patients with ICD. This is particularly interesting, since the underlying metabolic reprogramming induced by imiquimod was observed in both human (see Fig. 4 D in Garzorz-Stark et al. (2018)) and mouse (see Fig. E8 in Garzorz-Stark et al. (2018)) pDCs, hinting at a “shared cellular mechanism between mouse imiquimod-induced psoriasis-like reactions and ICD. [...] [This mechanism is also] in contrast to classical contact dermatitis, in which myeloid dendritic cells sense [small

5.2 APPLICATION IN IMIQUIMOD-INDUCED SKIN REACTIONS

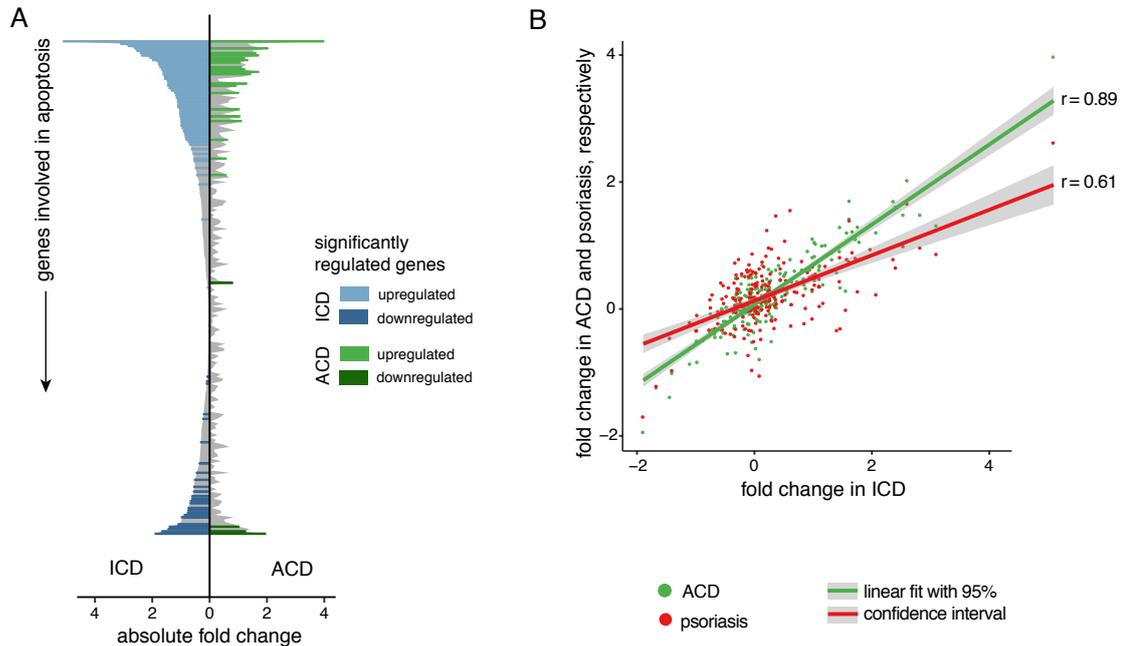


Fig. 5.7 – Differentially regulated genes within the pathway “apoptosis”. (A) Absolute fold change of each gene within the pathway “apoptosis” (Reactome database) in patients with ICD (left, blue) and those with ACD (right, green). Significantly up- and downregulated genes are highlighted in lighter and darker colors, respectively. (B) Correlation analysis of genes related to apoptosis. Shown are fold changes of genes in ICD lesions (x-axis) versus fold changes of genes in ACD (y-axis, green dots and line) or psoriasis (y-axis, red dots and line). Lines represent linear fits through the respective points and gray backgrounds are 95% confidence intervals. Pearson correlation coefficient is given by r . Figure adapted from Fig. E4 in Garzorz-Stark et al. (2018).

chemicals and not pDCs act as] [...] primary sensors (Kaplan et al., 2012).” (Garzorz-Stark et al., 2018)

“IL-23 mediated signaling events” is a key pathway upregulated in both psoriasis and ICD

“Because plasmacytoid dendritic cells are involved in the early pathogenesis of psoriasis, we next investigated the molecular overlap of ICD and psoriasis [...]: “IL-23 mediated signaling events” was identified as a key pathway upregulated in both psoriasis and ICD [(Fig. 5.9 A)].” (Garzorz-Stark et al., 2018) The pathway consists of 42 genes and 21 of these were significantly regulated in ICD. The overlap of significantly regulated genes between ICD and psoriasis consists of five genes (Fig. 5.9 B), nevertheless the pathway is also activated in psoriasis, just with different genes, 15 altogether. Especially, “IL-23A

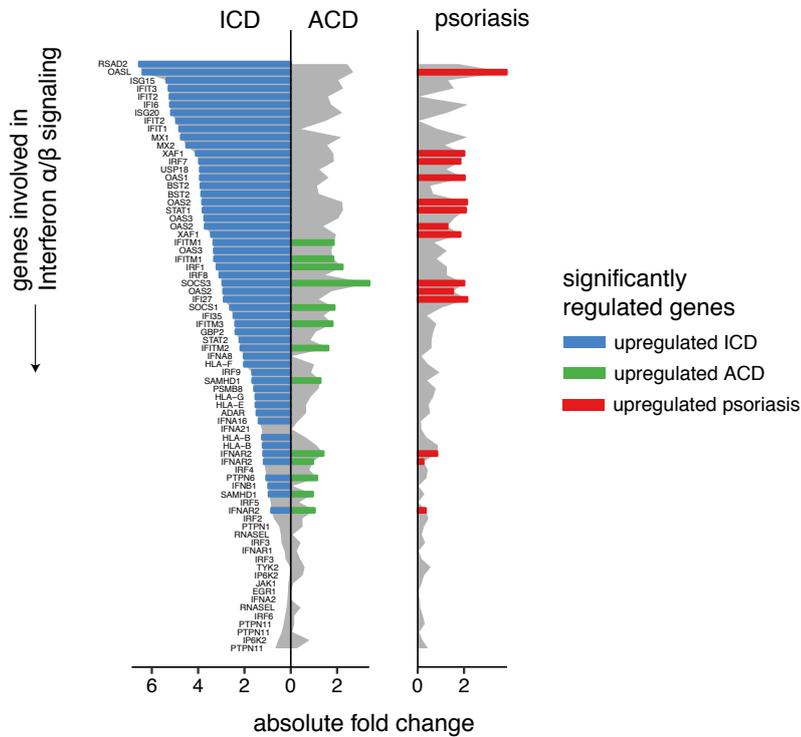


Fig. 5.8 – Regulation of Interferon- α/β signaling pathway. Shown is the absolute fold change in comparison of patients with ICD (left), ACD (middle) and psoriasis (right) in gray and significantly regulated genes are highlighted in colors (ICD blue, ACD green and psoriasis in red). Figure adapted from Fig. E6 in Garzorz-Stark et al. (2018).

was significantly upregulated in both patients with ICD and psoriasis but not in patients with ACD [(Fig. 5.9 A, highlighted)]. IL-23 is a key driver of Th17-immunity. Accordingly, [immunohistological stainings showed that] IL-17+ cells were more frequent in patients with ICD (10.3 ± 1.78) than in ACD (4.0 ± 0.83) and comparable [...] to psoriasis (9.2 ± 1.88) *in situ*.” (Garzorz-Stark et al., 2018) These findings were confirmed in T cells isolated from lesional skin of the respective diseases. ICD and psoriasis further show similarities in immune cell profiles: T cells derived from lesions secreted similar amounts of CXCL8 and the number of neutrophil granulocytes in lesions was comparable between ICD and psoriasis.

NOS2 is part of a new molecular classifier for psoriasis (see section 3.1 and Garzorz-Stark et al. (2016)). Gene expression analysis showed upregulation of NOS2 in psoriasis compared to intrinsic noninvolved skin (Fig. 5.9 A, highlighted), but no regulation in ICD and ACD.

5.2 APPLICATION IN IMIQUIMOD-INDUCED SKIN REACTIONS

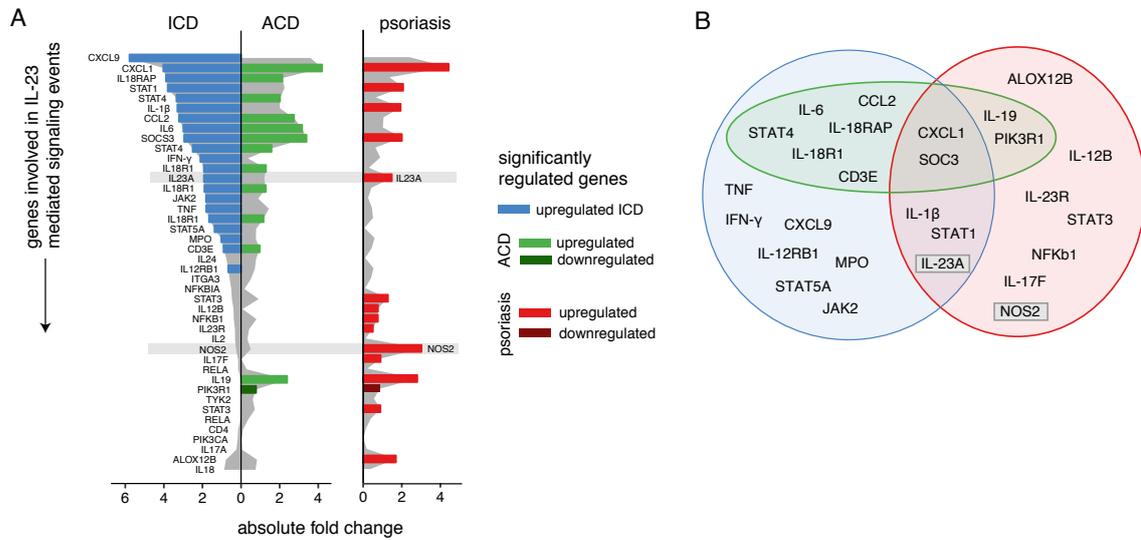


Fig. 5.9 – “IL-23 mediated signaling events” is a shared pathway among ICD and psoriasis. IL-23A is highlighted as a gene being regulated by ICD and psoriasis but not ACD. NOS2 was shown to differentiate between psoriasis and forms of eczema (Garzorz-Stark et al., 2016) and is exclusively regulated in psoriasis. (A) Differentially regulated genes from “IL-23 mediated signaling events” (Reactome database). Shown is the absolute fold change of each gene (gray background area). Significantly regulated genes are indicated in colors. ICD (blue, left), ACD (green, middle) and psoriasis (red, right) are compared. (B) Venn diagram to compare significantly regulated genes among ICD (blue), ACD (green) and psoriasis (red). Figure adapted from Fig. E9 in Garzorz-Stark et al. (2018).

However, using immunofluorescence staining the induction of NOS2 in lesional skin of both ICD and psoriasis was shown. In contrast, NOS2 was essentially missing in patients with ACD (Fig. 5, part I in Garzorz-Stark et al. (2018)).

In a one-patient trial the influence of IL-23 on ICD was examined in a prospective manner. The results showed that neutralizing the effect of IL-23 reduced the clinical ICD reaction which supports the hypothesis that IL-23 is important for imiquimod-induced inflammation.

5.2.4 Discussion

“A standardized human model of psoriasis is missing. In this study we evaluated a commonly used mouse model of psoriasis for its possible value in the human setting.” (Garzorz-Stark et al., 2018) The key challenge of the analysis was comparing several gene expression measurements in different conditions from the same person. It was solved with a linear mixed effects model which adjusted for inter-individual variability and revealed condition specific gene signatures which were experimentally validated. Visualizing pathways and

CHAPTER 5 ADJUSTING PATIENT-BIAS

contrasting expression levels within these pathways between different diseases allowed to detect shared and specific genes involved in underlying mechanisms.

After application of imiquimod, human skin showed more signs of acute contact dermatitis than of psoriasis. The same observation can be made on transcriptome level. Nevertheless, this study shows that the human imiquimod model still allows insights into the pathogenesis of psoriasis. A trait which is shared with human psoriasis is the trait that plasmacytoid dendritic cells act as primary sensors.

Chapter 6

Molecular characterization of human T helper cell subsets using multiple omics levels

T helper cells are part of the adaptive immune system and if their regulation is out of balance, it leads to complex diseases like psoriasis, an inflammatory skin disease. This chapter describes the comprehensive approach taken by the thesis author, her supervisors and her collaborator Stefanie Eyerich to provide a deeper molecular characterization of T helper cell phenotypes. Seventy-nine human T helper cell clones were established in the laboratory of Stefanie Eyerich. T helper cell clones consist of several hundred thousand cells which all descent from one parent cell and share the same phenotype. The clones were either stimulated with a generic T cell stimulant or left without stimulation, before secreted proteins and mRNA expression were measured. The aim was to understand what molecular markers are specific for each clone or groups of clones to potentially improve patient treatment.

The study has been performed with T cell clones and not single cells since the parallel measurement of secreted proteins and gene expression from single cells is still challenging, especially after stimulation of cells (Macaulay et al., 2017). The Human Cell Atlas (HCA) is an initiative which aims to measure all cells of the human body at single cell resolution (Regev et al., 2017). This resource will provide single cell data sets mainly from healthy individuals and only small cohorts of patients with relevant diseases are included (Regev et al., 2017). Activated T helper cells are mostly found during an active immune response, so those small cohorts of diseased patients are of particular interest for future analysis and characterization of human T helper cell subsets once the Human Cell Atlas data is fully measured and available.

CHAPTER 6 CHARACTERIZING T CELLS

Two methods, both first described in 2017, enable scientists to measure gene and protein expression of single cells. Both cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq, Stoeckius et al. (2017)) and RNA expression and protein sequencing assay (REAP-seq, Peterson et al. (2017)) use similar approaches. They use antibodies specific for the proteins of interest which are linked to a tripartite DNA sequence (Todorovic, 2017). The DNA sequence allows to measure the protein expression in parallel to the transcriptome. Applying these methods to T helper cells in human disease will improve their characterization substantially. The methods focused on surface protein expression. T helper cells interact with the environment through the release of proteins. Measuring those secreted proteins is still not possible with these novel techniques on single cell level.

Here, the focus is on the analysis of T helper cell clones. The thesis author in discussion with her supervisors proposed an approach to perform unbiased clustering of T helper cells into groups based on their measured cytokine secretion profile. Analyzing the relationship of computationally defined T helper cell groups to known T helper cell subsets revealed the presence of known subsets in the data set but also of mixed T helper cell subsets. Next, the thesis author used whole genome gene expression data to characterize all T helper cell subsets on a molecular level by applying six different statistical modeling approaches and identifying a small set of subset specific marker genes. Gene expression results showed consistence with expectations in skin diseases driven by specific T helper cell subsets. Mapping candidate genes to a protein-protein interaction network revealed possible target proteins important for several T cell subsets. Further experimental validation is currently ongoing before proposing the identified genes as new markers for T cell subsets.

In this study, the thesis author was responsible for all data analyses including analysis of secretome and whole-genome gene expression data measured by microarrays. She was also responsible for interpreting the results. Further, she performed clustering analyses using five different clustering algorithms. After going through the full microarray preprocessing workflow, the thesis author performed differential gene expression analysis using six different methods, some of them not commonly used for this task. The thesis author further came up with an elaborate approach to determine most differentially regulated genes per method and how to find consensus top hits. She also performed gene set enrichment analysis and mapping to a protein-protein interaction network. All figures presented in this section were designed and created by the thesis author. The thesis author did not perform any of the laboratory experiments. The biological interpretations of the results were researched and written by the thesis author and approved by her supervisors and biological collaborators.

6.1 Motivation

On a daily basis, the immune system and especially T helper cells (Th cells which are positive for the CD4 surface molecule, referred to as CD4+ T cells) help in fighting pathogens. Th cells coordinate the adaptive immune response by secreting key cytokines, small proteins for inter-cellular communication. Seven Th cell subsets are described in the literature with specialized roles in the human defense system.

In 1986 **Th1** and **Th2** were described by Mosmann et al. (1986) in mice as cells secreting IFN- γ and IL-4, respectively. Th1 cells are specialized to fight against intracellular pathogens, e.g. bacteria and viruses which reside inside cells (Fietta and Delsante, 2009). Interestingly, an overshooting Th1 response can lead to a lethal immune response by the induction of liver pathology and intense necrosis (Gazzinelli et al., 1996). Pathogens do not only attack cells but they can be extracellular parasites like worms. In the defense against these pathogens Th2 cells are involved. This beneficial activity itself can turn pathogenic since Th2 cells are implicated in allergic diseases like asthma (Romagnani, 1994).

For 20 years it was thought that only Th1 and Th2 subsets of T helper cells exist. Until 2006, Bettelli and colleagues have described the **Th17** cell, a T helper cell, which secretes IL-17 (Bettelli et al., 2006). In the following years, the group of Th cell subsets was expanded and now also includes **Th9** (Veldhoen et al., 2008) and **Th22** (Eyerich et al., 2009; Duhon et al., 2009; Trifari et al., 2009) cells, which secrete their signature cytokines IL-9 and IL-22, respectively. Th9 cells promote inflammation in several models but in particular they are important for allergic inflammation (Chang et al., 2010). Th17 cells promote the immune responses against fungi and extracellular bacteria and are thought to be involved in chronic inflammation in autoimmune diseases (Korn et al., 2007). Th22 cells can be found in the human skin and are implicated in epidermal reactions in inflammatory skin diseases (Eyerich et al., 2009).

T helper cells further include T follicular helper cells and regulatory T cells. T follicular helper cells (**Tfh**) were first described in 2000 as CD4+ cells in tonsils which express high levels of CXCR5 (Breitfeld et al., 2000). Tfh provide B cells help by supporting their survival and differentiation into plasma and memory cells (Crotty, 2014). They secrete IL-21 (Bentebibel et al., 2011), CD40L and IL-4 (Crotty, 2011) upon stimulation. Opposing these Th subsets which mainly induce inflammation, regulatory T cells (**Tregs**) prevent unwanted immune reactions, like in autoimmune diseases, and maintain tolerance to self (Geginat et al., 2013; Sakaguchi et al., 2008). Tregs are specialized for immune suppression and secrete TGF- β and IL-10.

CHAPTER 6 CHARACTERIZING T CELLS

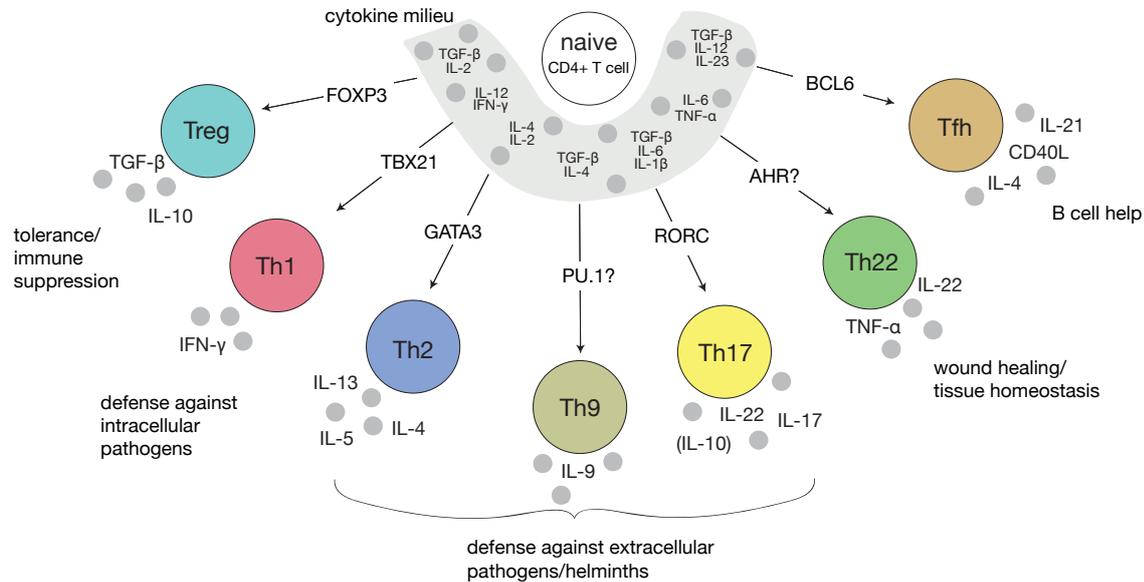


Fig. 6.1 – Overview of common T helper cell subsets. T helper cell subsets differentiate from naive CD4+ T cells in different cytokine milieus (indicated in grey) by switching on their respective transcription factor (name is visible along the arrow) and perform their function by secreting their signature cytokines. Transcription factors followed by a question mark are proposed but not yet accepted.

Figure 6.1 gives an overview of the yet known human Th subsets, their signature cytokines, transcription factors and function. Transcription factors for Th1, Th2, Th9, Th17, Th22, Treg and Tfh were already described as TBX21 (Szabo et al., 2000), GATA3 (Zheng and Flavell, 1997), PU.1 (Chang et al., 2010), RORC (Ivanov et al., 2006), AHR (Ramirez et al., 2010), FOXP3 (Fontenot et al., 2003) and BCL6 (Johnston et al., 2009; Nurieva et al., 2009; Yu et al., 2009), respectively. Where both AHR as a transcription factor for Th22 and PU.1 for Th9 have not been indisputably proven yet (Eyerich and Zielinski, 2014).

Specific T helper cell subsets are differentiated from naive CD4+ T cells *in vitro* in distinct cytokine milieus. Th1 cells are *in vitro* differentiated by IL-12 (Hsieh et al., 1993) and IFN-γ (Lighvani et al., 2001). Th2 need IL-2 and IL-4 (Le Gros et al., 1990) to differentiate from naive T cells. Th9 cells are *in vitro* differentiated by TGF-β and IL-4 (Veldhoen et al., 2008). For Th17 TGF-β (Veldhoen et al., 2006), IL-1β and IL-6 (Bettelli et al., 2006) have to be in the cytokine milieu. Th22 are *in vitro* differentiated by TNF-α and IL-6 (Duhon et al., 2009). Treg need TGF-β and IL-2 (Zhu et al., 2009) for *in vitro* differentiation. Tfh cells are *in vitro* differentiated by TGF-β, IL-12 and IL-23 (Schmitt et al., 2014).

Most of the experimental research on human T cells is performed in blood which only contains 2% of human T cells (Farber et al., 2014). Still, using state-of-the-art experimental techniques, new immune cell subsets are described regularly. Kunicki et al. (2018) identified new immune cell subsets using single-cell mass cytometry. Also single-cell RNA sequencing plays an increasing role in identifying new signaling molecules for known Th subsets and aiding in further understanding their phenotypes (Mahata et al., 2014). Villani et al. (2017) determined several new subtypes of human monocytes and dendritic cells from human blood using single-cell RNA sequencing. Since most of the T cells reside in peripheral tissue, the majority of human T cells are still unexplored and provide an interesting field of study (Zielinski, 2017).

State-of-the-art to distinguish Th subsets consists of three parts. First and most important for classification is the measurement of the **secreted cytokines** (Fig. 6.1 protein names next to differentiated T cells). Second, major **transcription factors** are taken into account (Fig. 6.1 protein names along arrows). Third, specific **surface molecules** are assessed. For example, Th1 cells are characterized by secreting IFN- γ (Mosmann et al., 1986), expressing the transcription factor TBX21 (Szabo et al., 2000) and the surface markers CXCR3 and CCR1 (Eyerich and Zielinski, 2014).

There are some problems with these three lines of classification. First, there are no clearly defined thresholds for the amount of a specific signature cytokine cells have to secrete in order to be assigned to a certain subset. Second, for some subsets the major transcription factor is not known. For TBX21, GATA3 and RORC it was shown that they are essential for the differentiation of naive cells into Th1, Th2 and Th17 cells, respectively. In contrast, for Th9 and Th22 cells there is no major transcriptional regulator known so far. Third, not for all Th subsets the surface molecules are clearly defined. For Tregs Bruder et al. (2004) described Neuropilin-1 as a specific surface marker which was later shown not to be specific in humans (Milpied et al., 2009).

6.2 Methods

Establishment of stable T helper cell clones

Skin biopsies were taken from the lesional skin of 13 patients suffering from inflammatory skin diseases and cultured in T cell medium plus IL-2 until T cells migrated out of the tissue. To perform T cell cloning, the obtained T cell suspension was diluted and dispersed to 96 well plates with one T cell per well on a feeder layer of irradiated peripheral blood mononuclear cells in the presence of a mitogen (phytohemagglutinin) and IL-2. This one T cell was then expanded into a T cell clone with a medium change every second day and IL-2 supplementation. All cells within this clone originated from the same single cell and thus show an identical phenotype. T cells were derived from several donors suffering from different diseases which are specific for different T cell imbalances and consequently not the whole spectrum of T cell subsets was obtained from one donor.

To check the phenotype stability of each clone, 10^6 cells were stimulated with anti-CD3 and anti-CD28 antibodies for six hours to analyze mRNA and for 48 hours to obtain culture supernatants for protein analysis by ELISA or Luminex (see methods section 2.3). Cells were then cultured for two weeks without further stimulus, but a medium change including IL-2 every second day. This procedure was repeated additional two times and finally 79 clones were included into the study since they showed identical results in all stimulations and therefore a stable phenotype.

Measurements of stable T helper cell clones and data processing

One million cells of each stable T cell clone were stimulated with anti-CD3 and anti-CD28 or remained unstimulated for six hours for whole genome expression analysis using Agilent SurePrint G3 Human Gene Expression 8x60K microarrays or for 48 hours to obtain culture supernatants for protein analysis by ELISA or Luminex. Microarray technology is explained in section 2.2.2. All processing steps applied for the microarray data are given in section 2.2.3. ELISA and Luminex data are analyzed as described in section 2.3.

Clustering

Principles of clustering and applied algorithms are described in detail in section 2.1.10. 79 T cell clones were clustered based on their cytokine secretion of IFN- γ , IL-4, IL-17

and IL-22 measured by ELISA (ELISA methodology explained in section 2.3.1). ELISA concentration were transformed according to procedure given in Table 2.2 which handles out of range values by transforming all measurements according to the standard values with known concentrations to arbitrary units of 1 - 10.

Five different clustering algorithms were applied on ELISA data (more details in section 2.1.10). First, k-means was applied which minimizes squared Euclidean distances between all data point within a cluster to the cluster center (Steinhaus, 1956; MacQueen et al., 1967). The cluster center are means within each cluster and are iteratively updated. Since k-means only finds a local optimum, the algorithm was executed 10,000 times with different starting positions and the result with the smallest error was used. Second, partitioning around medoids (pam, Kaufman and Rousseeuw (1987)) was used which employs a similar approach to k-means but cluster centers are data points, so it is more robust. Several distance measures for pam exist, here Euclidean distances were used. The third and fourth clustering algorithms are agglomerative (bottom-up) hierarchical clustering algorithms where at the beginning every observation is its own cluster and consecutively clusters are merged to larger clusters based on their similarity. The third clustering algorithm which was applied, uses complete linkage between two clusters which is the maximum distance between any two observations from the two clusters (formula in section 2.1.10, Johnson (1967); Lance and Williams (1967)). So the algorithm merges clusters if all observations in both clusters are similar. The fourth algorithm applies Ward's method (Joe and Ward, 1963) which minimizes increase in total within-cluster variance and leads to compact, spherical clusters. The last algorithm which was used was Gaussian mixture models which perform distribution-based clustering by fitting k Gaussians to the observations using expectation maximization (Dempster et al., 1977). Shape, volume and orientation of Gaussians were optimized using *mclust* package in R (Scrucca et al., 2017). For all k diagonal, equal volume and shape multivariate mixtures were used which corresponds to parameter "EEI" in *mclust*.

Number of clusters k was chosen by comparing average silhouette coefficients (formulas in section 2.1.10) between two and twenty clusters for all five methods. Here, k was chosen by balancing maximizing both agreement of all methods and average silhouette coefficient. To obtain clustering independent of chosen algorithm strict consensus clustering was performed. Samples were grouped into consensus clusters if all applied clustering algorithms grouped these samples together.

Differential gene expression analysis of Th cell clones

In gene expression analysis of Th cell clones the aim was identification of unique markers for each group of Th cells. So, not simply one tool for differential expression analysis was applied but six different ones. Those methods comprise standard tools for differential gene expression analysis like *limma* (Ritchie et al., 2015) and also newly introduced methods like the quantile approach proposed recently by Aran et al. (2017). Further, regularized regression models were applied to determine differentially expressed genes. They are not commonly used for differential gene expression analysis, but there are some examples: Wu (2005) proposed to detect differentially expressed genes by penalized linear regression models for classification tasks in microarray data. Ma et al. (2007) suggested to use supervised group lasso for predictive modeling of gene expression to account for clusters in gene expression data. Omranian et al. (2016) applied a fused lasso formulation to reconstruct gene regulatory networks from time-resolved microarray data. Also Zuo et al. (2017) used a special formulation of lasso, differentially weighted graphical LASSO, to incorporate prior knowledge network-based differential gene expression analysis.

Differential gene expression analysis for the Th cell clones posed two main challenges. First, none of the Th cell subsets was suited for being a reference or baseline so the definition of the baseline (= intercept) in the differential expression modeling approaches was not clear. Second, the collaborators were interested in a small but specific set of subset marker genes for each of the observed subsets. The thesis author applied six different methods to find these subset specific genes in a consensus approach (overview in Tab. 6.1).

T cell receptor genes were removed before differential gene expression analysis since it is known that particular T cell receptor genes are specific for some diseases (Attaf et al., 2015). Since disease specific genes were not of interest but subset specific ones, they were removed prior to analysis. The thesis author calculated differentially expressed genes for seven subsets, two stimulation conditions and both up-and downregulation, so altogether 28 sets of differentially regulated genes were determined.

Broadly, the applied methods represent two categories: methods which calculated a result separately per gene and those which investigated all measured genes in one model. The first category included linear models calculated with the *limma* package (Ritchie et al., 2015) but also quantile based approaches. Since the thesis author after discussions with her supervisors and collaborators did not want to set one Th subset as reference category, *limma* was used to calculate one linear model per subset per gene where the response was gene expression of all subsets for one of the genes. The method estimated two coefficients:

one for the intercept containing all subsets except for one and a second which described the difference in gene expression of the subset of interest compared to the remaining subsets. Since the size of some Th subsets was rather small two approaches to define what was represented by the intercept were applied. The first approach (called “**limma 1 vs. all**”) put all measured Th cell clones, except for the subset of interest, in the intercept. In the second approach (called “**limma 1 vs. sampled intercept**”) an intercept was sampled using a subgroup of Th cell clones of the same set size as the subset of interest. Sampling was performed 10,000 times and in two steps: first one of the six remaining subsets was sampled with equal probability, then within this subset one clone was sampled without replacement. Genes were considered significantly upregulated if the coefficient’s 0.05% quantile was above zero and downregulated if the coefficient’s 99.95% quantile was below zero.

The first category of methods generating a result for every gene separately also comprised the “**quantile approach**” which was used by Aran et al. (2017). The first step was to calculate for every Th subset the quantiles of its gene expression levels. 13 quantiles were calculated: 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85% and 90%. In the second step, for every gene it was checked whether the lower quantile in the subset of interest is higher than the maximum value of all higher quantiles of the remaining subsets. Here, both the lower and the higher quantile can be any one of the 13 quantiles, but the lower quantile should be smaller or equal to the higher quantile. E.g. if the lower quantile was chosen to be 10% and the higher quantile was chosen to be 90%, the 10% quantile of gene expression in the subset of interest was compared to the 90% quantile in the remaining subsets. If the maximum of all remaining 90% quantiles was lower than the 10% quantile of the subset of interest, this gene was called upregulated in the subset of interest, otherwise it was called not regulated. To determine downregulated genes, the minimum of 10% quantiles of the remaining subsets was compared to whether it was higher than the 90% quantile of the subset of interest. With this approach one subset was tested against all other subsets. The quantile approach required one parameter to be adjusted manually, namely the quantile where the above described comparison is performed. It was adjusted so that at least 20 genes fulfilled the criteria. For these 28 sets, seven times 15% (and 85%) were the quantiles chosen for defining most differentially expressed genes. Nine times it was 20% (and 80%), seven times 25% (and 75%) and five times 30% (and 70%).

The second category of methods, which estimated one model with all measured genes, included penalized regression methods to perform feature selection (see section 2.1.4). One binomial logistic regression with elastic-net penalty ($\alpha = 0.5$) was applied where one subset was tested versus all remaining ones and two multinomial logistic models were applied.

Binomial logistic regression with elastic-net penalty can give different results for different runs since the method performs internal cross-validation to define the regularization parameter λ . To overcome this problem Meinshausen and Bühlmann (2010) described “**stability selection**” which is implemented in R’s *stabs* package. It provides an upper bound to the per-family error rate, i.e., the expected number of false positives (see section 2.1.2 about multiple testing and 2.1.9 about feature selection with stability selection). The sampling scheme of Shah and Samworth (2013) with 50 complementary pairs leading to 100 subsamples was used. The result was a list of genes which fulfilled the criteria but without coefficients and thus no information about direction of regulation.

Also part of the second category of methods are multinomial logistic regression models. Applying penalized maximum likelihood in this setting allowed for two different kinds of lasso penalty. The first penalty tolerated for every group in the multinomial regression setting to select different variables (called “**elastic net multinomial ungrouped**”). The second is called grouped lasso penalty and ensures that for all groups which were commonly modeled the same variables were selected (called “**elastic net multinomial grouped**”). Both methods provided estimated coefficients for each variable but due to the internal cross-validation of the regularization parameter λ different results were calculated in each run. There is no stability selection procedure described for multinomial logistic regression. Each estimation was performed 1,000 times. How often each predictor was selected as well as range, mean, median, and confidence interval of the coefficients were evaluated.

Finding most differentially regulated genes for each applied differential gene expression method

For each of the six applied methods for differential gene expression analysis, most differentially regulated genes were determined differently depending on the method’s result structure. Forty most differentially regulated genes (20 up- and 20 downregulated) per stimulation condition were identified. For both limma-based methods, most differentially regulated was defined as having a significant regulation ($FDR < 0.05$ for “limma 1 vs. all”; 0.05% and 99.95% quantile of coefficient above or below zero for “limma 1 vs. sampled intercept”) and having the highest and lowest fold changes for up- and downregulated genes, respectively. In the sampling approach the median fold change was used.

To find most differentially regulated genes for the methods that calculated 1,000 elastic net runs (“elastic net multinomial grouped” and “elastic net multinomial ungrouped”) genes were chosen, which were selected in more than 50% of the runs and they were sorted by highest first quartile (upregulated genes) and lowest third quartile (downregulated genes)

Tab. 6.1 – Overview of applied differential gene expression analysis methods. The first column describes the name of the method in this thesis, the second gives more details about the method and the third depict the literature reference.

Method	Information	Reference
limma 1 vs. all	fits a linear model for each gene; tests one subset against all remaining	Ritchie et al. (2015)
limma 1 vs. sampled intercept	fits a linear model for each gene; tests one subset against a sampled subgroup of the same size equally distributed among the remaining subsets (10,000 runs)	Ritchie et al. (2015)
elastic net multinomial grouped	multinomial logistic model via penalized maximum likelihood ($\alpha = 0.5$); all genes as covariates; grouped lasso penalty ensuring that all variables are in or out together (1,000 runs)	Zou and Hastie (2005)
elastic net multinomial ungrouped	multinomial logistic model via penalized maximum likelihood ($\alpha = 0.5$); all genes as covariates; for each Th subset a different set of variables was selected (1,000 runs)	Zou and Hastie (2005)
stability selection	fits a binomial logistic regression per subset with all genes as covariates and selects influential variables with error control	Meinshausen and Bühlmann (2010)
quantile approach	finds genes where lowest quantile in one subset is higher than highest quantile in all other groups for 13 different quantiles	Aran et al. (2017)

of estimated coefficients. From both sorted lists the top 20 genes were defined as most differentially regulated. Since “stability selection” only selected less than 20 genes for all comparisons no further selection was necessary.

To determine most differentially regulated genes in the “quantile approach”, first the combination of quantiles with at least 20 genes was identified. Then, mean expression in all Th clones was calculated and genes with top 20 mean expression were selected.

Finding consensus top hits

The aim was finding differentially regulated genes independent of the applied statistical method by defining consensus top hits. So intersections, on microarray probe ID level,

CHAPTER 6 CHARACTERIZING T CELLS

of 20 most differentially regulated genes from each of the six methods were determined (example in Fig. 6.11). Top hits are those probes which are detected as most differentially regulated by at least three methods.

Pathway analysis

Gene set enrichment analysis of top hits for each subset and all top hits together was performed using *clusterProfiler* R-package and the “enrichGO” function on gene subontologies biological processes and molecular functions (Yu et al., 2012). Subset specific genes were tested against the universe of all genes which were measured on the microarray. Each tested Ontology term included at least a minimal size of five annotated genes. The cutoff for FDR-adjusted p-values was set at 10%. Up to six pathways per tested gene set and only pathways where at least two genes are in the pathway and in tested gene sets were displayed. If the same combination of genes was significantly enriched ($FDR < 10\%$) in several pathways, the pathway with the lowest adjusted p-value was displayed.

Mapping to protein-protein interaction network

To visualize and understand the interaction between top hits a protein interaction network analysis using the tool NetworkAnalyst was performed (Xia et al., 2015). The input were all top hits for each of the seven Th subsets. Protein-protein interactions were looked for based on the IMEx Interactome (Literature-curated comprehensive data from Breuer et al. (2012)) where no threshold has to be applied. Only first-order networks were investigated, which included the queried genes and their direct interaction partners in the network.

For clarity only proteins linking queried genes were depicted not those which interacted with the genes of interest but no further protein. So, all not queried proteins with degree one were removed for better visibility which decreased the network from 706 nodes connected by 842 edges to 162 nodes and 267 edges. Further, the protein Polyubiquitin-C was removed in the visualization (40 connections to the remaining 162 nodes) for a more comprehensive view.

6.3 Results

6.3.1 Computational identification of seven T helper cell subsets from major cytokine secretion

The aim was to identify T helper (Th) cell subsets in an objective manner. Traditionally, Th cell subsets were defined based on cytokine secretion profiles of four key signature cytokines: IFN- γ , IL-4, IL-17 and IL-22. Further information about each Th subset, like transcription factors and surface markers, were investigated and described after more detailed studies. To be in line with standard laboratory procedures of Th subset identification, protein secretion data was analyzed for the four key signature cytokines IFN- γ , IL-4, IL-17 and IL-22 to determine subsets which are present in Th cell clones. Cytokines were measured with ELISA in 79 stable Th cell clones following stimulation with anti-CD3/anti-CD28.

Identifying Th subsets from major cytokine secretion was regarded as a clustering task. To maximize objectivity, the aim was to find clusters that are independent to choice of clustering algorithm. **Five clustering algorithms** were applied to standardized protein secretion measurements of the Th cell clones not to bias the analysis on one method.

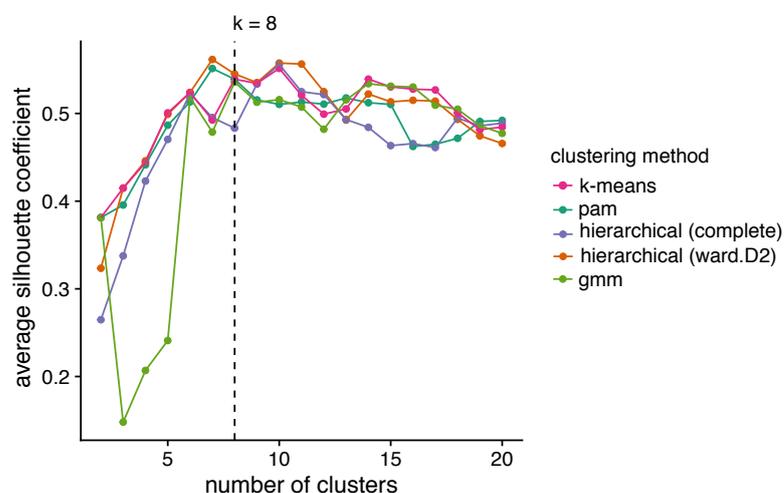


Fig. 6.2 – Determining number of clusters with average silhouette coefficients. Average silhouette coefficient for five clustering methods applied on ELISA concentrations of IFN- γ , IL-4, IL-17 and IL-22 of all Th cell clones and varying number of clusters. The best balance between minimal number of clusters, maximal agreement of the methods and maximizing the silhouette coefficients is reached for $k = 8$ clusters (dashed vertical line).

In every clustering approach the number of clusters has to be defined. Here, the average **silhouette coefficient** for two up to 20 clusters was calculated as a measure to determine

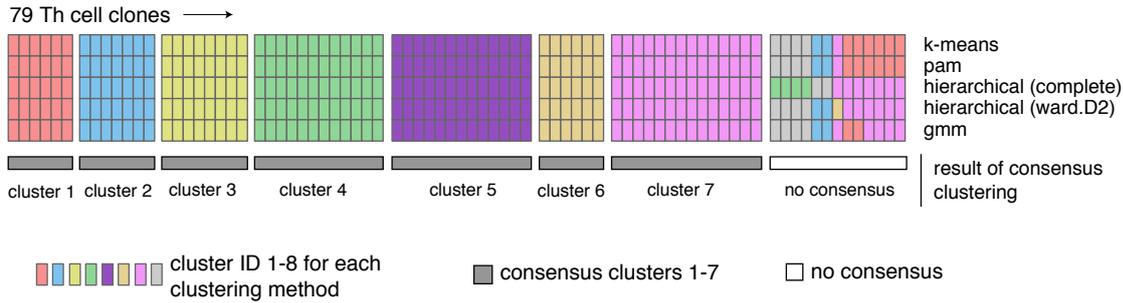


Fig. 6.3 – Clustering ELISA concentrations using $k = 8$ for five methods yielded slightly different results. Differences in clustering are visible by the cluster IDs indicated in colors. Cluster ID colors were manually matched for better visibility. Consensus clusters are those groups which were clustered together by all methods.

the optimal number of clusters. Balancing between maximal agreement of methods and maximizing average silhouette coefficient, resulted in eight as the optimal number of clusters for this data set of ELISA measurements (Fig. 6.2).

Since the clusters should not depend on the chosen clustering method, a **consensus clustering** result of five above mentioned methods and eight clusters was established (Fig. 6.3). Each clustering method yielded slightly different results, visible by the cluster IDs for each method indicated in colors. The cluster IDs were manually matched among the methods to make common clusters better visible. The interest lay in clear clusters so in the further steps only those consensus clusters were analyzed where all methods agree. Thereby clones were excluded where no consensus clustering was achieved. In total, seven consensus clusters were identified and for 13 clones no consensus was found. Note, that the approach was a strict consensus clustering as the interest was in very clear consensus clusters for further analysis.

So far secretion of IFN- γ , IL-4, IL-17 and IL-22 was used to group Th cell clones into seven consensus clusters. To assign clusters to already known Th cell subsets a closer look was taken at their individual cytokine secretion profiles (Fig. 6.4). Four consensus clusters could be assigned to known Th subset (for definition see section 6.1): cluster 1 only showed secretion of IFN- γ , representing **Th1** cells. In contrast, cells in cluster 2 secreted IL-4 but no IFN- γ , so they corresponded to **Th2** cells. **Th17** cells, secreting neither IFN- γ nor IL-4, but both IL-17 and IL-22, were cluster 3. Cluster 4 included cells which only secreted IL-22, so these cells were **Th22** cells.

Apart from the known Th subsets, three mixed subsets were found in the data set: cluster 5 comprised cells which secreted IFN- γ , IL-4 and IL-22, thus they were both or neither

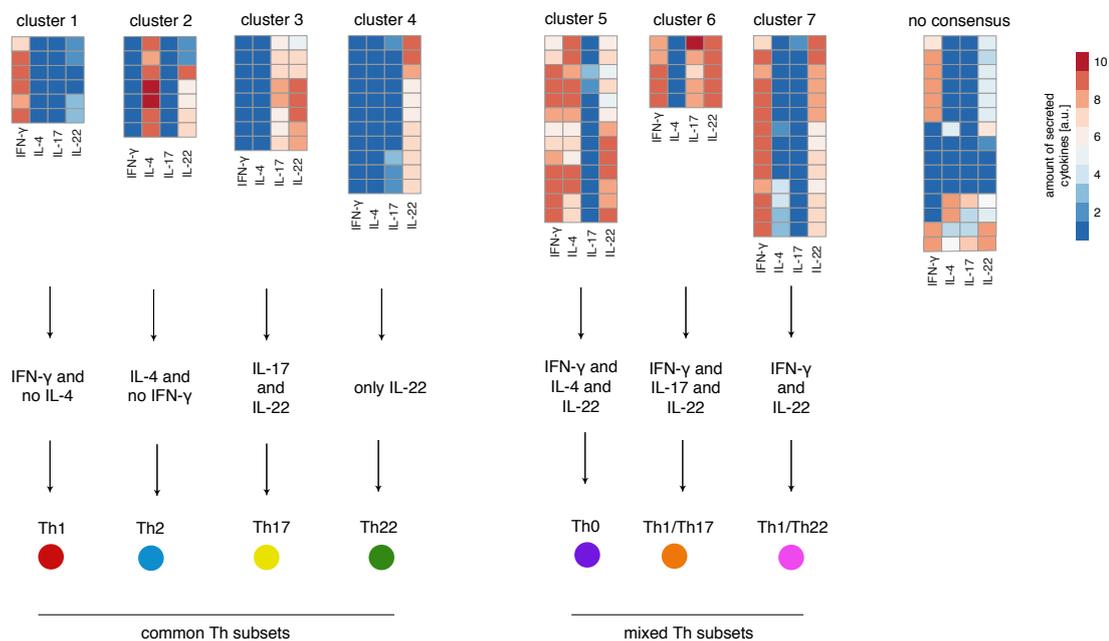


Fig. 6.4 – ELISA concentrations of IFN- γ , IL-4, IL-17 and IL-22. Concentration for all Th cell clones within each of the seven consensus clusters and remaining clones are given in scaled units transformed according to standard values. According to the definition of Th subsets based on secreted cytokines, the clusters were assigned to common Th subset and mixed subsets.

Th1, Th2 or Th22 and may reflect **Th0** cells. Cluster 6 contained cells which secreted the signature cytokine of Th1, IFN- γ , but also IL-17 and IL-22, so they were called **Th1/Th17** cells. Also cluster 7 combined signature cytokines of two major subsets: IFN- γ and IL-22. These cells were called **Th1/Th22** cells.

All three mixed subsets were already described in the literature. Th0 cells were characterized by the secretion of IL-4 and IFN- γ by Bendelac and Schwartz (1991). Miner and Croft (1998) showed that Th0 cells represent their own stable T cell subpopulation and can be derived from naive T cells. Th17 cells that also express IFN- γ were described as Th1/Th17 cells but their role and molecular phenotype is not yet fully understood (reviewed in Leung et al. (2010)). Tsanaktsi et al. (2018) detected an increase of Th1/Th17 cells in patients with active systemic lupus erythematosus. Th1/Th22 cells were described in peripheral blood of immune thrombocytopenic purpura patients (Zhan et al., 2018) and in peritumoral liver and hepatocellular carcinoma tissue (Kuang et al., 2014). In T cells isolated from pancreatic ductal adenocarcinoma patients 44 % of IL-22 producing T cell clones also secreted IFN- γ (Niccolai et al., 2016).

6.3.2 Relationship of identified Th cell subsets to known subsets and further secreted cytokines

T helper cell subsets are not only defined by their cytokine secretion profile, which were used in the previous section to define subsets, but also by transcription factors and surface molecules. Apart from the ELISA data, giving information about protein levels of signature cytokines IFN- γ , IL-4, IL-17 and IL-22, 24 additional secreted proteins were measured for a subset of 75 Th cell clones. Those were measured with the Luminex technology while three proteins were measured with both ELISA and Luminex technology (IFN- γ , IL-4 and IL-22). For all 79 Th cell clones whole-genome gene expression was measured. Since there was no access to protein measurements of transcription factors and surface markers, they were only investigated on transcriptional level.

Known transcription factors are not exclusive at transcriptional level

To confirm the identity and transcriptionally validate the common T cell subsets, the mRNA levels of previously described, subset-specific transcription factors were checked (see section 6.1). To test for differences among the Th subsets analysis of variance was performed followed by Tukey's Honest Significant Differences, which adjusted for multiple comparisons in the pairwise post-hoc comparisons (Fig. 6.5, summary in Tab. 6.2). Under the assumption that Th cells keep their specific subset identity, also when they are not stimulated, by expressing subset-specific transcription factors and surface molecules, both stimulated and unstimulated cells were analyzed for known transcription factors and surface molecules.

Th1 cells secrete IFN- γ which is regulated by the transcription factor **TBX21** (Szabo et al. (2000), Fig. 6.5 A, p-values for significant comparisons in Tab. 6.3). Expectedly, TBX21 was significantly upregulated in Th1 compared to Th2 and Th22 in stimulated cells. Th1/Th22 cells also showed a high expression of TBX21 in stimulated cells, significantly higher than Th0, Th2, Th17 and Th22 cells. As expected, Th2 showed the lowest expression of TBX21 in both stimulated and unstimulated cells, significantly lower than the remaining six subsets. Conversely, Th1 unstimulated cells were only significantly higher compared to Th2 and Th22, not for the remaining subsets. In this data set, Th1 transcription factor TBX21 was specifically downregulated in Th2 cells with all comparisons being significant and upregulated in Th1. On the basis of mean gene expression, TBX21 was similarly expressed in Th1 and Th1/Th22 cells.

6.3 RELATIONSHIP TO KNOWN SUBSETS

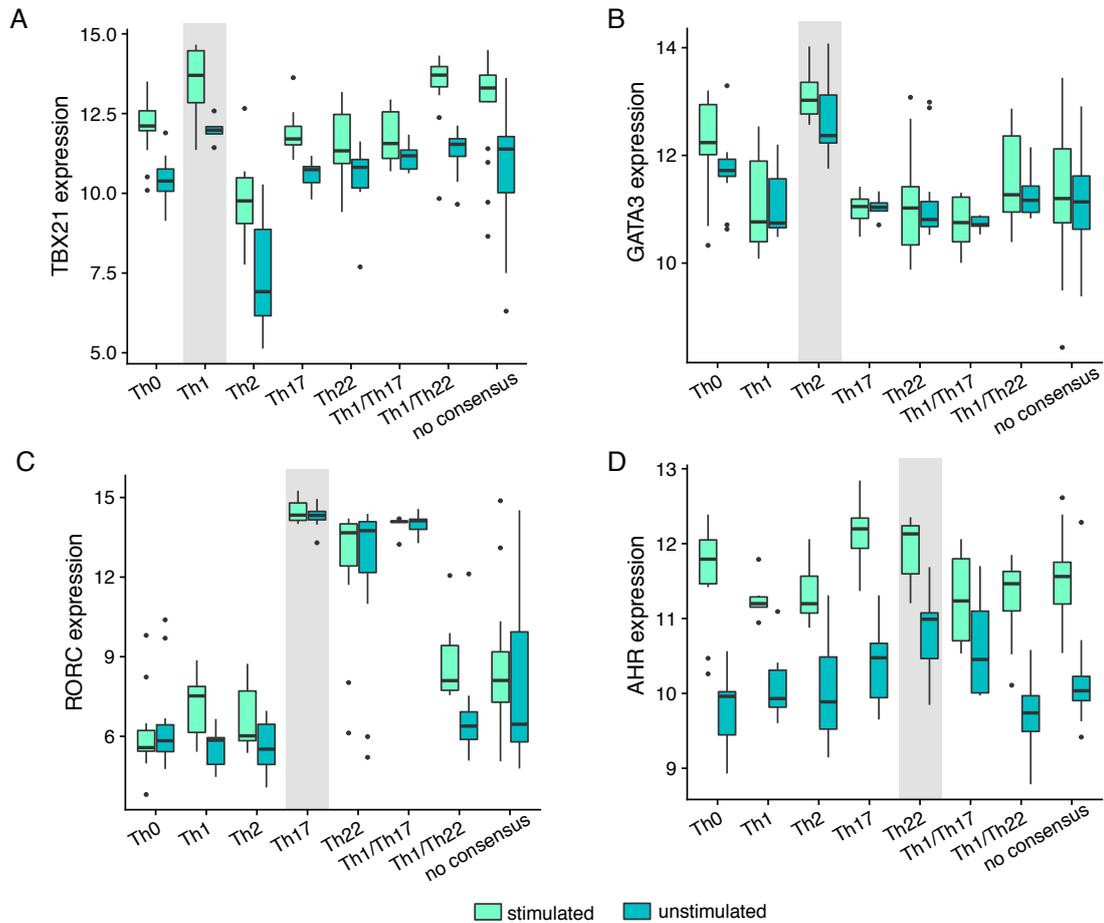


Fig. 6.5 – Gene expression levels of known transcription factors in the Th cell groups. TBX21 (A), GATA3 (B), RORC (C) and AHR (D) are thought to be the transcription factor of Th1, Th2, Th17 and Th22 cells, respectively (highlighted with grey box, high expression is expected for this subset). One representative of the two AHR probes is shown. The colors show stimulated and unstimulated cells.

IL-4 is the signature cytokine of Th2 cells and regulated by **GATA3** (Zheng and Flavell (1997), Fig. 6.5 B, p-values for significant comparisons in Tab. 6.4). Expectedly, GATA3 in stimulated cells was significantly higher expressed in Th2 cells compared to all subsets but Th0. In the unstimulated cells, GATA3 was significantly higher in Th2 than in all other subsets. Interestingly, in stimulated cells Th0 shows significantly higher expression of GATA3 compared to Th17, Th22 and Th1/Th17. In unstimulated cells this effect only remains for the comparison of Th0 and Th1/Th17. Since Th0 cells also secreted IL-4 (Fig. 6.4), it is not unexpected to see expression of Th2-specific transcription factor GATA3 known for inducing expression of IL-4.

Tab. 6.2 – Summary of transcriptional regulation of known transcription factors. Gene expression of transcription factors was compared among all identified Th subsets. Conditions were unstimulated and stimulated. Expression in Th subset named in first column is expected to be high compared to the remaining subsets for transcription factor named in second column.

Th subset	transcription factor (gene symbol)	transcriptional regulation
Th1	T-box transcription factor TBX21 (TBX21)	similar high expression levels in stimulated Th1 and Th1/Th22; lowest expression in Th2 compared to all subsets in both conditions
Th2	GATA-binding factor 3 (GATA3)	specific upregulation in Th2 in both conditions
Th17	Nuclear receptor ROR-gamma (RORC)	upregulated in Th17, Th22 and Th1/Th17 compared to other subsets in both conditions
Th22	Aryl hydrocarbon receptor (AHR)	higher expressed in unstimulated Th17 and stimulated Th22

Th17 cells are defined by secreting IL-17 which is regulated by **RORC** (Ivanov et al. (2006), Fig. 6.5 C, p-values for significant comparisons in Tab. 6.5). Expectedly, RORC was significantly higher expressed in stimulated Th17 cells compared to subsets not secreting IL-17, namely Th0, Th1, Th2 and Th1/Th22 cells. Also in the unstimulated condition Th17 expressed higher levels of RORC compared to the same subsets, Th0, Th1, Th2 and Th1/Th22. Since, Th1/Th17 cells also secreted IL-17 in the ELISA experiment, RORC was expectedly upregulated in Th1/Th17 compared to subsets not secreting IL-17 which were especially, both for stimulated and unstimulated gene expression, Th0. Interestingly, Th22 cells also showed higher expression of RORC in stimulated and unstimulated condition compared to Th0 even though Th22 cells did not secrete IL-17 in the ELISA experiment. Expectedly, stimulated Th1/Th22 cells expressed RORC significantly lower than Th17, Th1/Th17 and Th22 cells, but higher than Th0 cells.

IL-22 secretion is regulated by **AHR** and the cytokine is mainly secreted by Th17 and Th22 cells (Ramirez et al. (2010), 6.5 D, p-values for significant comparisons in Tab. 6.6). AHR is measured on the Agilent array with two different probes which showed similar behavior, for simplicity only one probe is shown and was analyzed (ID A_23.P215566). Expectedly, Th17 cells showed higher expression of AHR in stimulated cells compared to Th1 and Th2 cells. Conversely, Th17 cells also showed higher expression in stimulated cells compared to Th1/Th17 and Th1/Th22 cells which also secreted IL-22 in the ELISA. Surprisingly, Th22 cells only showed upregulation of AHR in stimulated cells compared to

6.3 RELATIONSHIP TO KNOWN SUBSETS

Th1/Th22 cells. Whereas, for unstimulated Th22 cells AHR was upregulated compared to Th0, Th2 and Th1/Th22. Despite the significant upregulation of AHR in stimulated Th17 cells, there was no regulation of AHR in unstimulated Th17 cells. However, Th1/Th17 cells showed significant upregulation of AHR in unstimulated cells compared to Th0 and Th1/Th22 cells.

Tab. 6.3 – Results of statistical comparison for Th1 specific transcription factor TBX21. Difference among Th subsets were tested with analysis of variance followed by Tukey’s Honest Significant Differences. Comparisons Tukey’s Honest Significant Differences p-value < 0.1 are shown since it adjusts for multiple testing.

Th subset pair	stimulation condition	p-value	expected?
Th1 > Th2	stimulated	$p = 1.1 \cdot 10^{-5}$	yes
Th1 > Th22	stimulated	$p = 0.030$	yes
Th1/Th22 > Th0	stimulated	$p = 0.045$	unknown
Th1/Th22 > Th2	stimulated	$p = 2.1 \cdot 10^{-7}$	unknown
Th1/Th22 > Th17	stimulated	$p = 0.071$	unknown
Th1/Th22 > Th22	stimulated	$p = 0.0035$	unknown
Th2 < Th0	stimulated	$p = 0.0027$	yes
Th2 < Th17	stimulated	$p = 0.015$	yes
Th2 < Th22	stimulated	$p = 0.046$	yes
Th2 < Th1/17	stimulated	$p = 0.087$	yes
Th1 > Th0	unstimulated	$p = 0.016$	yes
Th1 > Th2	unstimulated	$p = 7.0 \cdot 10^{-11}$	yes
Th1 > Th17	unstimulated	$p = 0.087$	yes
Th1 > Th22	unstimulated	$p = 0.030$	yes
Th1/Th22 > Th2	unstimulated	$p = 4.0 \cdot 10^{-11}$	unknown
Th2 < Th0	unstimulated	$p = 1.3 \cdot 10^{-7}$	yes
Th2 < Th17	unstimulated	$p = 3.6 \cdot 10^{-7}$	yes
Th2 < Th22	unstimulated	$p = 1.6 \cdot 10^{-7}$	yes
Th2 < Th1/17	unstimulated	$p = 1.2 \cdot 10^{-7}$	yes

Tab. 6.4 – Results of statistical comparison for Th2 specific transcription factor GATA3. Difference among Th subsets were tested with analysis of variance followed by Tukey’s Honest Significant Differences. Comparisons Tukey’s Honest Significant Differences p-value < 0.1 are shown since it adjusts for multiple testing.

Th subset pair	stimulation condition	p-value	expected?
Th0 > Th17	stimulated	p = 0.029	unknown
Th0 > Th22	stimulated	p = 0.034	unknown
Th0 > Th1/Th17	stimulated	p = 0.020	unknown
Th2 > Th1	stimulated	p = $8.4 \cdot 10^{-4}$	yes
Th2 > Th17	stimulated	p = $1.0 \cdot 10^{-4}$	yes
Th2 > Th22	stimulated	p = $9.1 \cdot 10^{-5}$	yes
Th2 > Th1/Th17	stimulated	p = $1.2 \cdot 10^{-4}$	yes
Th2 > Th1/Th22	stimulated	p = 0.0015	yes
Th0 > Th1/Th17	unstimulated	p = 0.069	unknown
Th2 > Th0	unstimulated	p = 0.031	unknown
Th2 > Th1	unstimulated	p = $7.3 \cdot 10^{-4}$	yes
Th2 > Th17	unstimulated	p = $1.1 \cdot 10^{-4}$	yes
Th2 > Th22	unstimulated	p = $1.9 \cdot 10^{-4}$	yes
Th2 > Th1/Th17	unstimulated	p = $5.0 \cdot 10^{-5}$	yes
Th2 > Th1/Th22	unstimulated	p = $3.3 \cdot 10^{-4}$	yes

6.3 RELATIONSHIP TO KNOWN SUBSETS

Tab. 6.5 – Results of statistical comparison for Th17 specific transcription factor RORC. Difference among Th subsets were tested with analysis of variance followed by Tukey’s Honest Significant Differences. Comparisons Tukey’s Honest Significant Differences p-value < 0.1 are shown since it adjusts for multiple testing.

Th subset pair	stimulation condition	p-value	expected?
Th17 > Th0	stimulated	$p = 1.1 \cdot 10^{-11}$	yes
Th17 > Th1	stimulated	$p = 3.3 \cdot 10^{-10}$	yes
Th17 > Th2	stimulated	$p = 2.3 \cdot 10^{-11}$	yes
Th17 > Th1/Th22	stimulated	$p = 7.3 \cdot 10^{-10}$	yes
Th1/Th17 > Th0	stimulated	$p = 2.1 \cdot 10^{-11}$	yes
Th1/Th17 > Th1	stimulated	$p = 8.2 \cdot 10^{-8}$	yes
Th1/Th17 > Th2	stimulated	$p = 6.0 \cdot 10^{-9}$	yes
Th1/Th17 > Th1/Th22	stimulated	$p = 8.6 \cdot 10^{-7}$	yes
Th1/Th22 > Th0	stimulated	$p = 2.0 \cdot 10^{-3}$	unknown
Th22 > Th0	stimulated	$p = 1.4 \cdot 10^{-11}$	unknown
Th22 > Th1	stimulated	$p = 6.1 \cdot 10^{-7}$	unknown
Th22 > Th2	stimulated	$p = 2.3 \cdot 10^{-8}$	unknown
Th22 > Th1/Th22	stimulated	$p = 5.6 \cdot 10^{-6}$	unknown
Th17 > Th0	unstimulated	$p = 1.8 \cdot 10^{-11}$	yes
Th17 > Th1	unstimulated	$p = 1.5 \cdot 10^{-10}$	yes
Th17 > Th2	unstimulated	$p = 5.2 \cdot 10^{-11}$	yes
Th17 > Th1/Th22	unstimulated	$p = 3.4 \cdot 10^{-11}$	yes
Th1/Th17 > Th0	unstimulated	$p = 3.8 \cdot 10^{-9}$	yes
Th1/Th17 > Th1	unstimulated	$p = 1.3 \cdot 10^{-8}$	yes
Th1/Th17 > Th2	unstimulated	$p = 5.8 \cdot 10^{-9}$	yes
Th1/Th17 > Th1/Th22	unstimulated	$p = 1.2 \cdot 10^{-8}$	yes
Th22 > Th0	unstimulated	$p = 8.8 \cdot 10^{-9}$	unknown
Th22 > Th1	unstimulated	$p = 9.9 \cdot 10^{-8}$	unknown
Th22 > Th2	unstimulated	$p = 3.2 \cdot 10^{-8}$	unknown
Th22 > Th1/Th22	unstimulated	$p = 3.7 \cdot 10^{-8}$	unknown

Tab. 6.6 – Results of statistical comparison for IL-22 associated transcription factor AHR. Difference among Th subsets were tested with analysis of variance followed by Tukey’s Honest Significant Differences. Comparisons Tukey’s Honest Significant Differences p-value < 0.1 are shown since it adjusts for multiple testing.

Th subset pair	stimulation condition	p-value	expected?
Th17 > Th1	stimulated	$p = 0.036$	yes
Th17 > Th2	stimulated	$p = 0.056$	yes
Th17 > Th1/Th17	stimulated	$p = 0.056$	unknown
Th17 > Th1/Th22	stimulated	$p = 0.0060$	unknown
Th22 > Th1/Th22	stimulated	$p = 0.030$	unknown
Th22 > Th0	unstimulated	$p = 5.4 \cdot 10^{-4}$	yes
Th1/Th17 > Th0	unstimulated	$p = 0.059$	yes
Th22 > Th2	unstimulated	$p = 0.088$	yes
Th22 > Th1/Th22	unstimulated	$p = 3.3 \cdot 10^{-4}$	unknown
Th1/Th17 > Th1/Th22	unstimulated	$p = 0.048$	unknown

A subset of known surface markers showed expected transcriptional regulation

Surface markers are commonly investigated using fluorescence-activated cell sorting (FACS) and specific combinations of markers are described for certain Th subsets (Eyerich and Zielinski, 2014). Since cell sorting data was not available, the transcriptional level was investigated. Even though surface markers are not exclusively regulated on transcriptional level and in addition downregulated when cells are taken out of the chemokine environment that specifically binds to chemokine receptors, the expression levels of known surface markers was checked in the Th cell subsets. Some show clear, anticipated patterns, others, however, do not (Fig. 6.6, summary Tab. 6.7). First an analysis of variance comparing all Th groups was performed per surface marker. If this showed a significant p-value, the appropriate post-hoc test, Tukey’s Honest Significance Differences, was applied. Results are summarized in the text and not all p-values are explicitly given.

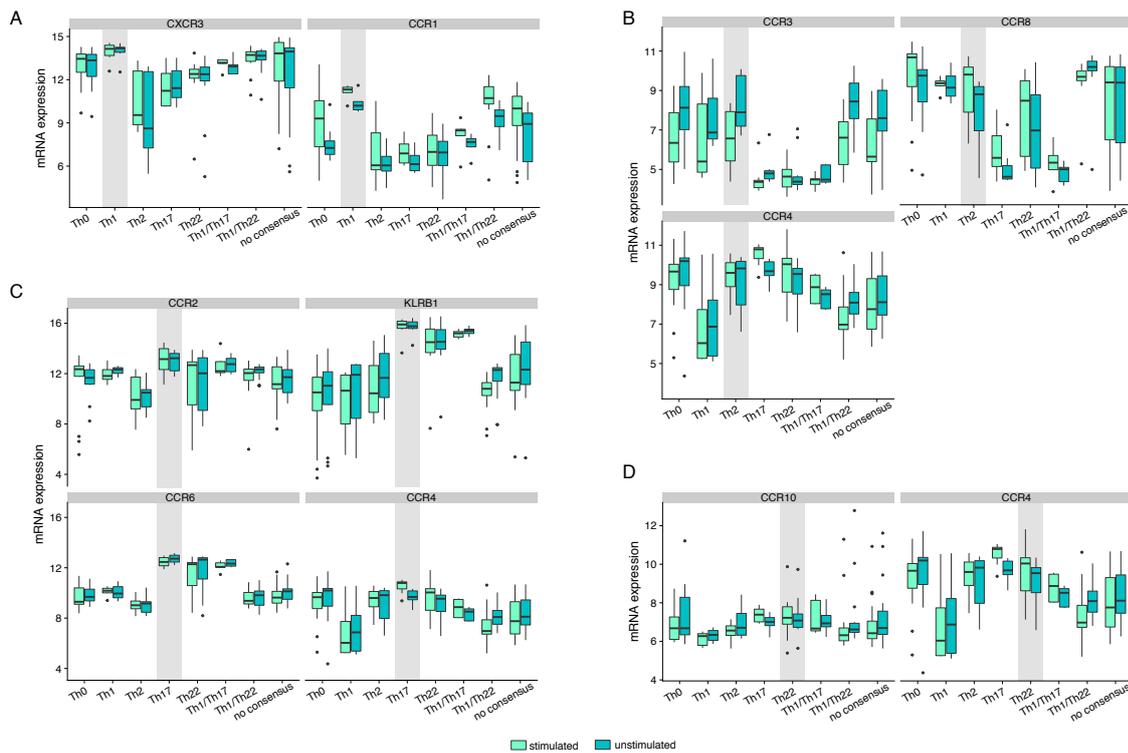


Fig. 6.6 – Transcriptional regulation of surface markers in Th cell groups. For Th1 (A), Th2 (B), Th17 (C) and Th22 (D) surface markers are described in the literature and their gene expression is visualized using boxplots (corresponding Th subset indicated in grey, high expression expected). One representative of two CXCR3 and two CCR2 probes is shown.

Two surface markers of **Th1** cells were investigated, CXCR3 and CCR1 (Fig. 6.6 A). Th1 and Th1/Th22 cells showed a similar pattern for both surface markers in comparison

6.3 RELATIONSHIP TO KNOWN SUBSETS

Tab. 6.7 – Overview of transcriptional regulation of surface molecules. Gene expression of surface molecules in all Th cell subsets were compared and results are summarized. Expression in Th subset named in first column is expected to be high compared to the remaining subsets for surface molecule named in second column.

Th subset	surface molecule (gene symbol)	transcriptional regulation
Th1	CXCR3	low in Th2
	CCR1	specifically high for Th1 and Th1/Th22
Th2	CCR3	low in Th1/Th17, Th17 and Th22; similar high levels in Th0, Th1, Th2 and Th1/Th22
	CCR8	low in Th17 and Th1/Th17; similar high levels in Th0, Th1, Th2 and Th1/Th22
	CCR4	higher in Th2 compared to Th1
Th17	CCR2	low in Th2, not Th17 specific
	CCR6, NK1.1 (=KLRB1)	low in Th0, Th1, Th2 and Th1/Th22; similar high levels in Th17, Th22 and Th1/Th17
	CCR4	higher in Th17 compared to Th1
Th22	CCR10	no differences detected
	CCR4	higher in Th22 compared to Th1

to the remaining subsets. While CXCR3 was specifically downregulated in Th2 cells, CCR1 clearly distinguished Th1 and Th1/Th22 cells from the remaining subsets on a transcriptional level.

For **Th2** cells three surface markers were looked at: CCR3, CCR8 and CCR4 (Fig. 6.6 B). CCR3 splits the Th cell clones in two groups: Th0, Th1, Th2 and Th1/Th22 showed a high expression compared to Th17, Th22 and Th1/Th17. CCR4 separated Th2 from Th1 cells. CCR8 differentiated cells which secreted IL-17 (Th17 and Th1/Th17) from those who did not, except for Th22 which showed a large spread.

Four surface markers for **Th17** cells were explored: CCR2, KLRB1 (= NK1.1), CCR6 and CCR4 (Fig. 6.6 C). CCR2 was not a specific marker for Th17 but only showed downregulation in Th2 cells. KLRB1 and CCR6 separated Th17, Th22 and Th1/Th17 cells from Th0, Th1, Th2 and Th1/Th22. CCR4 separated Th17 from Th1 cells.

CHAPTER 6 CHARACTERIZING T CELLS

For **Th22** cells two surface markers are described: CCR10 and CCR4 (Fig. 6.6 D). CCR10 is not differentially regulated on a transcriptional level in the Th subsets. For stimulated cells CCR4 separated Th22 from Th1 cells.

CCR4 is listed as a surface molecule for Th2, Th17 and Th22 cells to separate them from Th1 cells (Eyerich and Zielinski, 2014). Expectedly, CCR4 was significantly downregulated in stimulated Th1 cells compared to Th0 ($p = 0.025$), Th2 ($p = 0.032$), Th17 ($p = 2.1 \cdot 10^{-4}$) and Th22 ($p = 0.0058$). Similarly for stimulated Th1/Th22, CCR4 was significantly lower expressed in comparison to Th0 ($p = 0.034$), Th17 ($p = 1.3 \cdot 10^{-4}$) and Th22 ($p = 0.0062$) cells. CCR4 in unstimulated cells only showed a significant decrease for Th1 compared to Th0 ($p = 0.018$) and Th17 ($p = 0.021$). So CCR4 differentiated not only Th1 from Th2, Th17 and Th22 cells but also from Th0 cells.

Combining transcription factor and surface marker gene expression revealed similarity between Th1 and Th1/Th22 and between Th17 and Th1/Th17

In order to get an overall understanding of where the intermediate subtypes Th1/Th17 and Th1/Th22 were positioned among the four main subsets Th1, Th2, Th17 and Th22 when looking only at transcription factor (TF) and surface marker gene expression principal component analysis (PCA) was performed. Giving the gene expression of four subset-specific transcription factors (see Fig. 6.5) and nine surface markers (see Fig. 6.6) as input, PCA assigned in total 70% of the variance to the first three principal components (PC). PC1 (28.4% variance explained) and PC2 (25.4% variance explained) separated the Th cell clones into groups of subsets (Fig. 6.7 A). In the space spanned by PC1 and PC2 Th2 cells were separated from the remaining cells, whereas Th17 and Th1/Th17 cells were closely together. Th22 cells showed a broader spread but were in the vicinity of Th17 and Th1/Th17 cells. Th1 cells and Th1/Th22 cells clustered together. Th0 were positioned between Th1 and Th2 cells.

Variable loadings from principal component analyses explain which factors are responsible for observed patterns. Loadings of subset-specific transcription factors (TF) explained observed grouping in two dimensional space (square shapes in Fig. 6.7 C). Expression of GATA3 (Th2-specific TF) separated Th2 cells from the rest, whereas RORC (Th17-specific TF) and AHR (Th22-specific TF) grouped Th17, Th22 and Th1/Th17 cells together. TBX21 (Th1-specific TF) was located where more Th1 cells were found.

Since surface markers act in combination and not alone like transcription factors, their patterns were not as clear (oval shapes in Fig. 6.7 C). Th2-specific surface molecules CCR3

6.3 RELATIONSHIP TO KNOWN SUBSETS

and CCR8 had positive PC2 components, like all Th2 cells, but negative PC1 components. CCR2, CCR6 and KLRB1 which are Th17-specific surface molecules all had negative loadings in PC2 which explained the position of Th17 cells. However, the PC1 component varied between positive and negative values. Th22 cells all had a positive PC1 which was in line with positive PC1 loading of CCR10. CCR4, as a marker for Th2, Th17 and Th22, had a high positive PC1 loading and almost no PC2 component, which agrees perfectly with Th1 being on the negative side of PC1 and Th2, Th17 and Th22 all being on the positive side.

The third principal component (15.9% variance explained) does not reveal information about the subsets but the stimulation condition (Fig. 6.7 B). For unstimulated cells PC3 was high and for stimulated it was low. Variable loadings explained why this separation occurred (Fig. 6.7 D). AHR and TBX21 were upregulated in stimulated cells whereas CCR3, CCR2 and CCR4 were upregulated in unstimulated cells, an observation already observable in figure 6.6.

Summarizing the observations of the principal component analysis on expression levels of subset-specific transcription factors and surface molecules, similarities of mixed subsets were assigned. Th1/Th22 cells were more similar to Th1 cells than to Th22 cells on the level of transcription factor activity and surface molecule gene expression. Th0 cells lay right in between Th1 and Th2 cells. Th1/Th17 cells were more similar to Th17 cells than to Th1 cells. Th22 cells showed a broader distribution but aligned more with Th17 and Th1/Th17 than with the other subsets.

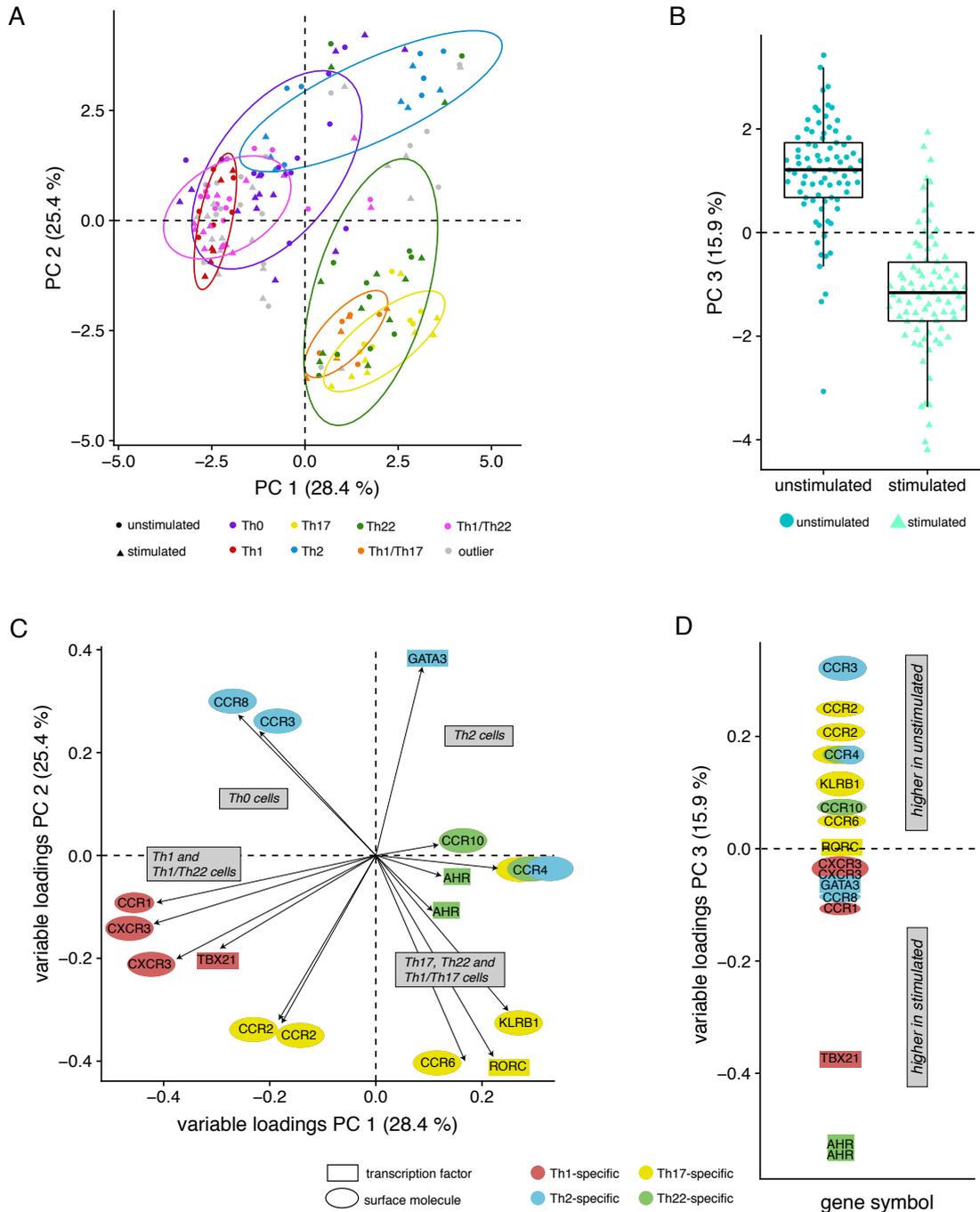


Fig. 6.7 – Principal component analysis (PCA) of gene expression of known subset-specific transcription factors and surface markers in all Th cell clones. The first two principal components (PC) split Th cell clones into groups according to their subset ((A), subsets indicated in colors, stimulation indicated with shape). The third principal component separated unstimulated from stimulated cells (B). Variable loadings are weights per variable in the new PCA space (PC1 and PC2 in (C) and PC3 in (D)) and explain why points separated in the PCA. Square shapes represent transcription factors and oval shapes surface molecules.

Protein secretion analysis validated and expanded using Luminex platform

Grouping of T cell clones into subsets was performed solely based on four cytokines whose expression was measured with ELISA like it is routinely done in laboratories. The next question was whether secretion of proteins used for clustering (IFN- γ , IL-4, IL-17 and IL-22) were validated on Luminex platform (IL-22 not available, complete Luminex not available for four out of 79 T cell clones). In general, absolute concentration values were not always comparable between both ELISA and Luminex quantification (data not shown), consequently Kruskal Wallis test followed by Dunn's post hoc test was used to test for pairwise differences of two Th subsets within each measurement technique and resulting p-values were compared.

Statistical analysis of IFN- γ , IL-4 and IL-17 measured with ELISA data showed expected results (three columns in right part of Fig. 6.8), exactly as defined by the assignment of the clusters to known Th subsets (Fig. 6.4). For example, the comparison *Th17* - *Th22* was highly significant with a positive difference (dark red) for IL-17, since IL-17 secretion in Th17 was significantly higher than in Th22 which follows the definition of Th22 cells that secrete IL-22 but not IL-17. For IL-22, which is only measured in ELISA, the overall picture for all pairwise comparisons was not as clear due to the fact that some Th2 cells and all Th0 cells in the data set also secreted IL-22 to some degree. For example, the comparison *Th2* - *Th22* showed no significant difference in IL-22. From concentration values in figure 6.4 finding a difference was not expected.

Performing the same analysis on protein secretion data measured with Luminex technology 47 of 63 (75%) pairwise comparisons within IFN- γ , IL-4 and IL-17 were in concordance with ELISA data results (Fig. 6.8). *Th0* - *Th22* in the Luminex data showed significant upregulation of IFN- γ , IL-4 and IL-17 in Th0 compared to Th22. Th0 was defined as secreting IFN- γ , IL-4 and IL-22 but no IL-17. Whereas, Th22 cells secreted only IL-22. Luminex and ELISA results agreed for IFN- γ and IL-4. IL-17 measured by ELISA in Th0 cells did not show a difference to Th22 cells. However, Luminex measurements of IL-17 did show a significantly higher secretion of IL-17 in Th0 than Th22 cells.

The remaining 24 proteins measured in Luminex all showed a significantly different pattern across seven Th cell groups (Kruskal Wallis test, FDR-corrected p-value < 0.05). The post-hoc pairwise comparison revealed stronger similarities between some groups but also groups with opposing protein secretion pattern. Th17 and Th22 only showed differentially secretion of IL-17 (as expected) and IL-6. Whereas, Th0 and Th22 differed in all of the

CHAPTER 6 CHARACTERIZING T CELLS

27 measured proteins. Where more IL-2 is secreted in Th22 cells compared to Th0, the remaining 26 cytokines have higher secretion in Th0 compared to Th22.

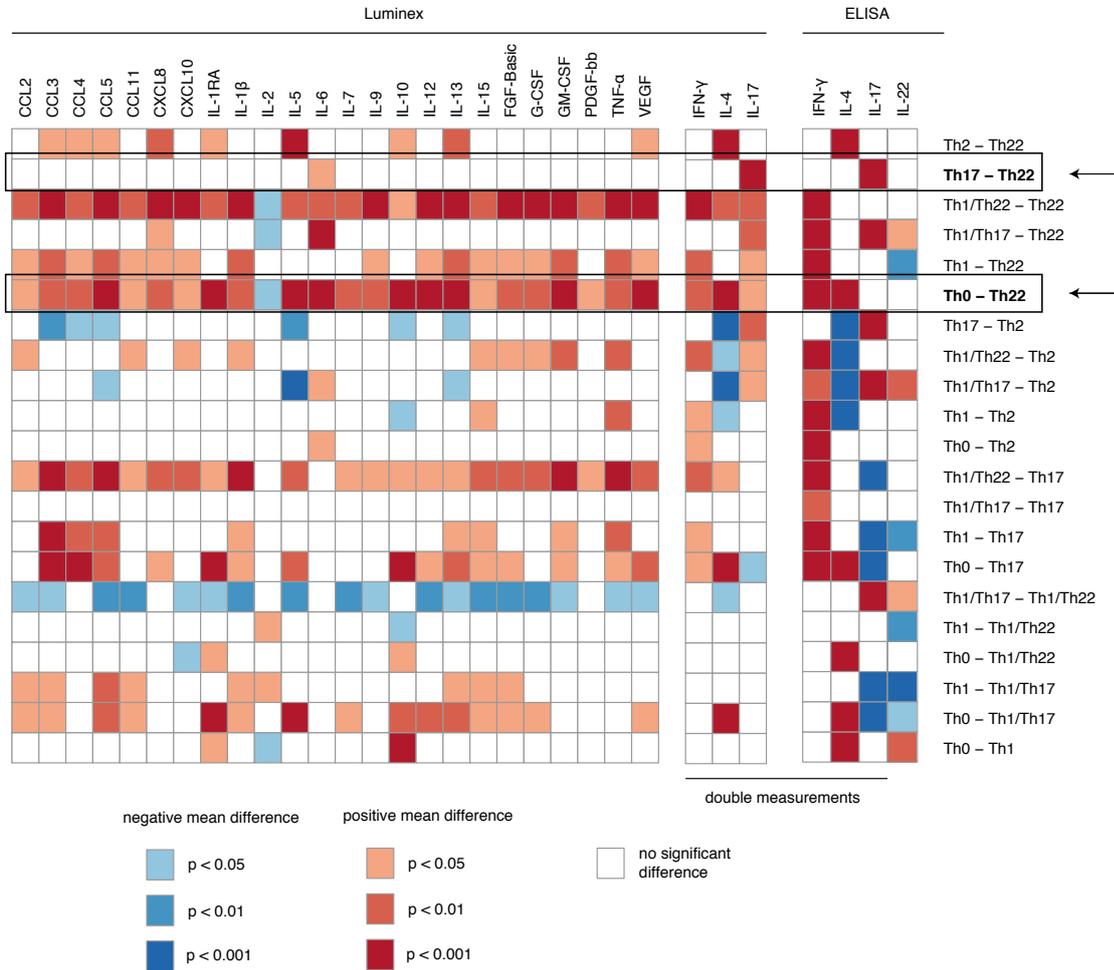


Fig. 6.8 – Difference in secreted cytokines among all pairs of the seven Th cell subsets. The ELISA measurements on the right side were used for clustering in section 6.3.1. Twenty-seven Luminex measurements gave a broader overview of the secretion patterns, IFN-γ, IL-4 and IL-17 were measured twice. Indicated in blue are negative mean differences, positive mean differences between two groups are in shades of red. The lower the significance the darker the color. The arrows indicate one comparisons with only few significant differences (upper arrow) and one with many differences (lower arrow) which are discussed in the text.

6.3.3 New transcriptional markers in Th subsets

In the previous section (6.3.2) the Th cell subsets in the data set were compared to known markers of Th cell subsets on the levels of protein secretion, transcription factor and surface molecule expression. Th cell subsets were assigned to the Th1, Th2, Th17 and Th22 phenotype and further combinations of these. In this section, the aim was to understand the phenotype of Th cell subsets further by analyzing whole genome gene expression data.

PCA showed no clear clustering of Th subsets in gene expression

To get an overview of the expression patterns principal component analysis (PCA) of whole genome expression data of all Th cell clones was performed. It showed a clear separation along the first principal component (PC), which explained 13.9% of variance, into unstimulated (lower PC1) and stimulated (higher PC1) cells (Fig. 6.9 A). A further principal component analysis separately for both conditions did not reveal clusters, also highlighting the Th subsets did not lead to visible clusters (Fig. 6.9 B for unstimulated and Fig. 6.9 C for stimulated cells). Looking further into more principal components, which explained lower variances, did not reveal a visible clustering either (data not shown).

Since most of the variance within the whole genome gene expression lay between stimulated and unstimulated cells (PCA in Fig. 6.9 A), differentially gene expression analysis was performed separately for each condition.

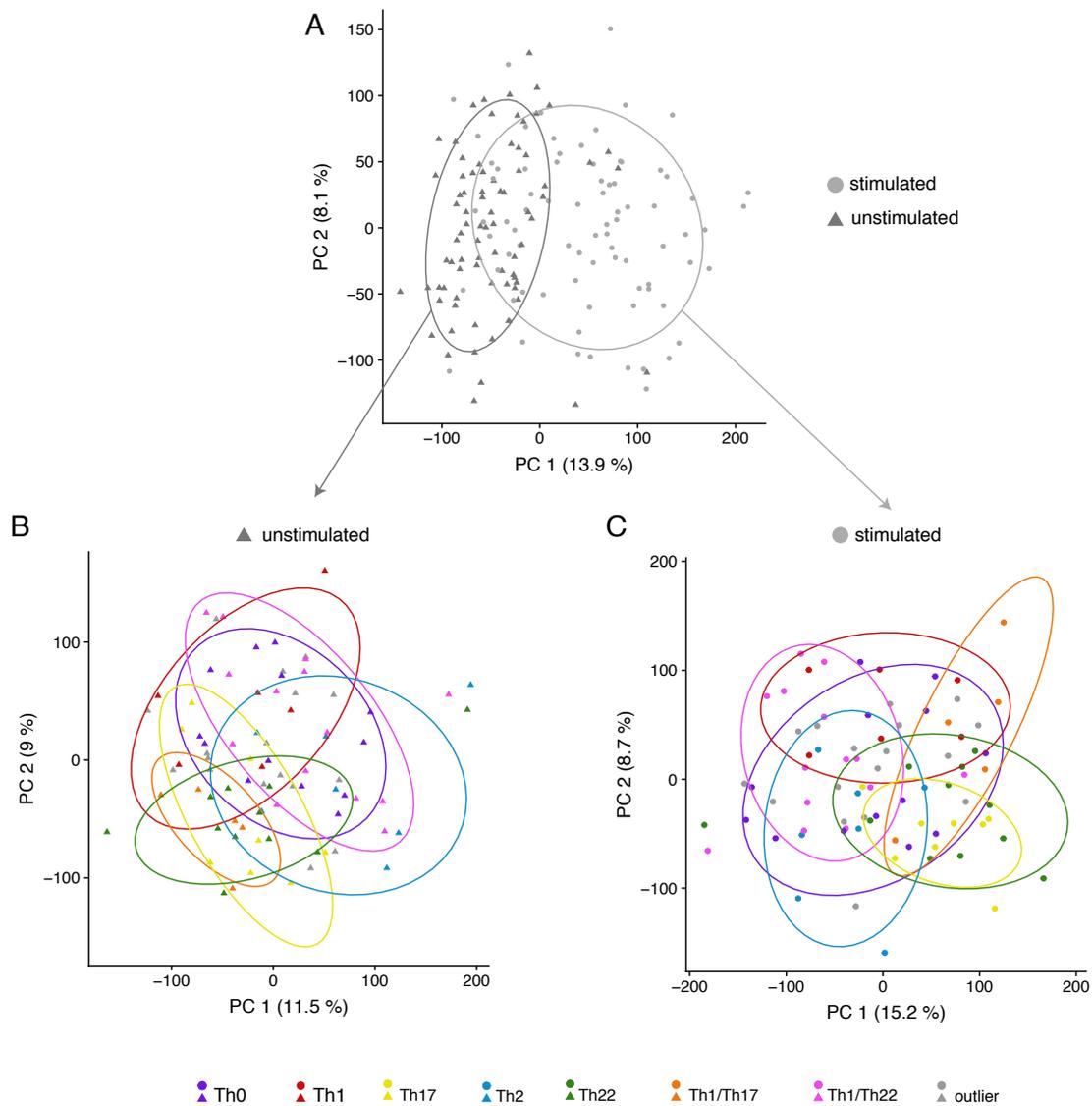


Fig. 6.9 – Principal component analysis of whole genome gene expression of Th cell clones. (A) The main difference between all clones was the stimulation condition. Within each condition ((B) for unstimulated and (C) for stimulated) the T cell subsets did not cluster. Lines represent multivariate t-distribution ellipses with 80% confidence interval.

Application of six differential gene expression methods led to different numbers of regulated genes

In order to find differentially expressed genes for each Th subset in a robust manner, six differential gene expression methods were applied to whole genome gene expression measurements from Agilent microarrays. The methods are described in detail in section 6.2 and an overview is given in table 6.1. The applied methods led to varying numbers of differentially regulated genes (Fig. 6.10). Most methods define up- and downregulated genes, except for stability selection. Stability selection was the method which determined least genes, on average around six genes per subset and condition. Multinomial grouped elastic net led to almost the same number of genes, around 170, for every subset and condition since the model searched for the same set of genes to model all seven subsets in one penalized multinomial regression model. The number still varied slightly, since the procedure was performed 1,000 times and only genes were termed “hits” if they were selected in 90% of the runs with the first quantile of all coefficients for this subset being larger than zero (upregulated genes) or the third quantile lower than zero (downregulated genes). Multinomial elastic net without grouping allows different gene sets to be selected for each subset and resulted in nine to 34 genes. Notably, for Th2 in the stimulated condition no downregulated genes were found.

The “quantile approach” follows an idea by Aran et al. (2017) and parameters had to be adjusted for the analysis of each subset separately. In short, genes were identified where the lower quantile of expression (e.g. 10% quantile) of the subset of interest was higher than the maximum equivalent higher quantile of expression (e.g. 90% quantile) of all other subsets. For every subset, condition and direction, the quantile combination (e.g. 10% and 90%) was determined where at least 20 genes satisfied the condition which lead to 28 combinations (seven subsets times two conditions times two directions). For Th0 in the stimulated, downregulated condition, 353 hits were found since loose quantiles had to be chosen, 30% as the lower quantile and 70% as the upper quantile of expression. The same quantiles led to 20 upregulated hits for Th0 in the stimulated condition.

Both approaches based on the *limma* package showed a different number of hits for the subsets. The difference in number of hits was largest for the “limma 1 vs. all” procedure with only 14 hits for Th1 but 6,392 hits for Th1/Th22. Here, one subset was compared to all other subsets as the intercept. There were 14 Th1/Th22 clones in the data set but only six Th1 clones. Which could be one reason for the differences in hits. Another could be that Th1/Th22 are more homogeneous than for example Th22 cells ($n = 11$ clones, 606 hits). To overcome the limit of differing number of clones and the implicit bias which

CHAPTER 6 CHARACTERIZING T CELLS

follows from that in differential gene expression analysis a “limma 1 vs. sampled intercept” approach was performed. In short, an intercept was sampled for every subset and condition which had the same sample size and was equally distributed among the remaining subsets. This procedure led to an increase in hits for all subsets, 1,862 hits for Th1 and 3,235 for Th22 in the stimulated condition.

Since every method used different aspects of the data to determine differentially regulated genes, leading to a varying number of hits, next the results of the methods were compared.

6.3 GENE EXPRESSION IN TH SUBSETS

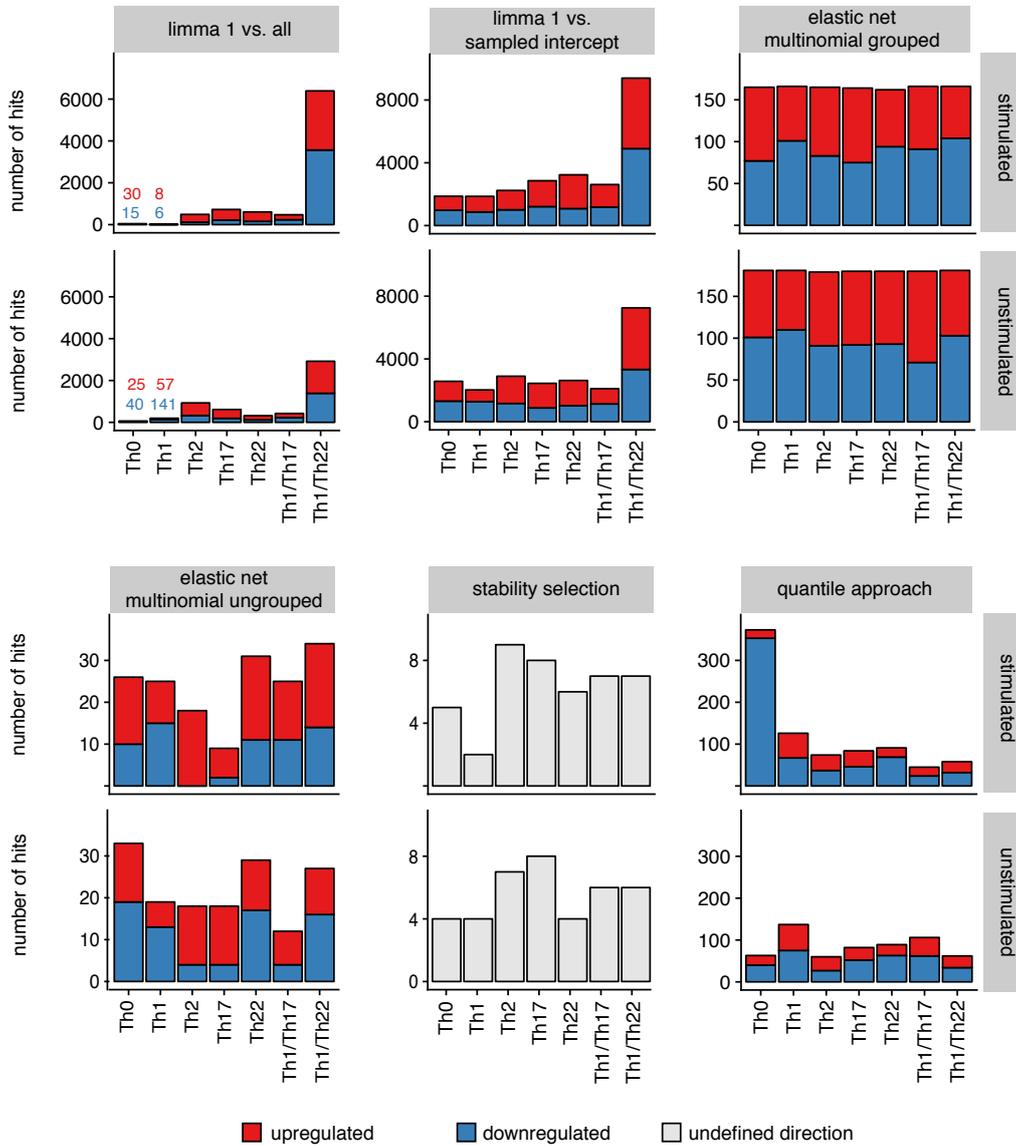


Fig. 6.10 – Results from six differential gene expression analysis methods. Methods are given in columns and stimulation conditions in rows. Most methods determine up- (red) and downregulated (blue) genes, except for stability selection (grey for undefined direction). Notably, the y-axes show number of hits but their limits vary between eight and 8,000 hits.

Top hit approach defined new subset-specific genes

The aim was to determine a small set of specific marker genes for each Th subset. Since the interest was in specific genes with the strongest signal independent of differential gene expression method choice, first the results of the six methods were made more comparable by defining 40 most differentially regulated genes (20 up- and 20 downregulated genes) per stimulation condition and subset. Then the intersection of these genes were inspected via overlap plots. Twenty-eight overlap plots were generated, for every subset, condition and direction of regulation one and here only one example is shown (Th1 unstimulated downregulated genes, Fig. 6.11).

Differentially regulated genes were defined as a **top hit** when they were detected by at least three of the six applied methods. In overlap plots, these are the genes with at least three

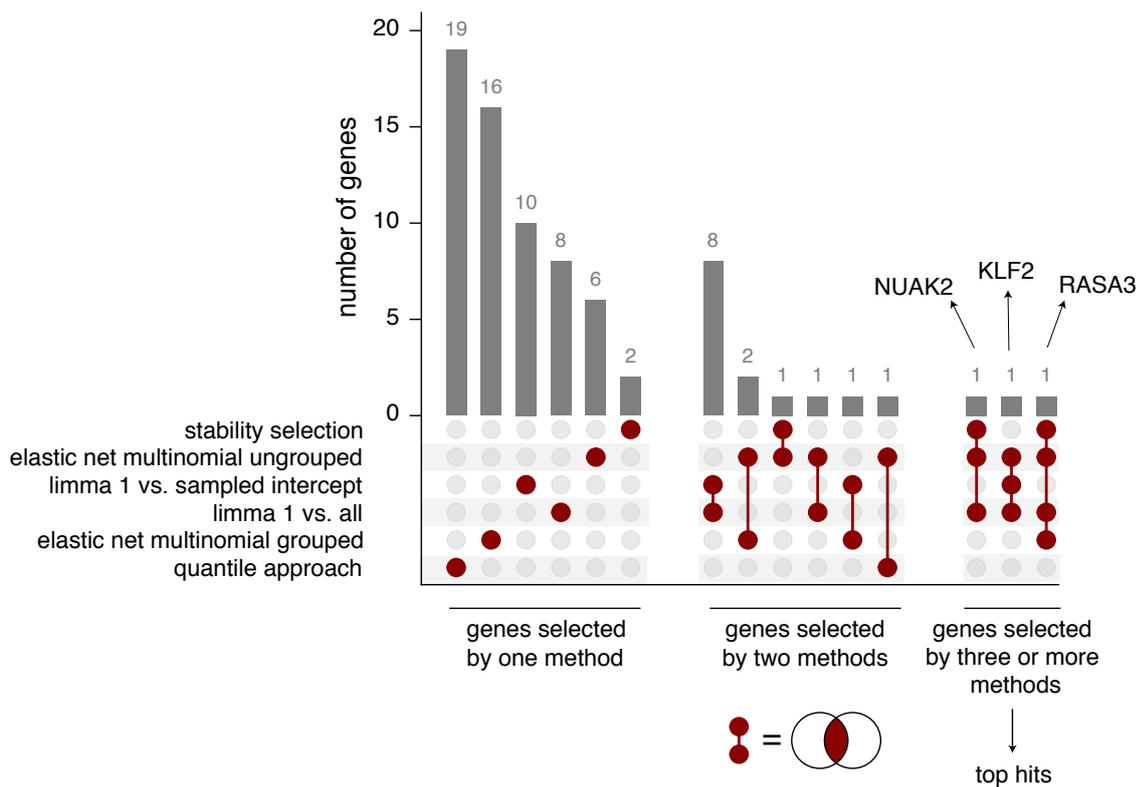


Fig. 6.11 – Overlap of six differential gene expression analysis methods results for downregulated genes in unstimulated Th1 cells. Methods are listed on the left. Single red dots and the corresponding grey bar on the top indicate how many genes were selected by just this method. Connected red dots show through the grey bar on top how many genes were commonly detected by all the connected methods. Top hits are those detected by three or more methods, in this case three genes fulfill this condition: NUAK2, KLF2 and RASA3.

6.3 GENE EXPRESSION IN TH SUBSETS

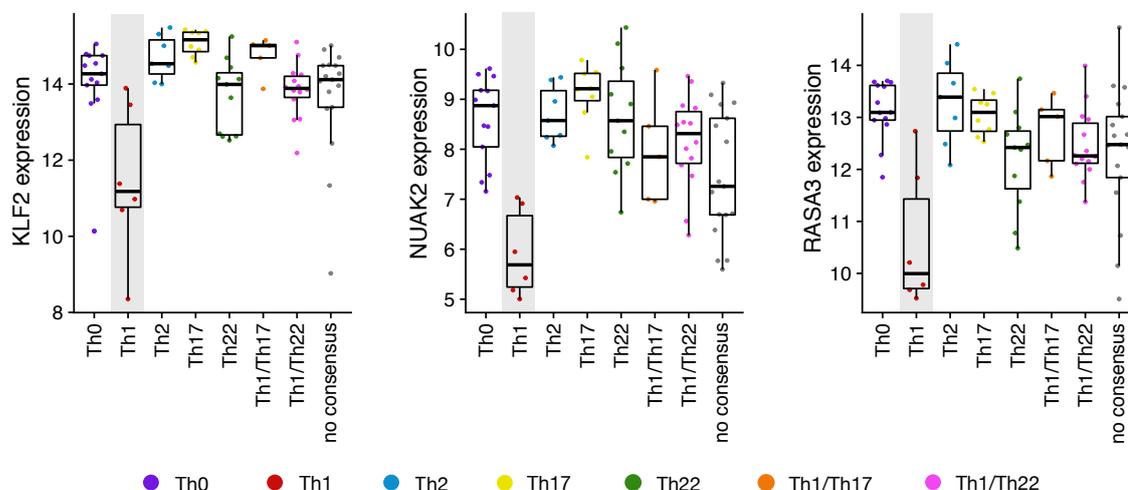


Fig. 6.12 – Gene expression comparison of Th1 specific downregulated top hits in unstimulated condition. Gene expression in all Th subsets of the three differentially downregulated genes for unstimulated Th1 cells which were detected by at least three of the applied methods, called top hits for this condition.

vertically connected dots. This revealed three top hits for Th1 unstimulated downregulated genes (Fig. 6.11), namely Kruppel Like Factor 2 (KLF2), NUA Family Kinase 2 (NUAK2) and RAS P21 Protein Activator 3 (RASA3). These top hits can further be assessed in more depth by inspecting the general gene expression patterns in all subsets via boxplots (Fig. 6.12). For all three top hits, the expression level in Th1 cells was, as expected, significantly downregulated compared to the remaining six subsets and no other comparison showed a significant difference.

Following the example of defining top hits for Th1 downregulated genes in the unstimulated condition (Fig. 6.11) using overlap plots, top hits were defined for all seven subsets, both regulation directions and stimulations by finding genes which were detected as differentially regulated by at least three of the six applied methods, resulting in 28 top hit lists.

An overview of all top hits showed different number of top hits per subset but a clear, anticipated pattern (Fig. 6.13). Downregulated top hits showed low expression (blue) for the respective subset whereas for the other subsets the gene was either not regulated (white) or upregulated (red) in scaled visualization. In contrast, upregulated top hits had high expression in the specific subset and lower expression in the other subsets. Some genes appear more than once in the plot either due to the fact that there were several probes on the microarray detecting expression of this specific gene and differential expression analysis was performed on probe-level. Another reason was that the gene was detected as a top

CHAPTER 6 CHARACTERIZING T CELLS

hit in unstimulated and stimulated cells giving even more support for the importance of this particular gene for the subset (e.g. SAMD3 in downregulated Th22 cells, two probes detected as top hit in unstimulated condition and one probe also detected in stimulated cells).

This definition of top hits was chosen since a small set of key genes important for each subset independent of the applied differential gene expression method was of interest. Starting to define top hits with more than 40 most differentially regulated genes per method, would not remove a gene from the current list but only expand it. So the identified genes (Fig. 6.13) would always be a strict subset even if the selection criteria were changed. The approach determined a list of subset-specific genes independent of statistical method and separately per direction of regulation and stimulation condition.

Even though each top hit search started per subset, stimulation condition and direction of regulation with 20 genes per method, Fig. 6.13 showed varying numbers of selected top hits. A different number of top hits was selected since at least three methods had to agree on the same gene. For example, no top hit for the Th1 specific downregulated stimulated condition was determined. Whereas, for Th2 specific upregulated and stimulated condition twelve top hits were identified. Overall, 64% of top hits were upregulated in the respective subset and 56% were identified in stimulated condition.

6.3 GENE EXPRESSION IN TH SUBSETS

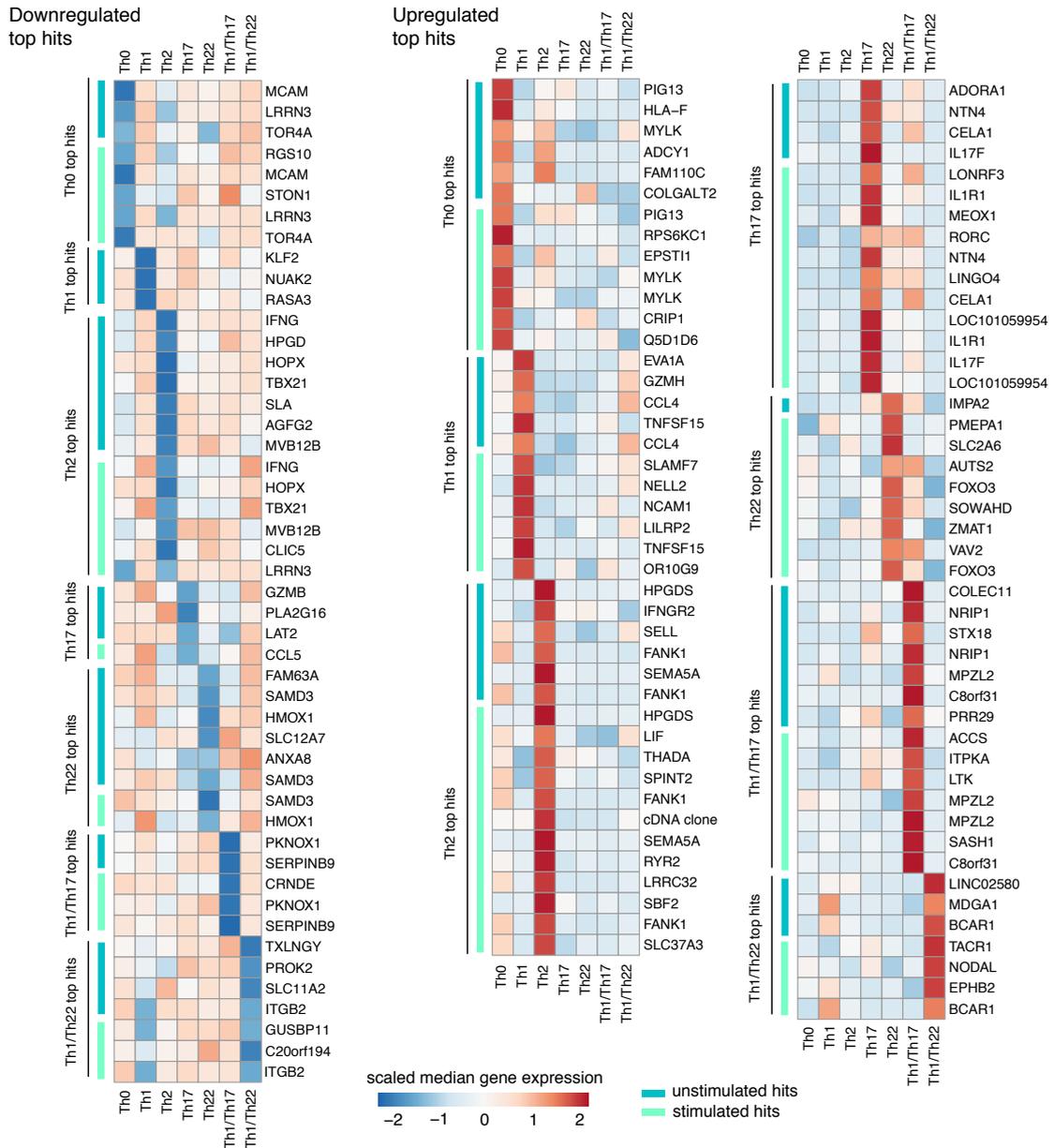


Fig. 6.13 – Heatmap of all top hits. Heatmap is split into up- (right, $n = 87$) and downregulated (left, $n = 48$) genes. Top hits are ordered by subset which they were specific for (top to bottom: Th0, Th1, Th2, Th17, Th22, Th1/Th17, Th1/Th22) first naming hits detected in unstimulated (dark blue) cells followed by stimulated (light blue) condition. Gene expression in rows are visualized as median calculated per subset and then scaled per gene (z-score). Red is high expression and blue low expression, scale is the same in both figures.

Importance of each differential expression method for top hit identification

To understand inter-relations of the six differential expression methods and to determine which one was most important for top hit identification, the results were compared from three different angles. A top hit was defined by being selected as most differentially regulated by at least three methods. Selected by exactly three methods were 64% (n = 87 of 135 total) of all top hits (Fig. 6.14 A), by four methods 26% (n = 35), by five methods 7% (n = 9) and only 3% (n = 4) of top hits were selected by all six methods. So for most of the top hits only three methods agreed.

For each method, it was analyzed separately how often that method was among the at least three methods defining a top hit. “Elastic net multinomial ungrouped” was the method which was involved most often, on average 77% of the top hits were among most differentially regulated genes for this method (Fig. 6.14 B). The “quantile approach”, however, was only important for identifying on average 43% of top hits. Interestingly, “elastic net multinomial ungrouped” was the method with second fewest hits in total compared to the remaining five methods. Only stability selection identified fewer hits overall (Fig. 6.10).

To assess the importance of each of the applied methods and their inter-relations it was compared which of the in total 135 top hits were identified by which methods. There are 20 possible combinations of choosing three methods out of six. Yet, only nine combinations were present in the 87 top hits which were identified by exactly three methods (Fig. 6.14 C). Comparing the number of genes which would have not been selected, if one method was left out in the top hit approach, revealed that “elastic net multinomial ungrouped” was the most important method, followed by “limma 1 vs. all” and “limma 1 vs. sampled intercept”. The number of genes not being selected would be 62, 58 and 46, respectively. But even leaving out the least important methods, which was either “stability selection” or the “quantile approach” with both leading to a reduction of 33 top hits, would reduce the top hit list substantially. Though, the criteria for top hits could be reduced if only five methods were to be compared. Among 48 genes which were selected by more than three methods, there were only three top hits which were selected by other methods than “elastic net multinomial ungrouped” (Fig. 6.14 C). In contrast, 26 of these 48 top hits were identified without “quantile approach”. In summary, each of the methods was almost equally important for top hit identification.

Pairwise comparison of methods identified both limma methods as most similar in their definition of most differentially regulated, since 81 top hits were identified based on the

6.3 GENE EXPRESSION IN TH SUBSETS

overlap of both limma methods plus at least one additional method. Whereas only 19 top hits were selected by overlap of “quantile approach” and “elastic net multinomial grouped” hinting at them being substantially different approaches for detecting differentially regulated genes.

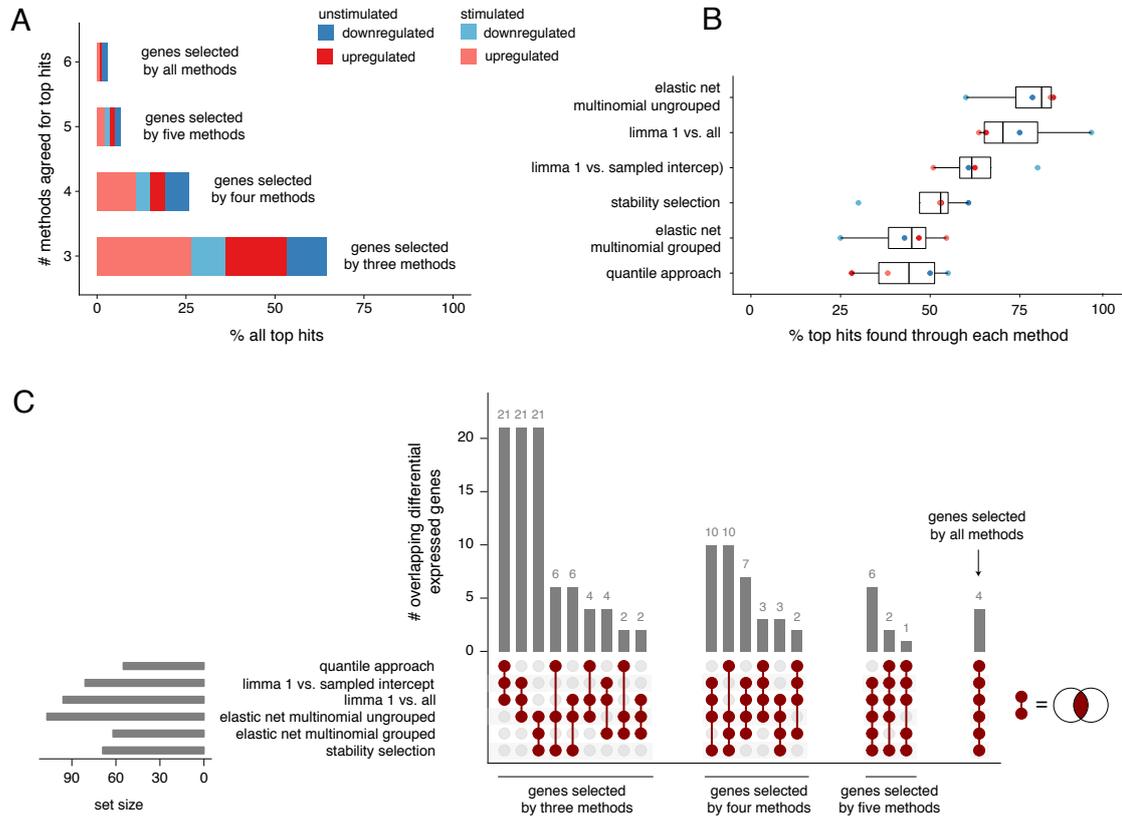


Fig. 6.14 – Comparison of importance of differential expression methods for top hit selection. At least three methods had to determine a gene as most differentially regulated, so that it was called a top hit (total n = 135). (A) Frequency of agreement of three or more methods. (B) Percentage of top hits which were initially selected by each of the methods as most differentially regulated. Normalized per stimulation condition and direction of regulation. (C) Detailed visualization of which combination of methods led to how many top hits. Connected red dots show through the grey bar on top how many top hits were detected by a combination of the connected methods.

Top hits implicated in angiogenesis and cytokine activity

To understand the molecular context of T cell subset specific top hits, pathway analysis was performed. The top hits per subset were either specifically up- or downregulated in the respective subset compared to the remaining subsets. First, top hits of each Th subset were investigated separately to determine, whether pathways were enriched (for $n = 7$ subsets each different number of genes). Next, the top hits for all subsets taken together were tested for enrichment for specific pathways ($n = 98$). Pathways in Gene Ontologies “molecular function” (Tab. 6.8) and “biological processes” were analyzed (Tab. 6.9, Gene Ontology Consortium (2016)).

Enriched molecular functions were especially processes associated to cytokine activity ($p = 0.025$) and phospholipidase activity ($p = 0.083$). Evidence was found in all top hits together, but also in Th1 (cytokine activity, $p = 0.093$) and Th17 (cytokine activity, $p = 0.043$; phospholipidase activity, $p = 0.0036$) top hits only (Tab. 6.8). Top hits for Th17, Th0 and Th1-Th22 were enriched for genes important for binding cytokines (Th17, $p = 0.043$), peptide binding (Th0, $p = 0.032$) and beta-amyloid binding (Th1-Th22, $p = 0.024$).

For the subontology “biological process” many enriched terms were detected, so only the five terms with smallest p-values are given and only one pathway for one specific set of enriched genes. For example the combination of Th1 specific genes CCL4, TNFSF15 and KLF2 is enriched in pathway “cellular response to tumor necrosis factor” ($p = 0.098$) but also in the pathway “response to tumor necrosis factor” ($p = 0.098$). Only the pathway with the smallest p-value is presented.

Angiogenesis ($p = 0.0015$) and cell migration ($p = 0.0024$) and their regulation ($p = 0.0033$ and $p = 0.014$, respectively) were enriched biological processes for all Th subset specific genes taken together (Tab. 6.9). Th1 specific genes were enriched for response to tumor necrosis factor ($p = 0.098$). Th2 specific genes were enriched for, among others, regulation of response to IFN- γ ($p = 0.059$). Even though IFN- γ is a Th1-specific secreted cytokine, it was detected as specific for Th2 cells by the analysis since Th2 cells were exclusively showing downregulation of IFN- γ compared to the remaining subsets (see Fig. 6.13).

Top hits for the Th17 subset were enriched for regulation of inflammatory response ($p = 0.082$) but also regulation of leukocyte migration ($p = 0.082$). Top hits for Th22 were enriched in lamellipodium assembly ($p = 0.073$) and also in Ras protein signal transduction ($p = 0.073$). Th0 top hits were specific for positive regulation of cell migration ($p = 0.075$) and response to abiotic or antibiotic stimulus ($p = 0.075$ and $p = 0.086$, respectively). The terms enriched in Th1/Th22 specific top hits were ameoboidal-type cell migration ($p =$

6.3 GENE EXPRESSION IN TH SUBSETS

0.015) which describes migration using extension and retraction of a pseudopodium and angiogenesis ($p = 0.020$).

Tab. 6.8 – Overview of pathway enrichment for gene ontology molecular function.
P-values are corrected for multiple testing following the Benjamini-Hochberg procedure (false discovery rate is given).

tested gene set	enriched pathway	p-value	genes in pathway
Th1 top hits	cytokine activity	0.093	CCL4, TNFSF15
Th17 top hits	phospholipidase activity	0.0036	ADORA1, PLA2G16, CCL5
	phospholipidase C activity	0.0055	ADORA1, CCL5
	cytokine binding	0.040	IL17F, IL1R1
	serine-type peptidase activity	0.043	CELA1, GZMB
	cytokine activity	0.043	IL17F, CCL5
	transcription factor activity	0.060	MEOX1, RORC
Th22 top hits	solute:cation symporter activity	0.052	SCL12A7, SLC2A6
	phosphoric ester hydrolase activity	0.064	IMPA2, HMOX1
Th0 top hits	calmodulin binding	0.032	MYLK, ADCY1
	peptide binding	0.032	HLA-F, CRIP1
Th1-Th22 top hits	beta-amyloid binding	0.024	EPHB2, ITGB2

all top hits together	cytokine activity	0.025	CCL4, TNFSF15, IFNG, LIF, IL17F, CCL5, NODAL
	CCR1 chemokine receptor binding	0.035	CCL4, CCL5
	phospholipidase activity	0.083	ADORA1, PLA2G16, CCL5, HMOX1

CHAPTER 6 CHARACTERIZING T CELLS

Tab. 6.9 – Overview of pathway enrichment for gene ontology biological processes.
P-values are corrected for multiple testing following the Benjamini-Hochberg procedure (false discovery rate is given).

tested gene set	enriched pathway	p-value	genes in pathway
Th1 top hits	cellular response to tumor necrosis factor	0.098	CCL4, TNFSF15, KLF2
Th2 top hits	regulation of calcium-transporting ATPase activity	0.043	THADA, RYR2
	positive regulation of peptidyl-serine phosphorylation of STAT protein	0.059	IFNG, LIF
	negative regulation of cell adhesion	0.059	SEMA5A, TBX21, SPINT2, LRRC32
	regulation of response to interferon-gamma	0.059	IFNGR2, IFNG
	cell differentiation involved in embryonic placenta development	0.061	LIF, SPINT2
	leukocyte cell-cell adhesion	0.0025	SELL, IFNG, TBX21, LRRC32
Th17 top hits	regulation of inflammatory response	0.082	ADORA1, IL17F, IL1R1, RORC, CCL5
	regulation of leukocyte migration	0.082	ADORA1, IL1R1, CCL5
	adipose tissue development	0.082	RORC, PLA2G16
	positive regulation of protein tyrosine kinase activity	0.082	ADORA1, CCL5
	exocrine system development	0.082	NTN4, CELA1
	connective tissue development	0.082	IL17F, RORC, PLA2G16
Th22 top hits	lamellipodium assembly	0.073	AUTS2, VAV2
	Ras protein signal transduction	0.073	SAMD3, AUTS2, VAV2
	regulation of anatomical structure size	0.073	HMOX1, SLC12A7, VAV2
	regulation of cell size	0.073	SLC12A7, VAV2
	cellular response to steroid hormone stimulus	0.080	PMEPA1, FOXO3
	signal transduction by p53 class mediator	0.082	FOXO3, ZMAT1
Th0 top hits	renal system process	0.075	ADCY1, MCAM
	positive regulation of cell migration	0.075	MYLK, FAM110C, MCAM
	cellular response to antibiotic	0.075	ADCY1, CRIP1
	cellular response to abiotic stimulus	0.086	MYLK, CRIP1
	response to antibiotic	0.087	ADCY1, CRIP1
	cellular response to drug	0.087	ADCY1, RGS10

6.3 GENE EXPRESSION IN TH SUBSETS

Tab. 6.9 – Overview of pathway enrichment for gene ontology biological processes - **continued**.

tested gene set	enriched pathway	p-value	genes in pathway
Th1-Th22 hits	top ameoboidal-type cell migration	0.015	BCAR1, ACR1, NODAL, ITGB2
	angiogenesis	0.020	NODAL, EPHB2, PROK2, ITGB2
	positive regulation of smooth muscle contraction	0.020	TACR1, PROK2
	learning or memory	0.020	TACR1, EPHB2, SLC11A2
	endodermal cell differentiation	0.020	NODAL, ITGB2
	epithelial cell migration	0.020	BCAR1, TACR1, ITGB2

all top hits together	angiogenesis	0.0015	KLF2, SEMA5A, CELA1, IL17F, HMOX1, VAV2, MCAM, SASH1, PKNOX1, NODAL, EPHB2, PROK2, ITGB2
	positive regulation of cell migration	0.0024	CCL4, SEMA5A, IFNG, IL1R1, CCL5, HMOX1, MYLK, FAM110C, MCAM, SASH1, BCAR1, TACR1
	regulation of angiogenesis	0.0033	KLF2, SEMA5A, CELA1, IL17F, HMOX1, SASH1, NODAL, PROK2, ITGB2
	positive regulation of epithelial cell migration	0.014	SEMA5A, IFNG, HMOX1, SASH1, BCAR1, TACR1
	positive regulation of reactive oxygen species biosynthetic process	0.022	KLF2, IFNG, FOXO3, ITGB2
	positive regulation of angiogenesis	0.032	SEMA5A, CELA1, HMOX1, SASH1, NODAL, ITGB2

Comparison to skin disease gene expression showed correct pattern for the majority of top hits

It is known that psoriasis is driven by an imbalance of more Th1 and Th17 cells and less Th2 cells (Guttman-Yassky et al., 2011). Whereas atopic eczema shows a higher Th2 response compared to Th1 (Guttman-Yassky et al., 2011). This set the ground truth for the comparison of Th subset top hits with gene expression patterns in skin diseases. All 48 down- and 87 upregulated top hits were tested for differences in the expression in biopsies derived from healthy, psoriasis and atopic eczema skin (ANOVA followed by post-hoc TukeyHSD) on the Agilent microarray probe-level. For 25 (18.5%) significant differences were detected among the three groups and in the post-hoc test between psoriasis and eczema ($FDR < 0.1$). For five of these, a difference was found but the gene is specific for Th0 or Th1/Th22 where the expected direction of regulation is not known. Of the remaining 20 comparisons, 14 (70%) show the right direction in the skin lesions, meaning a gene which is upregulated in Th1 subset should also be higher in psoriasis compared to atopic eczema (e.g. true for CCL4). Looking back to the genes of the example (top hits for unstimulated, downregulated Th1 cells, Fig. 6.12), KLF2, NUA2 and RASA3, for all three genes a trend towards lower expression in psoriasis skin compared to healthy and atopic eczema can be observed (Fig. 6.15), as expected since the gene is specifically downregulated in Th1 cells. KLF2 and RASA3 are significantly downregulated in psoriasis compared to atopic eczema and healthy ($FDR < 0.1$ in the post-hoc TukeyHSD).

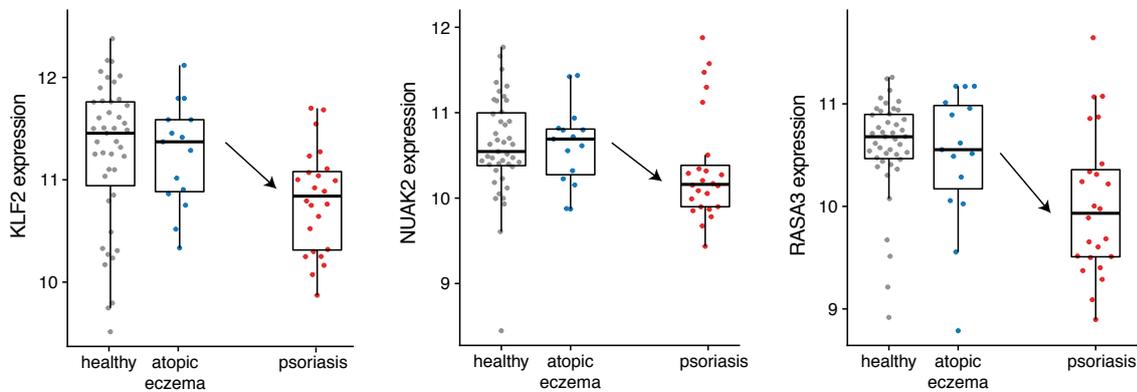


Fig. 6.15 – Gene expression in biopsies of healthy, atopic eczema and psoriasis skin of three top hits for downregulated genes for unstimulated Th1 cells. Since the genes were specifically downregulated in Th1, psoriasis skin is expected to have lower expression compared to atopic eczema and healthy (indicated by arrow).

Strong interaction of top hits in protein-protein-interaction network

To understand interconnections between subset specific top hits and identify possible transcriptional regulatory hubs, which are not regulated themselves but tightly linked, mapping of top hits to a protein-protein interaction network was performed. Protein-protein interactions between the top hits calculated using NetworkAnalyst (Xia et al., 2015) detected nine first-order subnetworks, one subnetwork where 67 of the top hits were included and eight further subnetworks with each only one of the top hits included (not shown). The subnetwork where 67 of the top hits were included is shown in figure 6.16. Stimulation condition is not coded but direction of regulation is coded through the shape (round for upregulation, square for downregulation, see Fig. 6.16). Where the terms “upregulated” and “downregulated” always refer to one subset compared to the remaining six subsets since top hits were defined as genes uniquely regulated in either direction in one of the seven Th subset compared to all others.

Only two top hits for the same subset were directly connected in the protein-protein interaction network: Interferon- γ (IFN- γ) and Interferon- γ receptor 2 (IFNGR2). Both were specific for Th2 cells but IFN- γ was downregulated in both stimulated and unstimulated cells (square shape in Fig. 6.16) and IFNGR2 was specifically upregulated in Th2 unstimulated cells compared to the remaining six subsets (round shape in Fig. 6.16).

One more direct interaction between top hits was observed but the interaction was between top hits for different subsets: intracellular protease inhibitor Serpin family B member 9 (SERPINB9, Th1/Th17 specific downregulated in both stimulated and unstimulated cells) and the cytotoxic lymphocyte protease granzyme B (GZMB, Th17 specific downregulated in unstimulated cells). GZMB is indirectly connected via four different proteins to granzyme H (GZMH, Th1 specific upregulated gene in unstimulated cells).

All other subset specific genes were always separated by at least one protein which interacts with both partners thus connects them. The network did not show clustering of top hit genes into subclusters. Still, some interesting interactions and accumulations could be observed which were investigated in more detail and discussed in section 6.4.

Early growth response 1 (EGR1) is a transcription factor which connects five subset specific top hits in the network. Arora et al. (2008) identified interaction partners of EGR1 using a chip-on-chip protocol. Three of them were identified as subset specific top hits, namely SLAM family member 7 (SLAMF7, upregulated in Th1, stimulated cells), Proliferation-inducing protein 13 (PIG13, upregulated in unstimulated Th0) and Collagen beta(1-O)galactosyltransferase 2 (COLGALT2, upregulated in unstimulated Th0). EGR1 is also connected to the transcription factor Krüppel-like factor 2 (KLF2, Th1 downregulated gene, unstimulated cells) in the network since EGR1 is known to bind the promoter region of KLF2 (Tang et al., 2010). EGR1 induction itself is also mediated by the Interleukin-1 receptor 1 (IL1R1, Th17 upregulated top hit in stimulated cells, Sells et al. (1995)).

CHAPTER 6 CHARACTERIZING T CELLS

Forkhead box O3 (FOXO3, Th22 specific upregulated in stimulated condition) induces serine-threonine protein kinase (AKT1) activity (Zhou et al., 2012). AKT1 was suggested to directly influence the localization of Family with sequence similarity 110 member C (FAM110C, Th0 specific upregulated in unstimulated condition, Hauge et al. (2009)).

Signal transducer and activator of transcription 1 (STAT1) connects three top hits for different subsets. It interacts with Th22 upregulated FOXO3 (Wu et al., 2007) and is known to regulate IFN- γ (Th2 specific downregulated in both conditions) expression via binding regulatory sites (Strengell et al., 2003). Whereas, STAT1 itself is regulated by T-box 21 (TBX21, Th2 specific downregulated in stimulated and unstimulated condition).

Cyclic AMP-responsive element-binding protein 1 (CREB1) is a transcription factor known to interact with one top hit for Th1 and one for Th2. CREB1 activates the promoter of C-C motif chemokine ligand 4 (CCL4, Th1 upregulated top hit in unstimulated cells, Mayer et al. (2013)). It is also known to bind the promoter sequence of IFN- γ (Th2 specific downregulated in stimulated and unstimulated condition, Liu et al. (2010)) and to upregulate expression of IFN- γ (Samten et al., 2005).

Vav 2 guanine nucleotide exchange factor (VAV2, Th22 specific upregulated in stimulated condition) was shown to interact with SMAD family member 3 (SMAD3, Brown et al. (2008)). SMAD3 also interacts with FOXO3 (Th22 specific upregulated in stimulated condition, Seoane et al. (2004)).

Hepatocyte nuclear factor 4-alpha (HNF4A) connects eight subset specific top hits. The transcription factor is known to regulate three of the Th1/Th17 specific upregulated top hits by binding their promoter regions (Odom et al., 2004): Syntaxin-18 (STX18, Th1/Th17 specific upregulated in unstimulated cells), Myelin protein zero like 2 (MPZL2, Th1/Th17 specific upregulated in unstimulated and stimulated cells) and 1-aminocyclopropane-1-carboxylate synthase homolog (ACCS, Th1/Th17 specific upregulated in stimulated cells). Nuclear receptor interacting protein 1 (NRIP1, Th1/Th17 specific upregulated in unstimulated cells) is a nuclear protein also interacting with HNF4A (Albers et al., 2005). HNF4A is further known to regulate the transcription of four additional T cell subset specific top hits (Odom et al., 2004), namely Major histocompatibility complex, class I, F (HLA-F, upregulated in Th0 unstimulated cells), Breast cancer anti-estrogen resistance protein 1 (BCAR1, upregulated in stimulated and unstimulated Th1/Th22 cells), NUAK family kinase 2 (NUAK2, Th1 downregulated in unstimulated cells) and Ribosomal protein S6 kinase C1 (RPS6KC1, T0 upregulated in stimulated cells).

6.3 GENE EXPRESSION IN TH SUBSETS

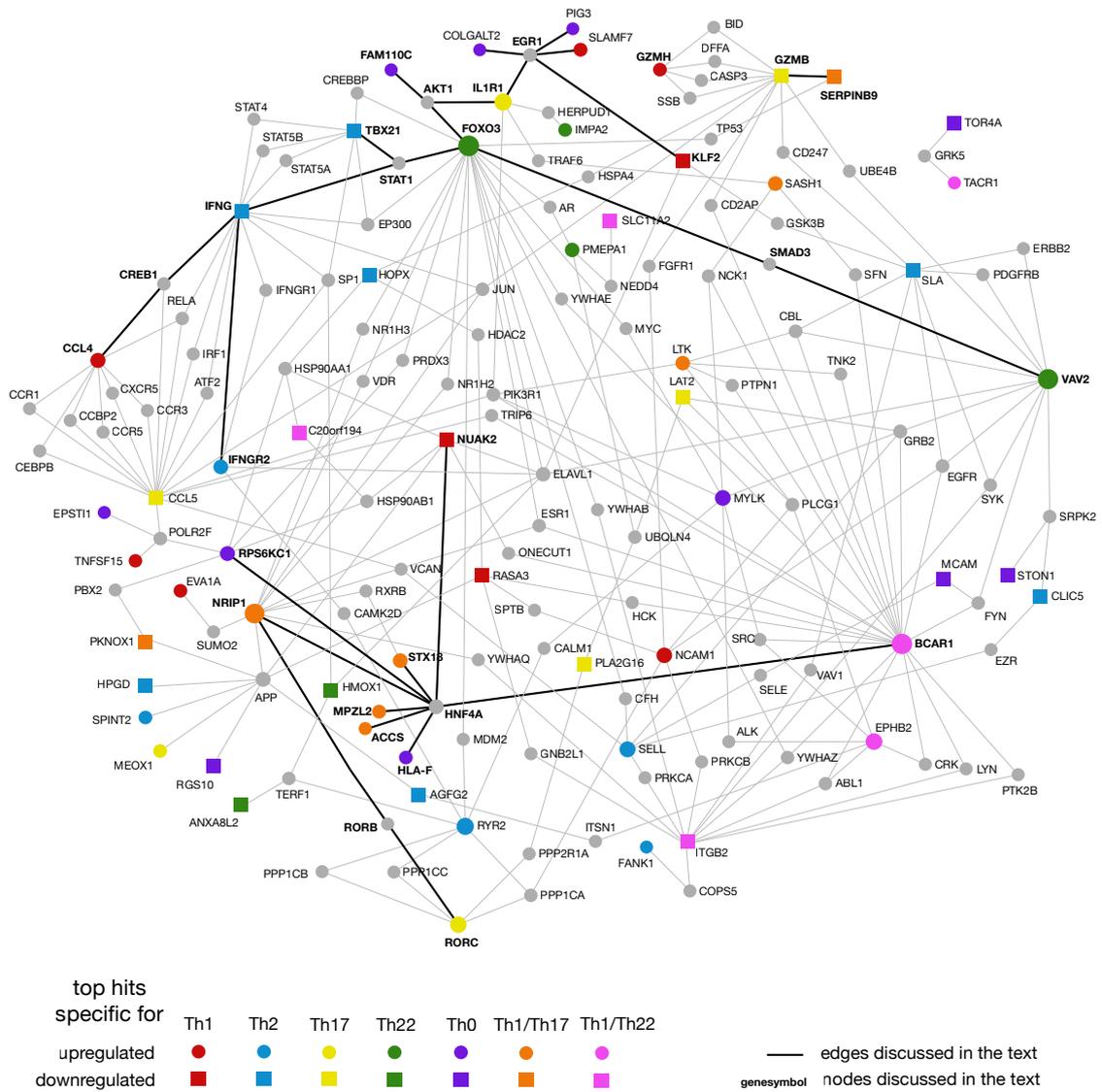


Fig. 6.16 – Network of protein-protein interactions among the top hits for Th cell subsets. Top hits in each subset are indicated with a different color. Gray nodes are proteins connecting the top hits determined by the protein-protein interaction network. Unregulated top hits are indicated in round shapes and downregulated genes in squared shapes. Stimulation condition is not given. Thick black lines and bold faced gene symbols represent proteins and their connections which are discussed in the text.

6.4 Discussion

T helper cells are important parts of the adaptive immune system. There are several different T helper cell subsets described in the literature (Eyerich and Zielinski (2014), Fig. 6.1). The list of known subsets was proposed as not being exhaustive (Zielinski, 2017) and not all subsets have yet been extensively studied. The analyzed data set consisting of 79 T cell clones had the unique property of including a large variety of T cell subsets simultaneously which were all handled and measured using the same protocols. The aim was to identify a unique set of marker genes for each T helper cell subset in a robust and unbiased way.

Instead of using conventional procedures to distinguish among T helper cell subsets a computational approach based on major cytokine secretion was proposed. The key idea was to perform consensus clustering of five clustering algorithms on ELISA measured cytokine secretion levels which revealed seven clusters in the data set (Fig. 6.3). The clusters were related to four known Th subsets and three mixed types based on major cytokine secretion patterns (Fig. 6.4). No previously unknown Th subsets were identified in this data set. Wong et al. (2016) described the same seven subsets, and additional ones, in a single cell experiment using mass cytometry.

To further identify and transcriptionally validate T cell subsets in the data set mRNA expression levels of known transcription factors and surface molecules were analyzed. Transcriptional regulation of Th1-specific transcription factor TBX21 showed similar expression in Th1 cells and the mixed type Th1/Th22 (Fig. 6.5 A). Th2-specific transcription factor GATA3 was distinctly upregulated in the cluster associated to Th2 cells (Fig. 6.5 B). RORC is described as Th17 specific transcriptional regulator (Ivanov et al., 2006). In the data set, RORC gene expression was comparable among Th17, Th22 and Th1/Th17 cells, hinting at the importance of RORC for IL-17 secretion (Fig. 6.5 C). Interestingly, Th22 upregulated RORC but only expresses IL-22 and not IL-17. The transcriptional regulation of known surface markers was less distinctive compared to transcription factors (Fig. 6.6), likely due to post-translational regulatory mechanisms. Combining mRNA information of transcription factors and surface molecules revealed clear patterns in principal component analysis which indicated that Th1/Th22 cells were more similar to Th1 than to Th22 cells and that Th1/Th17 cells were more similar to Th17 cells than to Th1 cells (Fig. 6.7 A). Distinct protein secretion patterns measured with ELISA and used for clustering were validated with the Luminex platform. Luminex measurements provided protein secretion levels of 24 additional proteins which showed interesting patterns and association between subsets (Fig. 6.8). For example, Th22 cells only differed from Th17 cells in the secretion of IL-17 and IL-6. Whereas, Th22 cells and Th0 cells differed in the secretion of every protein except for IL-22.

After validation of computationally defined T cell subsets, the aim was to identify new transcriptional markers for each subset. Th subsets have been analyzed for their transcriptional profiles before. Most studies only compared two subsets (Horiuchi et al., 2011; Boniface et al., 2010; Plank et al., 2017) and not a wide variety like the seven subsets in this study. Further, differentiated T cells were not directly isolated from involved tissues but *in vitro* differentiated from naive cells which might induce biases (Horiuchi et al., 2011; Plank et al., 2017). The holistic approach of analyzing

seven subsets for two stimulation conditions in parallel provides a unique potential for detecting novel marker genes.

Principal component analysis of whole-transcriptome data did not reveal clusters for subsets but only for stimulated and unstimulated condition (Fig. 6.9 A). So new transcriptional markers were searched for separately for unstimulated and stimulated cells. To determine subset specific marker genes six different methods for differential gene expression analysis were applied separately per subset, stimulation condition and direction of regulation since both specific up- and downregulated genes were of interest. Methods showed very different numbers of differentially expressed genes (Fig. 6.10). Since the aim was to find a small set of specific markers robust to the choice of differential analysis method, a consensus approach to define common top hits was performed. Twenty most differentially regulated genes per method, subset, direction and stimulation condition were intersected and those genes which were selected by at least three of the six applied methods were termed top hits (Fig. 6.11).

Different numbers of top hits per Th subset, stimulation condition and direction of regulation were identified (Fig. 6.13). Almost two-third of the top hits (64%) were upregulated in the respective subset and condition. Whereas similar numbers of top hits were determined for stimulated (56%) and unstimulated (46%) condition. For one combination, no top hit was identified: Th1 specific downregulated genes in the stimulated condition. Whereas, for Th2 specific upregulated and stimulated condition twelve top hits were found. The varying amount of top hits was not dependent on sample size ($n = 6$ for Th1 and $n = 7$ for Th2, see Fig. 6.4) but might rather reflect the intrinsic heterogeneity within each computationally defined subset.

A subset of the identified marker genes have been previously described for their relevance in T cell proliferation or the immune system and human diseases in general. Most research has been performed in comparing Th1 and Th2 phenotypes. There were 42,087 and 37,415 hits on PubMed for Th1 and Th2, respectively, and 23,885 for the combined search of Th1 and Th2 in February 2019 (National Institutes of Health, 2019). Th17 had 14,386 PubMed hits but in combination with Th1 only 6,524 research items were found (National Institutes of Health, 2019). Especially Th22 and the mixed Th subsets are less described in the literature (e.g. 486 hits in PubMed for Th22, 824 for Th0) so it was more challenging to associate identified genes for Th22 and mixed subsets to the literature.

Krüppel like factor 2 (KLF2) was identified as Th1 specific downregulated top hit in unstimulated cells and was previously shown to be important for coordination of CD4+ T cell differentiation through promoting the expression of trafficking receptors and lineage defining transcription factors (Lee et al., 2015). Granzyme H (GZMH) was previously suggested as a novel marker for Th1 genes (Ono et al., 2014), backing the finding of GZMH being specifically upregulated in Th1 cells.

For Th2 cells IFN- γ and T-box 21 (TBX21), among others, were identified as specifically downregulated and IFN- γ receptor 2 (IFNGR2) was identified as upregulated in Th2. IFN- γ is known to inhibit the proliferation of Th2 cells through binding the IFN- γ receptor (Gajewski and Fitch, 1988; Gajewski et al., 1989). The IFN- γ receptor itself consists of two chains, the alpha (IFNGR1) and the beta

CHAPTER 6 CHARACTERIZING T CELLS

(IFNGR2) chain (Hemmi et al., 1994; Soh et al., 1994). The beta chain (IFNGR2) was previously described to be upregulated in Th2 cells, which was also shown in this analysis, and absent in Th1 cells (Pernis et al., 1995) which explains why IFN- γ inhibits Th2 but not Th1 proliferation.

In Th17 cells, four downregulated and eleven upregulated top hits were detected. Cytotoxic lymphocyte protease granzyme B (GZMB) was specifically downregulated in Th17. Supporting this finding, GZMB was previously described as being downregulated in human Th17 cells compared to Th1 and Th1/Th17 cells (Hu et al., 2017). Interleukin 1 receptor 1 (IL-1R1) was identified as an upregulated top hit for Th17 which is supported by the finding of Engelbertsen et al. (2017) that IL-1R1 promotes Th17 immunity in CD4+ T cells in atherosclerosis. Further, Interleukin 17 (IL-17) and RAR related orphan receptor C (RORC) were identified as Th17-specific upregulated genes which are the known signature cytokine (Bettelli et al., 2006) and transcription factor (Ivanov et al., 2006) of Th17.

For Th1/Th17 cells Serpin family B member 9 (SERPINB9), among others, was identified as downregulated. SERPINB9 is known to inhibit the activity of GZMB (Sun et al., 1996). With SERPINB9 being downregulated in Th1/Th17 cells, the inhibition of GZMB is reduced and thus more GZMB can be expressed in Th1/Th17 as was described in Hu et al. (2017).

Two of the Th22 specific upregulated top hits were described for their effect in T cells. VAV proteins are important in T-cell differentiation and activation (Tybulewicz, 2005). FOXO3 is known to inhibit inflammatory cytokines from being produced by dendritic cells and to influence T cell survival (Dejean et al., 2009).

Subset specific top hits were determined using a consensus approach of six methods for differential analysis where at least three had to agree. Comparing the results in light of similarity between methods, revealed that both limma based approaches are expectedly most similar and the quantile approach was least similar to the remaining five methods (Fig. 6.14). The quantile approach adapted from (Aran et al., 2017) is purely based on distribution of values and cut-offs that have to be set manually which explains the difference to the other applied methods.

In order to set subset specific top hits into a biological context, pathway analysis was performed and top hits were compared to skin disease gene expressions. Several of the enriched molecular functions and biological processes were involved in cytokine activity and angiogenesis (Tab. 6.8 and 6.9). T cells are regulated by cytokines and secrete them, so the context given by the pathway analysis results agreed with prior expectations. Importance of T cells for angiogenesis, especially in the tumor context, was described by Freeman et al. (1995). Th1 specific genes were enriched for response to tumor necrosis factor which is an important cytokine for Th cell subset differentiation and especially promotes expression of IL-22 (Schmitt and Ueno, 2015). Genes specific for Th22 were enriched in lamellipodium assembly which are important for T cell signaling (Dustin, 2009) and also in Ras protein signal transduction which induces cell differentiation, growth and survival (Halfar et al., 2001). Comparison to gene expression skin diseases showed correct, expected patterns for 70% of the top hits (Fig. 6.15).

To understand the interconnections between T cell subset specific top hits and indicate possible validation targets, top hits were mapped to a protein-protein interaction network with particular focus on immune system relevant interactions (Breuer et al., 2012). The network revealed several interesting relationships among the identified T cell subset specific genes (Fig. 6.16). Early growth response 1 (EGR1) connects five subset specific top hits in the network. It is a zinc finger transcription factor that is upregulated in T cells after stimulation of the T cell receptor (Shao et al., 1997). Additionally, EGR1 was shown to be an upstream regulator of TBX21 expression by binding the promoter region of TBX21 (Shin et al., 2009) which is the transcription factor for Th1 cells (Szabo et al., 2000). Serine-threonine protein kinase (AKT1) activity is induced by Th22-specific Forkhead box O3 (FOXO3, Zhou et al. (2012)) and was suggested to directly influence the localization of the Th0-specific gene Family with sequence similarity 110 member C (FAM110C, Hauge et al. (2007)). AKT1 regulates important biological processes including angiogenesis (Chen et al., 2005), proliferation (Héron-Milhavet et al., 2006) and cell survival (Koseoglu et al., 2007). AKT is further involved in IL-1 signal transduction (Teshima et al., 2004). They also described an interesting connection to inducible nitric oxide synthase (iNOS), namely that IL-1R1 was required for iNOS transcription and AKT increased the transcriptional activity of iNOS gene promoter (Teshima et al., 2004). Section 3.1 showed that iNOS together with CCL27 is able to classify psoriasis from atopic eczema patients. Notably, iNOS was higher expressed in psoriasis patients. Psoriasis skin was shown to be enriched for Th17 cells (Lowe et al., 2008), thus the finding that IL-1R1 is upregulated in Th17 cells is supported by the connection of IL-1R1 to iNOS transcription via AKT activity.

The association of signal transducer and activator of transcription 1 (STAT1) with Th22-specific FOXO3 was shown to increase in IFN- γ treated cells (Wu et al., 2007). STAT1 is known to regulate IFN- γ expression (Strengell et al., 2003). Whereas, STAT1 itself is regulated by TBX21 and also GATA Binding Protein 3 (GATA3) which both bind its promoter region (Kanhere et al., 2012) and which are transcription factors for Th1 and Th2 cells, respectively (Szabo et al., 2000; Zheng and Flavell, 1997). Taken together, STAT1 is possibly in the center of Th1/Th2/Th22 phenotypes.

Cyclic AMP-responsive element-binding protein 1 (CREB1) is a transcription factor associated to both Th1 upregulated C-C motif chemokine ligand 4 (CCL4, Mayer et al. (2013)) and Th2 downregulated IFN- γ (Liu et al., 2010). It was shown that methylation in the promoter region of IFN- γ in Th2 cells inhibits binding of CREB1 to that promoters and leads to downregulation of IFN- γ specifically in Th2 cells (Williams et al., 2013). Thus giving an epigenetic explanation for IFN- γ being specifically downregulated in Th2 cells. Whereas in Th1 cells CREB1 can bind the promoter region and activate expression of IFN- γ (Samten et al., 2005; Williams et al., 2013). CREB1 was shown to regulate Th1/Th2 response in leprosy patients (Upadhyay et al., 2019). Thus, being both important for specific downregulation in Th2 cells and upregulation in Th1 cells it possibly presents an interesting candidate for further study.

SMAD family member 3 (SMAD3) was shown to interact with two Th22 upregulated genes Vav 2 guanine nucleotide exchange factor (VAV2, Brown et al. (2008)) and FOXO3 (Seoane et al., 2004). SMAD3 was described to be involved in differentiation of regulatory and inflammatory T cells,

CHAPTER 6 CHARACTERIZING T CELLS

e.g. Th17 cells (Martinez et al., 2009). In particular, SMAD3 was shown to interact with and decrease the transcriptional activity of RAR related orphan receptor C (RORC, Martinez et al. (2009)), which is the transcription factor of Th17 cells (Ivanov et al., 2006). The interaction of SMAD3 to two Th22 upregulated genes possibly indicates its importance in separating Th17 from Th22 phenotypes. RORC is also associated to the Th1/Th17 upregulated gene Nuclear receptor interacting protein 1 (NRIP1) since RORC interacts with RAR-related orphan receptor B (RORB, Wallach et al. (2013)) which is described to interact with NRIP1 (Albers et al., 2005).

Hepatocyte nuclear factor 4-alpha (HNF4A) was connected to eight T cell subset specific top hits. HNF4A was proposed as a regulatory hub in CD4+ T cells in patients suffering from systemic lupus erythematosus via genome-wide DNA methylation pattern analysis (Jeffries et al., 2011). In systemic lupus erythematosus an imbalance of T cell cytokines was previously described (Talaat et al., 2015). Especially, serum IL-17 levels are increased in these patients. Mouse studies about lupus show that inhibiting Th1 and Th17 responses in lupus-prone mice has therapeutic effects (Hou et al., 2011). These studies also show that in mice with an established lupus-like disease treated with Apigenin, that inhibits autoantigen presentation for expansion of autoreactive Th1 and Th17 cells, the disease was suppressed (Kang et al., 2009).

In summary, the analysis revealed previously described but also putative novel markers for T helper cell subsets but experimental validation is still missing. Computational consensus clustering of cells based on protein secretion measured by ELISA and not based on manual cut-offs led to robust, uniform T cell groups which were associated to known T helper cell subsets. Combining results of six differential gene expression methods uncovered a core set of T cell subset specific genes which were compared to skin disease gene expression, analyzed using pathways and protein-protein interactions to identify possible targets for experimental validation.

Chapter 7

Discussion and outlook

7.1 Discussion

Applying computational methods to answer clinical questions and gain knowledge about human health and disease was at the core of all described studies. The thesis author integrated data from different molecular omics levels with clinical data. Available tools and methods were tailored to the precise clinical question to find correct answers. Compared to the whole process involved in translational medicine, the thesis author, together with her supervisors and collaboration partners only worked on the first step: bridging the gap from laboratory work to first tests in the clinics. The range of clinical applications was broad, from describing the molecular phenotype of T helper cells as a basic unit in the immune system to finding disease classifiers and calculating signatures for inflammatory skin diseases.

The overall aim of this thesis was to improve translational medicine research and understanding of human diseases by applying statistical and bioinformatic tools to perform statistical integration of biomedical data. General challenges were identified in translational medicine research (section 1.3) and four specific research questions around these challenges were formulated (section 1.4). The thesis author together with her supervisors pursued the aim and dealt with the challenges on different levels and from different angles depending on available data and specific clinical interest of the collaborators in overall 14 studies of which eight are already published papers (see pages xi-xiii). Another subset of eight studies (five of them published) were discussed in this thesis since they jointly answer the research questions.

The first research question asked whether robust and interpretable molecular disease classifiers for unbiased patient diagnosis can be found by means of disease subtypes and independent patient cohorts. Molecular markers for differentiating psoriasis and atopic eczema (section 3.1, published in Garzorz-Stark et al. (2016)) and for diagnosing allergic asthma patients (section 3.2, published in Milger et al. (2017)) were found. The skin disease classifier's robustness was assessed using six disease subtypes and 29 of 31 patients who suffered from subtypes of psoriasis and atopic eczema were classified correctly. Further, the classifier was tested on ten clinically unclear cases and all patients were assigned to the correct diagnosis which was validated by a positive response to

CHAPTER 7 DISCUSSION AND OUTLOOK

disease-specific drug applications. For the asthma classifier study, an independent test cohort was measured which consisted of patients recruited at a different location. The classifier's robustness was evaluated by applying the trained model to the test cohort which yielded an AUC of 0.92. In addition to measuring the robustness, the aim was to find interpretable molecular markers which are understandable for clinicians and relate to their prior knowledge of the diseases. To this end, the thesis author integrated the molecular markers with patient's clinical records. For example for the skin classifier, NOS2 expression in lesional skin was shown to correlate positively with patients' BMI which is in line with the findings that NOS2 is upregulated in psoriatic lesional skin and psoriasis patients having an increased BMI (Jensen and Skov, 2016). In the asthma classifier study, an association of the molecular marker to oral corticosteroid treatment was identified. The same association was previously described in healthy patients under steroid treatment (Igaz et al., 2015).

The answer to the first research question showed one possible solution to the translational medicine challenge of basing discoveries on well designed cohorts. Since, there was no access to prospective longitudinal studies or population based cohorts, the models were tested in independent cohorts. In one study, two cohorts were available: one cohort including clinically clear cases and a cohort of special sub-diagnosis and unclear patients. Both groups were recruited at the same clinic and measured in the same laboratory (section 3.1). In another study, there were three different cohorts: exploratory measurements in mice, one cohort of patients from medicine practices and one patient cohort collected at the University Hospital Munich (section 3.2). In both studies, the established models were trained in one cohort and applied in the other to assess their performance. Apart from testing models in independent cohorts, directly visualizing selected variables and their distribution within clinical classes helped to promote comprehensibility and approval of statistical results by the collaborators. Further, associating these variables with known disease-specific factors shed light into why the molecular markers differentiated the patient groups.

The second research question asked whether serum proteins can be used as markers for disease monitoring and prognosis. The aim was to determine factors which are obtained in a minimally invasive way and which improve and standardize clinical characterization. In three studies the thesis author used statistical methods to associate serum protein measurements to clinically relevant patient outcomes. Since severity of atopic eczema is needed to monitor the disease, it was predicted from 33 serum proteins (section 4.1, published in Krause et al. (2016)). The computational model reached a leave-one-out cross validation error of 20%. The model error is too large to be clinically applicable but the model gives hints for possible markers for future research. To assess the possibility of using serum markers for prognosis, 33 serum proteins and total and specific IgE levels were analyzed to predict the persistence of atopic eczema in children (section 4.2, paper in preparation) which only gave results close to random guessing. After combining this data with clinical attributes, a test AUC of 0.68 was achieved in 10-fold cross validation averaged over twelve models. Even though no perfect predictive model was identified, a handful of factors were still determined which are currently tested for their applicability in clinics. Also serum proteins were analyzed to predict the prognosis of patients who undergo major liver resection (section 4.3, paper in preparation). Time series patterns in serum proteins were identified which were associated to clinical outcomes. Statistical models combining serum measurements and clinical attributes were calculated which

were predictive of outcome. An independent test cohort is currently measured by the collaborators and will help to evaluate the quality of the models with regard to their clinical relevance.

Even if retrospective studies, like those described in this thesis, do not fulfill the perfect requirements to answer research questions unbiasedly, they hint in the direction of what questions to ask and what samples to collect in future longitudinal cohorts. Especially, for studies in the medical field, where medical students have one or two years to collect patients and measure data for their doctoral thesis, retrospective studies are ideal if access to e.g. biobanks is given. Both retrospective and longitudinal, prospective studies together help to identify risks and intervention targets to improve human health and fight diseases. Studies described in this thesis either used snapshot data or retrospectively collected information.

The answer to the first two research questions shed light on another challenge in translational medicine research which is the characterization and monitoring of patients. Patient characterization may vary due to subjective evaluation criteria. In two studies, molecular markers for patient classification were identified (chapter 3). In three other studies, approaches to overcome subjective patient evaluation criteria by using objective markers in human serum to monitor human diseases and prognosis were proposed (chapter 4). Variability due to subjectivity during data collection was minimized through having all patients characterized by the same doctor. Whereas, histological analyses for the studies were always assessed by two independent scientists.

In contrast to clinical patient characterizations, measured molecular data is not systematically biased but only subject to controllable variability introduced by limitations of detection or natural, random fluctuations. The variability in molecular data was handled by carefully evaluating imputation methods, like GSimp (Wei et al. (2018), sections 4.2 and 4.3) and MICE (Raghunathan et al. (2001), section 4.2). Imputation methods had to be applied in order to include patients in the models who had values outside the detection limits or missing data. The evaluation was based on direct inspection and dimension reduction methods, e.g. factor analysis for mixed data. To overcome batch effects in the data, the thesis author and her supervisors assisted their collaborators in designing experiments, e.g. optimizing plate design in microarray studies, or applied batch correction methods for known batches (section 4.3) and surrogate variable analysis (section 5.2) in case of unknown batches.

The third research question asked whether inter-individual variability in patients could be adjusted for. The inter-individual variability masks common, underlying disease characteristics in complex phenotypes which is another challenge in translational medicine research of complex diseases. For the analysis of gene expression in human skin tissue described in this thesis, the thesis author adjusted for this variability using linear mixed effects models. The approach was published in two papers which involved gene expression in lesional and authologous non involved skin (section 5.1, published in Lauffer et al. (2018) and section 5.2, published in Garzorz-Stark et al. (2018)). The key benefit provided by linear mixed effects models is that every subject included in the data is modeled as a random intercept but all subjects together are used to obtain the effect sizes of the fixed effects. The random intercept adjusts for individual variability while the fixed effect is what is ultimately

CHAPTER 7 DISCUSSION AND OUTLOOK

of interest. Both publications included experimental validation of the computational results which underpinned the approach's validity. In Lauffer et al. (2018) the linear mixed effects model identified a shared type I immune response in interface diseases which was validated *in vitro* in T cells isolated from lesional skin. The other application (Garzorz-Stark et al., 2018) resulted in a proof-of-concept study where the finding that IL-23, a gene detected through computational analysis, influenced imiquimod-induced inflammation was supported in one patient using a neutralizing antibody.

The fourth research question asked whether new marker genes describing T helper cell subsets can be described to better understand their phenotypes and their function in the immune system. Phenotypes of immune cells, in particular T helper cells, are poorly understood on a deep molecular level. The aim was to describe these cells using protein secretion and gene expression data. T helper cell subsets were distinguished by applying a computational approach based on major cytokine secretion. Seven computationally defined subsets were associated to known, biological T helper cell subsets or combinations thereof. To improve the molecular understanding of different T helper cell subsets, gene expression levels were investigated. Building a consensus of six differential gene expression methods revealed a core set of specific genes for each of the subsets. Some genes were previously described, e.g. Granzyme H was previously suggested as a novel marker for Th1 genes (Ono et al., 2014) and the analysis described in this thesis also identified Granzyme H as a Th1-specific gene. Others represent novel candidates possibly useful for experimental validations like Cyclic AMP-responsive element-binding protein 1 (CREB1) which might regulate the Th1/Th2 axis. Novel marker genes for T helper cell subsets were described but their implications for immune system functions still have to be experimentally tested.

Research leads to new hypotheses, which need testing and validation. In the field of translational medicine, often ideal models of human disease are non-existent. Section 5.2 describes the approach taken by the thesis author, her supervisors and her collaborations partners on finding a human model system to study early pathogenesis of psoriasis using imiquimod. For the study described in section 5.1 no human model was available, three dimensional skin equivalents were used for validating the hypotheses. Immune cells play an important role in human health and disease. The work on improving the understanding of T helper cell subsets (chapter 6) might have implications for optimizing experimental validations and facilitating laboratory models where different cell types are combined, like in the experiments of Van Den Bogaard et al. (2014) where the crosstalk of T cells and keratinocytes in 3D skin microenvironments was described.

7.2 Outlook

Many translational medicine studies are not completed after publication but further steps are taken. The molecular classifier to differentiate between psoriasis and atopic eczema patients (section 3.1) was already proposed in an earlier study by Quaranta et al. (2014b). For the study described in (Garzorz-Stark et al., 2016), the analysis was extended, the classifier was applied to disease subtypes and the associations of molecular markers to clinical attributes were investigated. The classifier is currently tested in a large cohort of around 700 paraffin-embedded skin samples which are collected from local dermatology practices. This potentially opens the diagnostic tool to practices with no access to state-of-the-art experimental laboratories. The results gained in the study on persistence of atopic eczema in children (section 4.2) are planned to be tested in the university clinic of the collaborators. The idea is to have clinicians fill short questionnaires with parents of affected children targeting the variables identified through the performed analysis. Over the next years, the disease courses of these children will be monitored and it will be possible to evaluate the predictive power of the selected variables.

To gain deeper understanding of the interplay of different cells types within tissues, like in lesional skin of inflammatory skin disease, RNA sequencing of single cells in these tissues presents an exciting opportunity (Ramsköld et al., 2012; Shalek et al., 2013; Patel et al., 2014). Besides understanding interplay of cells, single cell RNA sequencing is also used for disease biomarkers discovery (Zhu et al., 2014) and validation (Niu et al., 2016). Tirosh et al. (2016) applied single cell RNA sequencing in skin cells of melanoma tumors to investigate the complex cellular heterogeneity in this skin disease. Der et al. (2017) also applied single cell RNA sequencing in skin cells, and in parallel renal cells, to obtain a biomarker for Lupus nephritis. Both examples prove the ability of interrogating skin diseases with single cell RNA sequencing. It thus provides a next step to improve the understanding of skin diseases.

Using serum proteins as biomarkers to diagnose and monitor human diseases is currently extensively studied but mostly in rather small local cohorts ($n = 193$ in Thijs et al. (2017a) and $n = 72$ in Krause et al. (2016)). For gene expression measurements performed with microarrays or RNA-sequencing, journals require authors to publish the raw data on databases like Gene Expression Omnibus maintained by the National Center for Biotechnology Information (NCBI's GEO, Barrett et al. (2012)). Releasing raw data enables other researchers to reproduce results or integrate the data with their own measurements. In contrast, protein secretion data is usually not published. The thesis author got into contact with Thijs et al. (2017a) right after they published their paper about severity prediction of atopic eczema. The thesis author in discussion with her supervisors wrote a proposal on how sharing their data with them would be useful to improve biomarker discovery and validation. Nevertheless, after over eighteen months, there is still no access to the data mostly due to legal issues and reasons of data security and protection. Promoting open science and encouraging scientists to share their data could substantially improve reproducibility and statistical power in finding disease biomarkers. The field of psychology has already advanced in this topic and journals

CHAPTER 7 DISCUSSION AND OUTLOOK

like “Judgment and Decision Making” require authors to publish all raw data (Society for Judgment and Decision Making, 2019).

A lot of studies in the field of translational medicine or personalized medicine use statistical methods to build disease classifiers or find biomarkers (Thijs et al., 2015b; Inkeles et al., 2015; Zissler et al., 2018). Strikingly, results of these model are commonly presented without proper assessment or validation. Thijs et al. (2015b) claimed they found a multivariate signature which “robustly predicts the [severity] score” without applying the model to a test cohort or estimating an error, e.g. through cross-validation. Only the correlation coefficient between original and predicted values was given where the same cohort was used for training and testing. So the generalizability of the proposed signature cannot be assessed. Even if researchers are unable to recruit an independent test cohort themselves, education about error estimations, generalizability and quality assessments of statistical results, would improve the objective presentation of results in translational medicine.

Apart from not presenting proper testing or validation of disease classifiers or biomarkers which are based on statistical methods, the choice of method is rarely discussed. Often, one particular mathematical method is chosen to perform the analysis without mentioning advantages or caveats. Mostly, not even reasons were given why this particular method was chosen and whether others were also tested. For example, Inkeles et al. (2015) built a multi-disease classifier for 16 different skin conditions based on publicly available microarray data. They only applied random forest as a classification method and did not give reasons for their choice. Discussing the reasoning and openly stating if different methods were checked first, would enable readers to form a scientifically sound opinion and use other researchers’ experience for their own analysis. A positive example is given by the paper Redmon and Farhadi (2018) about real-time object detection. The paper has a section called “Things We Tried That Didn’t Work” where the authors explicitly discussed several approaches which did not lead to the expected success and were discarded. Even though the paper is rather colloquially written and specifically mentions failed attempts to solve the problem, it has already been cited 242 times (according to google scholar in February 2019) indicating its acceptance by the scientific community. One idea to improve the situation is to encourage authors to explain their method choice and to elaborate on attempts which did not succeed but without forcing them to find an explanation why one method worked better than the other since that is not always easy to pinpoint.

Instead of mentioning methods which were applied but did not reveal satisfactory outcomes, in two studies results of different statistical methods were compared to obtain a core finding which is independent to method choice. In section 4.2 a core subset of commonly selected features was extracted. For section 6 a consensus clustering was performed where all applied method had to agree and an approach for combining differential gene expression methods was proposed where only a subset of methods had to be consistent. The comparisons sometimes revealed substantial differences due to method choice even if the application was to the exact same data set. The thesis author definitely did not compare among all applicable methods but even the small subset highlighted the spectrum of possible results obtained from the same data. Evaluating the differences in more detail and in a more structured way could shine light on reasons for the variability in results.

Analyzing data with several methods is not a new idea. Simonsohn et al. (2015) proposed a procedure called “specification curve analysis” aiming for application, visualization and comparison of all reasonable model specifications. They defined reasonable specifications as those which are “theoretically justified, statistically valid, and non-redundant” (Simonsohn et al., 2015). To determine all reasonable specifications, all analytical steps are considered, from exclusion of outliers, categorizing of variables, inclusion of covariates to types of regression model, which can lead to a large number of specifications and thus models to estimate (e.g. 1,728 in the example in Simonsohn et al. (2015)). The proposal to calculate specification curves instead of presenting only the result of one particular specification is discussed by the scientific community (LeBel et al., 2018; George et al., 2016) and recommended by guideline authors (Christensen and Miguel, 2018; Forstmeier et al., 2017). However, only one publication citing Simonsohn et al. (2015) actually performed a specification curve analysis: Rohrer et al. (2017) estimated between 720 and 2,160 model specifications to answer the question whether birth-order in siblings influences personality traits. The number of specifications varied since the authors investigated different outcomes. If the number of specifications is impracticably large, Simonsohn et al. (2015) proposed to use a random subset of them to generate specification curves. Currently, specification curves are mostly discussed in the psychology field (Rohrer et al., 2017; Carter et al., 2017; Milfont and Klein, 2018), probably due to the fact that the “inventors” are psychologists themselves, e.g. Uri Simonsohn is a Professor of Behavioral Science and both co-authors have a PhD in psychology. Transferring their idea to biomedical research could positively impact its reproducibility and reliability. But it should be kept in mind, to present the analysis in a way that the readers, e.g. medical doctors, are able to comprehend the analysis and the final result.

Apart from the importance of making raw data publicly available and discussing statistical method choices, sharing the underlying code of the analysis is also crucial for reproducible research. Several options for code sharing already exist, e.g. uploading the scripts to GitHub or creating executable Jupyter Notebooks. Both options require some computational expertise from the user who would like to understand or reproduce the analysis. Another way is programming a web service which creates easy access for users but computational details are often hidden and the web service itself is a work overhead which might not be justified by the actual benefit. Using R markdown (Allaire et al., 2018) and the corresponding output files in html or pdf format allows to save both source code and resulting figures but is not interactive if the users do not want to use R. An easy to use combination of making all methodological details available and reproducibility of results and figures which is open, interactive and easily usable for both computational biologists and wet lab scientists is still missing.

In the experience of the thesis author, her supervisors and collaborators, during the review process of clinical journals the choice of modeling or statistical evaluation is rarely questioned by editors or reviewers. Making everyone more aware of how choice of statistical method influences results, how biomarker signature models can be assessed and how important publishing raw data and the code used for analysis is, would improve the reproducibility in translational medicine research and thus ultimately bring benefit to the patients. The awareness could be raised by identifying a way to bring across the pitfalls of misusing statistics in a clear and concise way to the collaborators by

CHAPTER 7 DISCUSSION AND OUTLOOK

e.g. simulations or explicit, relatable examples. Only if everyone involved in the process trusts the procedure used to infer knowledge from data, the gained knowledge will be trusted and used to everyone's benefit.

Bibliography

- Mashallah Aghilinejad, Elahe Kabir-Mokamelkhah, Zahra Imanizade, and Hossein Danesh. Occupational class groups as a risk factor for gastrointestinal cancer: A case-control study. *The international journal of occupational and environmental medicine*, 8(1 January):851–21, 2017.
- Hirotsugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- Michael Albers, Harald Kranz, Ingo Kober, Carmen Kaiser, Martin Klink, Jörg Suckow, Rainer Kern, and Manfred Koegl. Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. *Molecular & Cellular Proteomics*, 4(2):205–213, 2005.
- SS Alessi, Marcello Menta Simonsen Nico, Juliana Dumet Fernandes, and Silvia Vanessa Lourenço. Reflectance confocal microscopy as a new tool in the in vivo evaluation of desquamative gingivitis: patterns in mucous membrane pemphigoid, pemphigus vulgaris and oral lichen planus. *British Journal of Dermatology*, 168(2):257–264, 2013.
- JJ Allaire, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, and Winston Chang. *rmarkdown: Dynamic Documents for R*, 2018. URL <https://CRAN.R-project.org/package=rmarkdown>. R package version 1.10.
- André Altmann, Laura Tološi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010.
- Stephen F Altschul, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.
- Ronald M Andersen. Revisiting the behavioral model and access to medical care: does it matter? *Journal of health and social behavior*, pages 1–10, 1995.
- N Leigh Anderson. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clinical chemistry*, 56(2):177–185, 2010.
- Chittaranjan Andrade. Understanding relative risk, odds ratio, and related terms: as simple as it can get. *The Journal of clinical psychiatry*, 76(7):857–861, 2015.

BIBLIOGRAPHY

- Ilias Angelidis, Lukas M Simon, Isis E Fernandez, Maximilian Strunz, Christoph H Mayr, Flavia R Greiffo, George Tsitsiridis, Meshal Ansari, Elisabeth Graf, Tim-Matthias Strom, et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nature communications*, 10(1):963, 2019.
- Dvir Aran, Zicheng Hu, and Atul J Butte. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome biology*, 18(1):220, 2017.
- Shilpi Arora, Yipeng Wang, Zhenyu Jia, Saynur Vardar-Sengul, Ayla Munawar, Kutbuddin S Doctor, Michael Birrer, Michael McClelland, Eileen Adamson, and Dan Mercola. Egr1 regulates the coordinated expression of numerous EGF receptor target genes as identified by ChIP-on-chip. *Genome biology*, 9(11):R166, 2008.
- Meriem Attaf, Eric Huseby, and Andrew K Sewell. $\alpha\beta$ T cell receptors as predictors of health and disease. *Cellular & molecular immunology*, 12(4):391, 2015.
- Ovgu Aydin, Burhan Engin, Oya Oğuz, Şennur İlvan, and Cuyan Demirkesen. Non-pustular palmoplantar psoriasis: is histologic differentiation from eczematous dermatitis possible? *Journal of cutaneous pathology*, 35(2):169–173, 2008.
- Tanya Barrett, Stephen E Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F Kim, Maxim Tomashevsky, Kimberly A Marshall, Katherine H Phillippy, Patti M Sherman, Michelle Holko, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*, 41(D1):D991–D995, 2012.
- Cristina Bascones-Ilundain, Miguel Angel Gonzalez-Moles, German Esparza-Gomez, Jose Antonio Gil-Montoya, and Antonio Bascones-Martinez. Importance of apoptotic mechanisms in inflammatory infiltrate of oral lichen planus lesions. *Anticancer research*, 26(1A):357–362, 2006.
- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles*, 67(1):1–48, 2015. ISSN 1548-7660. doi: 10.18637/jss.v067.i01. URL <https://www.jstatsoft.org/v067/i01>.
- Sebastian Bauer, Julien Gagneur, and Peter N Robinson. GOing Bayesian: model-based gene set analysis of genome-scale data. *Nucleic acids research*, 38(11):3523–3532, 2010.
- Etienne Becht, Charles-Antoine Dutertre, Immanuel WH Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W Newell. Evaluation of umap as an alternative to t-sne for single-cell data. *bioRxiv*, page 298430, 2018.
- Lisa A. Beck, Diamant Thaçi, Jennifer D. Hamilton, Neil M. Graham, Thomas Bieber, Ross Rocklin, Jeffrey E. Ming, Haobo Ren, Richard Kao, Eric Simpson, Marius Ardeleanu, Steven P. Weinstein, Gianluca Pirozzi, Emma Guttman-Yassky, Mayte Suárez-Fariñas, Melissa D. Hager, Neil Stahl, George D. Yancopoulos, and Allen R. Radin. Dupilumab Treatment in Adults with

BIBLIOGRAPHY

- Moderate-to-Severe Atopic Dermatitis. *New England Journal of Medicine*, 371(2):130–139, jul 2014. doi: 10.1056/nejmoa1314768. URL <https://doi.org/10.1056/nejmoa1314768>.
- Albert Bendelac and Ronald H Schwartz. Th0 cells in the thymus: the question of T-helper lineages. *Immunological reviews*, 123:169–188, 1991.
- Elisa Benedetti, Maja Pučić-Baković, Toma Keser, Annika Wahl, Antti Hassinen, Jeong-Yeh Yang, Lin Liu, Irena Trbojević-Akmačić, Genadij Razdorov, Jerko Štambuk, et al. Network inference from glycoproteomics data reveals new reactions in the igg glycosylation pathway. *Nature communications*, 8(1):1483, 2017.
- 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)*, pages 289–300, 1995.
- Salah-Eddine Bentebibel, Nathalie Schmitt, Jacques Banchereau, and Hideki Ueno. Human tonsil B-cell lymphoma 6 (BCL6)-expressing CD4+ T-cell subset specialized for B-cell help outside germinal centers. *Proceedings of the National Academy of Sciences*, 108(33):E488–E497, 2011.
- Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.
- Klea D Bertakis. The influence of gender on the doctor–patient interaction. *Patient education and counseling*, 76(3):356–360, 2009.
- Estelle Bettelli, Yijun Carrier, Wenda Gao, Thomas Korn, Terry B Strom, Mohamed Oukka, Howard L Weiner, and Vijay K Kuchroo. Reciprocal developmental pathways for the generation of pathogenic effector T H 17 and regulatory T cells. *Nature*, 441(7090):235, 2006.
- DE Bidwell, Ann Bartlett, and A Voller. Enzyme immunoassays for viral diseases. *Journal of Infectious Diseases*, 136(Supplement 2):S274–S278, 1977.
- Thomas Bieber. Atopic Dermatitis. *New England Journal of Medicine*, 358(14):1483–1494, apr 2008. doi: 10.1056/nejmra074081. URL <https://doi.org/10.1056/nejmra074081>.
- Chester Ittner Bliss. The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22(1):134–167, 1935.
- Wolf-Henning Boehncke and Michael P Schön. Psoriasis. *The Lancet*, 386(9997):983 – 994, 2015. ISSN 0140-6736.
- Karsten Boehnke, Philip W Iversen, Dirk Schumacher, María José Lallena, Rubén Haro, Joaquín Amat, Johannes Haybaeck, Sandra Liebs, Martin Lange, Reinhold Schäfer, et al. Assay establishment and validation of a high-throughput screening platform for three-dimensional patient-derived colon cancer organoid cultures. *Journal of biomolecular screening*, 21(9):931–941, 2016.

BIBLIOGRAPHY

- Benjamin M Bolstad, Rafael A Irizarry, Magnus Åstrand, and Terence P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–193, 2003.
- Katia Boniface, Wendy M Blumenschein, Katherine Brovont-Porth, Mandy J McGeachy, Beth Basham, Bela Desai, Robert Pierce, Terrill K McClanahan, Svetlana Sadekova, and René de Waal Malefyt. Human th17 cells comprise heterogeneous subsets including ifn- γ -producing cells with distinct properties from the th1 lineage. *The Journal of Immunology*, 185(1):679–687, 2010.
- Jan C Brase, Daniela Wuttig, Ruprecht Kuner, and Holger Sültmann. Serum microRNAs as non-invasive biomarkers for cancer. *Molecular cancer*, 9(1):306, 2010.
- Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- Dagmar Breitfeld, Lars Ohl, Elisabeth Kremmer, Joachim Ellwart, Federica Sallusto, Martin Lipp, and Reinhold Förster. Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. *Journal of Experimental Medicine*, 192(11):1545–1552, 2000.
- Karin Breuer, Amir K Foroushani, Matthew R Laird, Carol Chen, Anastasia Sribnaia, Raymond Lo, Geoffrey L Winsor, Robert EW Hancock, Fiona SL Brinkman, and David J Lynn. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic acids research*, 41(D1):D1228–D1233, 2012.
- Kimberly A Brown, Amy-Joan L Ham, Cara N Clark, Nahum Meller, Brian K Law, Anna Chytil, Nikki Cheng, Jennifer A Pietenpol, and Harold L Moses. Identification of novel Smad2 and Smad3 associated proteins in response to TGF- β 1. *Journal of cellular biochemistry*, 105(2):596–611, 2008.
- Daniela Bruch-Gerharz, Karin Fehsel, Christoph Suschek, Guenter Michel, Thomas Ruzicka, and Victoria Kolb-Bachofen. A proinflammatory activity of interleukin 8 in human skin: expression of the inducible nitric oxide synthase in psoriatic lesions and cultured keratinocytes. *Journal of Experimental Medicine*, 184(5):2007–2012, 1996.
- Dunja Bruder, Michael Probst-Kepper, Astrid M Westendorf, Robert Geffers, Stefan Beissert, Karin Loser, Harald von Boehmer, Jan Buer, and Wiebke Hansen. Frontline: Neuropilin-1: a surface marker of regulatory T cells. *European journal of immunology*, 34(3):623–630, 2004.
- Patrick M Brunner, Mayte Suárez-Fariñas, Helen He, Kunal Malik, Huei-Chi Wen, Juana Gonzalez, Tom Chih-Chieh Chan, Yeriel Estrada, Xiuzhong Zheng, Saakshi Khattri, et al. The atopic dermatitis blood signature is characterized by increases in inflammatory and cardiovascular risk proteins. *Scientific reports*, 7(1):8707, 2017.

BIBLIOGRAPHY

- Roger Bumgarner. Overview of dna microarrays: types, applications, and their future. *Current protocols in molecular biology*, 101(1):22–1, 2013.
- Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O’Connell, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203, 2018.
- Evan Carter, Felix Schönbrodt, Will M Gervais, and Joseph Hilgard. Correcting for bias in psychology: A comparison of meta-analytic methods. 2017.
- Centers for Disease Control and Prevention. Improving public health practice through translation research (R18). Request for application, 2007. URL <https://grants.nih.gov/grants/guide/rfa-files/rfa-cd-07-005.html>. [Online; accessed 30-October-2018].
- John M Chambers, Trevor J Hastie, et al. *Statistical models in S*, volume 251. Wadsworth & Brooks/Cole Advanced Books & Software Pacific Grove, CA, 1992.
- Hua-Chen Chang, Sarita Sehra, Ritobrata Goswami, Weiguo Yao, Qing Yu, Gretta L Stritesky, Rukhsana Jabeen, Carl McKinley, Ayele-Nati Ahyi, Ling Han, et al. The transcription factor PU. 1 is required for the development of IL-9-producing T cells and allergic inflammation. *Nature immunology*, 11(6):527, 2010.
- Andrew Chatr-Aryamontri, Rose Oughtred, Lorrie Boucher, Jennifer Rust, Christie Chang, Nadine K Kolas, Lara O’Donnell, Sara Oster, Chandra Theesfeld, Adnane Sellam, et al. The biogrid interaction database: 2017 update. *Nucleic acids research*, 45(D1):D369–D379, 2017.
- Juhua Chen, Payaningal R Somanath, Olga Razorenova, William S Chen, Nissim Hay, Paul Bornstein, and Tatiana V Byzova. Akt1 regulates pathological angiogenesis, vascular maturation and permeability in vivo. *Nature medicine*, 11(11):1188, 2005.
- Li Chen, Fenghao Sun, Xiaodong Yang, Yulin Jin, Mengkun Shi, Lin Wang, Yu Shi, Cheng Zhan, and Qun Wang. Correlation between rna-seq and microarrays results using tcga data. *Gene*, 628: 200–204, 2017.
- Jae Eun Choi, Ga Na Oh, Jong Yeob Kim, Soo Hong Seo, Hyo Hyun Ahn, and Young Chul Kye. Ablative fractional laser treatment for hypertrophic scars: comparison between Er: YAG and CO2 fractional lasers. *Journal of Dermatological Treatment*, 25(4):299–303, 2014.
- Ahlame Douzal Chouakria and Panduranga Naidu Nagabhushan. Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, 1(1):5–21, 2007.
- Garret Christensen and Edward Miguel. Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3):920–80, 2018.

BIBLIOGRAPHY

- Rachael A Clark, Benjamin Chong, Nina Mirchandani, Nooshin K Brinster, Kei-ichi Yamanaka, Rebecca K Dowgiert, and Thomas S Kupper. The vast majority of CLA⁺ T cells are resident in normal skin. *The Journal of Immunology*, 176(7):4431–4439, 2006.
- Randall J Cohrs, Tyler Martin, Parviz Ghahramani, Luc Bidaut, Paul J Higgins, and Aamir Shahzad. Translational medicine definition by the European Society for Translational Medicine, 2015.
- Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
- Jana Cole, Raymond Tsou, Ken Wallace, Nicole Gibran, and Frank Isik. Comparison of normal human skin gene expression using cDNA microarrays. *Wound Repair and Regeneration*, 9(2):77–85, 2001.
- Robert B Cotter, Jeffrey D Burke, Magda Stouthamer-Loeber, and Rolf Loeber. Contacting participants for follow-up: how much effort is required to retain participants in longitudinal studies? *Evaluation and Program Planning*, 28(1):15–21, 2005.
- David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 215–242, 1958.
- Shane Crotty. Follicular helper CD4 T cells (T_{fh}). *Annual review of immunology*, 29:621–663, 2011.
- Shane Crotty. T follicular helper cell differentiation, function, and roles in disease. *Immunity*, 41(4):529–542, 2014.
- Frauke Degenhardt, Stephan Seifert, and Silke Szymczak. Evaluation of variable selection methods for random forests and omics data sets. *Briefings in bioinformatics*, 2017.
- Anne S Dejean, Daniel R Beisner, Irene L Ch’En, Yann M Kerdiles, Anna Babour, Karen C Arden, Diego H Castrillon, Ronald A DePinho, and Stephen M Hedrick. Transcription factor Foxo3 controls the magnitude of T cell immune responses by modulating the function of dendritic cells. *Nature immunology*, 10(5):504, 2009.
- Spiros G Delis and Christos Dervenis. Selection criteria for liver resection in patients with hepatocellular carcinoma and chronic liver disease. *World Journal of Gastroenterology: WJG*, 14(22):3452, 2008.
- Jonas Demeulemeester, Parveen Kumar, Elen K Møller, Silje Nord, David C Wedge, April Peterson, Randi R Mathiesen, Renathe Fjellidal, Masoud Zamani Esteki, Koen Theunis, et al. Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *Genome biology*, 17(1):250, 2016.

BIBLIOGRAPHY

- Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.
- Evan Der, Saritha Ranabothu, Hemant Suryawanshi, Kemal M Akat, Robert Clancy, Pavel Morozov, Manjunath Kustagi, Mareike Czuppa, Peter Izmirly, H Michael Belmont, et al. Single cell rna sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI insight*, 2(9), 2017.
- Angelo Massimiliano D’Erme, Dagmar Wilsmann-Theis, Julia Wagenpfeil, Michael Hö lz el, Sandra Ferring-Schmitt, Sonja Sternberg, Miriam Wittmann, Bettina Peters, Andreas Bosio, Thomas Bieber, and Joerg Wenzel. IL-36 γ (IL-1F9) Is a Biomarker for Psoriasis Skin Lesions. *Journal of Investigative Dermatology*, 135(4):1025–1032, apr 2015. doi: 10.1038/jid.2014.532. URL <https://doi.org/10.1038/jid.2014.532>.
- Valeer J Desmet, Michael Gerber, Jay H Hoofnagle, Michael Manns, and Peter J Scheuer. Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology*, 19(6):1513–1520, 1994.
- Thomas G Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine learning*, 32:1–22, 1998.
- Alexis Dinno. *dunn.test: Dunn’s Test of Multiple Comparisons Using Rank Sums*, 2017. URL <https://CRAN.R-project.org/package=dunn.test>. R package version 1.3.5.
- Kieu Trinh Do, Simone Wahl, Johannes Raffler, Sophie Molnos, Michael Laimighofer, Jerzy Adamski, Karsten Suhre, Konstantin Strauch, Annette Peters, Christian Gieger, et al. Characterization of missing values in untargeted ms-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics*, 14(10):128, 2018.
- James G Dolan, Donald R Bordley, and Alvin I Mushlin. An Evaluation of Clinicians’ Subjective Prior Probability Estimates. *Medical Decision Making*, 6(4):216–223, 1986.
- Fritz Drasgow. Polychoric and polyserial correlations. *Encyclopedia of statistical sciences*, 9, 2004.
- Harold Edson Driver and Alfred Louis Kroeber. *Quantitative expression of cultural relationships*, volume 31. University of California Press, 1932.
- Sandrine Dudoit, Juliet Popper Shaffer, and Jennifer C Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, pages 71–103, 2003.
- Thomas Duhren, Rebekka Geiger, David Jarrossay, Antonio Lanzavecchia, and Federica Sallusto. Production of interleukin 22 but not interleukin 17 by a subset of human skin-homing memory T cells. *Nature immunology*, 10(8):857, 2009.

BIBLIOGRAPHY

- Ben W Dulken, Dena S Leeman, Stéphane C Boutet, Katja Hebestreit, and Anne Brunet. Single-cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage. *Cell reports*, 18(3):777–790, 2017.
- Sherry A Dunbar. Applications of Luminex® xMAP™ technology for rapid, high-throughput multiplexed nucleic acid detection. *Clinica chimica acta*, 363(1-2):71–82, 2006.
- Olive Jean Dunn. Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64, 1961.
- Olive Jean Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.
- Alexandra Duque-Fernandez, Lydia Gauthier, Mélissa Simard, Jessica Jean, Isabelle Gendreau, Alexandre Morin, Jacques Soucy, Michèle Auger, and Roxane Pouliot. A 3D-psoriatic skin model for dermatological testing: The impact of culture conditions. *Biochemistry and biophysics reports*, 8:268–276, 2016.
- Steffen Durinck, Paul T Spellman, Ewan Birney, and Wolfgang Huber. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols*, 4(8):1184, 2009.
- Michael L Dustin. The cellular context of T cell signaling. *Immunity*, 30(4):482–492, 2009.
- John M Eisenberg. Sociologic influences on decision-making by clinicians. *Annals of Internal Medicine*, 90(6):957–964, 1979.
- Khalifa El Malki, Susanne H Karbach, Jula Huppert, Morad Zayoud, Sonja Reißig, Rebecca Schüler, Alexej Nikolaev, Khalad Karram, Thomas Münzel, Christoph RW Kuhlmann, et al. An alternative pathway of imiquimod-induced psoriasis-like skin inflammation in the absence of interleukin-17 receptor a signaling. *Journal of Investigative Dermatology*, 133(2):441–451, 2013.
- Daniel Engelbertsen, Sara Rattik, Maria Wigren, Jenifer Vallejo, Goran Marinkovic, Alexandru Schiopu, Harry Björkbacka, Jan Nilsson, and Eva Bengtsson. IL-1R and MyD88 signalling in CD4+ T cells promote Th17 immunity and atherosclerosis. *Cardiovascular research*, 114(1):180–187, 2017.
- Eva Engvall and Peter Perlmann. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry*, 8(9):871–874, 1971.
- Ashraf A Ewis, Zhivko Zhelev, Rumiana Bakalova, Satoshi Fukuoka, Yasuo Shinohara, Mitsuru Ishikawa, and Yoshinobu Baba. A history of microarrays in biomedicine. *Expert review of molecular diagnostics*, 5(3):315–328, 2005.
- Kilian Eyerich, Johannes Huss-Marp, Ulf Darsow, Andreas Wollenberg, Stefanie Foerster, Johannes Ring, Heidrun Behrendt, and Claudia Traidl-Hoffmann. Pollen grains induce a rapid and biphasic

BIBLIOGRAPHY

- eczematous immune response in atopic eczema patients. *International archives of allergy and immunology*, 145(3):213–223, 2008.
- Stefanie Eyerich and Christina E Zielinski. Defining Th-cell subsets in a classical and tissue-specific manner: Examples from the skin. *European journal of immunology*, 44(12):3475–3483, 2014.
- Stefanie Eyerich, Kilian Eyerich, Davide Pennino, Teresa Carbone, Francesca Nasorri, Sabatino Pallotta, Francesca Cianfarani, Teresa Odorisio, Claudia Traidl-Hoffmann, Heidrun Behrendt, et al. Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. *The Journal of clinical investigation*, 119(12):3573–3585, 2009.
- Stefanie Eyerich, Anna T. Onken, Stephan Weidinger, Andre Franke, Francesca Nasorri, Davide Pennino, Martine Grosber, Florian Pfab, Carsten B. Schmidt-Weber, Martin Mempel, Ruediger Hein, Johannes Ring, Andrea Cavani, and Kilian Eyerich. Mutual Antagonism of T Cells Causing Psoriasis and Atopic Eczema. *New England Journal of Medicine*, 365(3):231–238, jul 2011. doi: 10.1056/nejmoa1104200. URL <https://doi.org/10.1056/nejmoa1104200>.
- Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May, Marija Milacic, Corina Duenas Roca, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Solomon Shorser, Thawfeek Varusai, Guilherme Viteri, Joel Weiser, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D’Eustachio. The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, 46 (D1):D649–D655, nov 2017. doi: 10.1093/nar/gkx1132. URL <https://doi.org/10.1093/nar/gkx1132>.
- MA Fadel and AA Tawfik. New topical photodynamic therapy for treatment of hidradenitis suppurativa using methylene blue niosomal gel: a single-blind, randomized, comparative study. *Clinical and experimental dermatology*, 40(2):116–122, 2015.
- Donna L Farber, Naomi A Yudanin, and Nicholas P Restifo. Human memory T cells: generation, compartmentalization and homeostasis. *Nature Reviews Immunology*, 14(1):24, 2014.
- A Fauster, M Rebsamen, KVM Huber, JW Bigenzahn, A Stukalov, CH Lardeau, S Scorzoni, M Bruckner, M Gridling, K Parapatics, et al. A cellular screen identifies ponatinib and pazopanib as inhibitors of necroptosis. *Cell death & disease*, 6(5):e1767, 2015.
- P Fietta and G Delsante. The effector T helper cell triade. *Rivista di biologia*, 102(1):61–74, 2009.
- Christine Fink, Christina Alt, Lorenz Uhlmann, Christina Klose, Alexander Enk, and Holger A Haenssle. Intra- and interobserver variability of image-based PASI assessments in 120 patients suffering from plaque-type psoriasis. *Journal of the European Academy of Dermatology and Venereology*, 2018.

BIBLIOGRAPHY

- Claudio Fiorino, Michele Reni, Angelo Bolognesi, Giovanni Mauro Cattaneo, and Riccardo Calandrino. Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiotherapy and oncology*, 47(3):285–292, 1998.
- Ronald A Fisher. XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.
- Ronald A Fisher. Statistical methods for research workers, 1925a.
- Ronald A Fisher. The logic of inductive inference. *Journal of the Royal Statistical Society*, 98(1): 39–82, 1935.
- Ronald Aylmer Fisher. Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 22, pages 700–725. Cambridge University Press, 1925b.
- Jason D Fontenot, Marc A Gavin, and Alexander Y Rudensky. Foxp3 programs the development and function of CD4+ CD25+ regulatory T cells. *Nature immunology*, 4(4):330, 2003.
- Thomas A Forbes, Sara E Howden, Kynan Lawlor, Belinda Phipson, Jovana Maksimovic, Lorna Hale, Sean Wilson, Catherine Quinlan, Gladys Ho, Katherine Holman, et al. Patient-iPSC-Derived Kidney Organoids Show Functional Validation of a Ciliopathic Renal Phenotype and Reveal Underlying Pathogenetic Mechanisms. *The American Journal of Human Genetics*, 102(5): 816–831, 2018.
- Wolfgang Forstmeier, Eric-Jan Wagenmakers, and Timothy H Parker. Detecting and avoiding likely false-positive findings—a practical guide. *Biological Reviews*, 92(4):1941–1968, 2017.
- Daniel G Fort, Timothy M Herr, Pamela L Shaw, Karen E Gutzman, and Justin B Starren. Mapping the evolving definitions of translational research. *Journal of clinical and translational science*, 1(1):60–66, 2017.
- T Fredriksson and U Pettersson. Severe psoriasis—oral therapy with a new retinoid. *Dermatology*, 157(4):238–244, 1978.
- Michael R Freeman, Francis X Schneck, Michael L Gagnon, Christopher Corless, Shay Soker, Kathy Niknejad, George E Peoples, and Michael Klagsbrun. Peripheral blood T lymphocytes and lymphocytes infiltrating human cancers express vascular endothelial growth factor: a potential role for T cells in angiogenesis. *Cancer Research*, 55(18):4140–4145, 1995.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.
- Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

BIBLIOGRAPHY

- Robin C Friedman, Kyle Kai-How Farh, Christopher B Burge, and David P Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*, 19(1):92–105, 2009.
- Leon A Furchtgott, Carson C Chow, and Vipul Periwal. A model of liver regeneration. *Biophysical journal*, 96(10):3926–3935, 2009.
- TF Gajewski, J Joyce, and FW Fitch. Antiproliferative effect of IFN-gamma in immune regulation. III. Differential selection of TH1 and TH2 murine helper T lymphocyte clones using recombinant IL-2 and recombinant IFN-gamma. *The Journal of Immunology*, 143(1):15–22, 1989.
- Thomas F Gajewski and Frank W Fitch. Anti-proliferative effect of IFN-gamma in immune regulation. I. IFN-gamma inhibits the proliferation of Th2 but not Th1 murine helper T lymphocyte clones. *The Journal of Immunology*, 140(12):4245–4252, 1988.
- Francis Galton. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263, 1886.
- Francis Galton. I. Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 45(273-279):135–145, 1889.
- Nancy Garbacki, Emmanuel Di Valentin, Pierre Geurts, Alexandre Irrthum, Céline Crahay, Thierry Arnould, Christophe Deroanne, Jacques Piette, Didier Cataldo, Alain Colige, et al. MicroRNAs profiling in murine models of acute and chronic asthma: a relationship with mRNAs targets. *PloS one*, 6(1):e16509, 2011.
- Paul J Gardina, Tyson A Clark, Brian Shimada, Michelle K Staples, Qing Yang, James Veitch, Anthony Schweitzer, Tarif Awad, Charles Sugnet, Suzanne Dee, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC genomics*, 7(1):325, 2006.
- D Garmhausen, T Hagemann, T Bieber, I Dimitriou, R Fimmers, T Diepgen, and N Novak. Characterization of different courses of atopic dermatitis in adolescent and adult patients. *Allergy*, 68(4):498–506, 2013.
- Natalie Garzorz-Stark, Linda Krause, Felix Lauffer, Anne Atenhan, Jenny Thomas, Sebastian P Stark, Regina Franz, Stephan Weidinger, Anna Balato, Nikola S Mueller, et al. A novel molecular disease classifier for psoriasis and eczema. *Experimental dermatology*, 25(10):767–774, 2016.
- Natalie Garzorz-Stark, Felix Lauffer, Linda Krause, Jenny Thomas, Anne Atenhan, Regina Franz, Sophie Roenneberg, Alexander Boehner, Manja Jargosch, Richa Batra, et al. Toll-like receptor 7/8 agonists stimulate plasmacytoid dendritic cells to initiate TH17-deviated acute contact dermatitis in human subjects. *Journal of Allergy and Clinical Immunology*, 141(4):1320–1333, 2018.
- Carl Friedrich Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, volume 7. Perthes et Besser, 1809.

BIBLIOGRAPHY

- Ricardo T Gazzinelli, Maria Wysocka, Sara Hieny, Tanya Schariton-Kersten, Allen Cheever, R Kühn, W Müller, Giorgio Trinchieri, and Alan Sher. In the absence of endogenous IL-10, mice acutely infected with *Toxoplasma gondii* succumb to a lethal immune response dependent on CD4+ T cells and accompanied by overproduction of IL-12, IFN-gamma and TNF-alpha. *The Journal of Immunology*, 157(2):798–805, 1996.
- J Geginat, M Paroni, F Facciotti, P Gruarin, I Kastirr, F Caprioli, M Pagani, and S Abrignani. The CD4-centered universe of human T cell subsets. In *Seminars in immunology*, volume 25, pages 252–262. Elsevier, 2013.
- Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*, 45(D1):D331–D338, 2016.
- Gerard George, Ernst C Osinga, Dovev Lavie, and Brent A Scott. Big data and data science methods for management research, 2016.
- Philipp E Geyer, Lesca M Holdt, Daniel Teupser, and Matthias Mann. Revisiting biomarker discovery by plasma proteomics. *Molecular systems biology*, 13(9):942, 2017.
- Mohammad M Ghassemi, Tuka Al-Hanai, Jesse D Raffa, Roger G Mark, Shamim Nemati, and Falgun H Chokshi. How is the Doctor Feeling? ICU Provider Sentiment is Associated with Diagnostic Imaging Utilization. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4058–4064. IEEE, 2018.
- Alice Giustacchini, Supat Thongjuea, Nikolaos Barkas, Petter S Woll, Benjamin J Povinelli, Christopher AG Booth, Paul Sopp, Ruggiero Norfo, Alba Rodriguez-Meira, Neil Ashley, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nature medicine*, 23(6):692, 2017.
- David J Gladstone, Melanie Spring, Paul Dorian, Val Panzov, Kevin E Thorpe, Judith Hall, Haris Vaid, Martin O’donnell, Andreas Laupacis, Robert Côté, et al. Atrial fibrillation in patients with cryptogenic stroke. *New England Journal of Medicine*, 370(26):2467–2477, 2014.
- Peter J Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192, 1984.
- Lynda Grine, Lien Dejager, Claude Libert, and Roosmarijn E Vandenbroucke. Dual inhibition of TNFR1 and IFNAR1 in imiquimod-induced psoriasiform skin inflammation in mice. *The Journal of Immunology*, page 1403015, 2015.
- Emma Guttman-Yassky, Mayte Suárez-Fariñas, Andrea Chiricozzi, Kristine E. Nogales, Avner Shemer, Judilyn Fuentes-Duculan, Irma Cardinale, Peng Lin, Reuven Bergman, Anne M. Bowcock, and James G. Krueger. Broad defects in epidermal cornification in atopic dermatitis identified

BIBLIOGRAPHY

- through genomic analysis. *Journal of Allergy and Clinical Immunology*, 124(6):1235–1244.e58, dec 2009. doi: 10.1016/j.jaci.2009.09.031. URL <https://doi.org/10.1016/j.jaci.2009.09.031>.
- Emma Guttman-Yassky, Kristine E Nograles, and James G Krueger. Contrasting pathogenesis of atopic dermatitis and psoriasis—part II: immune cell subsets and therapeutic concepts. *Journal of Allergy and Clinical Immunology*, 127(6):1420–1432, 2011.
- Laleh Haghverdi, Florian Buettner, and Fabian J Theis. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, 31(18):2989–2998, 2015.
- Kristine Halfar, Christian Rommel, Hugo Stocker, and Ernst Hafen. Ras controls growth, survival and differentiation in the Drosophila eye by different thresholds of MAP kinase activity. *Development*, 128(9):1687–1696, 2001.
- Brigham J Hartley and Kristen J Brennard. Neural organoids for disease phenotyping, drug screening and developmental biology studies. *Neurochemistry international*, 106:85–93, 2017.
- Helena Hauge, Sebastian Patzke, and Hans-Christian Aasheim. Characterization of the FAM110 gene family. *Genomics*, 90(1):14–27, 2007.
- Helena Hauge, Kristine Engelsen Fjelland, Mouldy Sioud, and Hans-Christian Aasheim. Evidence for the involvement of FAM110C protein in cell spreading and migration. *Cellular signalling*, 21(12):1866–1873, 2009.
- Jason E Hawkes, Johann E Gudjonsson, and Nicole L Ward. The snowballing literature on imiquimod-induced skin inflammation in mice: a critical appraisal. *Journal of Investigative Dermatology*, 137(3):546–549, 2017.
- Roderick J Hay, Nicole E Johns, Hywel C Williams, Ian W Bolliger, Robert P Dellavalle, David J Margolis, Robin Marks, Luigi Naldi, Martin A Weinstock, Sarah K Wulf, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *Journal of Investigative Dermatology*, 134(6):1527–1534, 2014.
- J Heinrich, G Bolte, B Hölscher, J Douwes, I Lehmann, B Fahlbusch, W Bischof, M Weiss, M Borte, HE Wichmann, et al. Allergens and endotoxin on mothers’ mattresses and total immunoglobulin E in cord blood of neonates. *European Respiratory Journal*, 20(3):617–623, 2002.
- Silvio Hemmi, Ruth Böhni, Gerlinde Stark, Francesco DI Marco, and Michel Aguet. A novel member of the interferon receptor family complements functionality of the murine interferon γ receptor in human cells. *Cell*, 76(5):803–810, 1994.
- Lisa Héron-Milhavet, Celine Franckhauser, Vanessa Rana, Cyril Berthenet, Daniel Fisher, Brian A Hemmings, Anne Fernandez, and Ned JC Lamb. Only Akt1 is required for proliferation, while Akt2 promotes cell cycle exit through p21 binding. *Molecular and cellular biology*, 26(22):8267–8280, 2006.

BIBLIOGRAPHY

- Tony Hey, Stewart Tansley, Kristin M Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA, 2009.
- Russell Higuchi, Carita Fockler, Gavin Dollinger, and Robert Watson. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nature Biotechnology*, 11(9):1026, 1993.
- Joshua WK Ho, Maurizio Stefani, Cristobal G dos Remedios, and Michael A Charleston. Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics*, 24(13):i390–i398, 2008.
- Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- Benjamin Hofner and Torsten Hothorn. *stabs: Stability Selection with Error Control*, 2017. URL <https://CRAN.R-project.org/package=stabs>. R package version 0.6-3.
- Benjamin Hofner, Luigi Boccutto, and Markus Göker. Controlling false discoveries in high-dimensional situations: boosting with stability selection. *BMC bioinformatics*, 16(1):144, 2015.
- Rolf Holle, Michael Happich, Hannelore Löwel, Heinz-Erich Wichmann, null for the MONICA/KORA Study Group, et al. Kora-a research platform for population based health research. *Das Gesundheitswesen*, 67(S 01):19–25, 2005.
- Bernhard Homey, Harri Alenius, Anja Müller, Hortensia Soto, Edward P. Bowman, Wei Yuan, Leslie McEvoy, Antti I. Lauerma, Till Assmann, Erich Bünemann, Maili Lehto, Henrik Wolff, David Yen, Heather Marxhausen, Wayne To, Jonathon Sedgwick, Thomas Ruzicka, Percy Lehmann, and Albert Zlotnik. CCL27–CCR10 interactions regulate T cell-mediated skin inflammation. *Nature Medicine*, 8(2):157–165, feb 2002. doi: 10.1038/nm0202-157. URL <https://doi.org/10.1038/nm0202-157>.
- Philipp S Hoppe, Michael Schwarzfischer, Dirk Loeffler, Konstantinos D Kokkaliaris, Oliver Hilsenbeck, Nadine Moritz, Max Endeke, Adam Filipczyk, Adriana Gambardella, Nouraz Ahmed, et al. Early myeloid lineage choice is not initiated by random pu. 1 to gata1 protein ratios. *Nature*, 535(7611):299, 2016.
- Heidi Hörig, Elizabeth Marincola, and Francesco M Marincola. Obstacles and opportunities in translational research. *Nature medicine*, 11(7):705, 2005.
- Shu Horiuchi, Atsushi Onodera, Hiroyuki Hosokawa, Yukiko Watanabe, Tomoaki Tanaka, Sumio Sugano, Yutaka Suzuki, and Toshinori Nakayama. Genome-wide analysis reveals unique regulation of transcription of th2-specific genes by gata3. *The Journal of Immunology*, 186(11):6378–6389, 2011.
- Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

BIBLIOGRAPHY

- Li-Fei Hou, Shi-Jun He, Xin Li, Yang Yang, Pei-Lan He, Yu Zhou, Feng-Hua Zhu, Yi-Fu Yang, Ying Li, Wei Tang, et al. Oral administration of artemisinin analog SM934 ameliorates lupus syndromes in MRL/lpr mice by inhibiting Th1 and Th17 cell responses. *Arthritis & Rheumatism*, 63(8):2445–2455, 2011.
- Chyi-Song Hsieh, Steven E Macatonia, Catherine S Tripp, Stanley F Wolf, Anne O’garra, and Kenneth M Murphy. Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages. *Science*, 260(5107):547–549, 1993.
- Dan Hu, Samuele Notarbartolo, Tom Croonenborghs, Bonny Patel, Ron Cialic, Tun-Hsiang Yang, Dominik Aschenbrenner, Karin M Andersson, Marco Gattorno, Minh Pham, et al. Transcriptional signature of human pro-inflammatory TH 17 cells identifies reduced IL10 gene expression in multiple sclerosis. *Nature communications*, 8(1):1600, 2017.
- David J Hunter. Gene–environment interactions in human diseases. *Nature Reviews Genetics*, 6(4): 287, 2005.
- Ivan Igaz, Gábor Nyíró, Zoltán Nagy, Henriett Butz, Zsolt Nagy, Pál Perge, Peter Sahin, Miklós Tóth, Károly Rácz, Peter Igaz, et al. Analysis of circulating MicroRNAs in vivo following administration of dexamethasone and adrenocorticotropin. *International Journal of Endocrinology*, 2015, 2015.
- Megan S. Inkeles, Philip O. Scumpia, William R. Swindell, David Lopez, Rosane M.B. Teles, Thomas G. Graeber, Stephan Meller, Bernhard Homey, James T. Elder, Michel Gilliet, Robert L. Modlin, and Matteo Pellegrini. Comparison of Molecular Signatures from Multiple Skin Diseases Identifies Mechanisms of Immunopathogenesis. *Journal of Investigative Dermatology*, 135(1): 151–159, jan 2015. doi: 10.1038/jid.2014.352. URL <https://doi.org/10.1038/jid.2014.352>.
- Ivaylo I Ivanov, Brent S McKenzie, Liang Zhou, Carlos E Tadokoro, Alice Lepelley, Juan J Lafaille, Daniel J Cua, and Dan R Littman. The orphan nuclear receptor ROR γ t directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell*, 126(6):1121–1133, 2006.
- Anil K Jain. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8):651–666, 2010.
- Silke Janitza, Ender Celik, and Anne-Laure Boulesteix. A computationally fast variable importance test for random forests for high-dimensional data. *Advances in Data Analysis and Classification*, pages 1–31, 2015.
- Matlock Jeffries, Mikhail Dozmorov, Yuhong Tang, Joan T Merrill, Jonathan D Wren, and Amr H Sawalha. Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus. *Epigenetics*, 6(5):593–601, 2011.
- Peter Jensen and Lone Skov. Psoriasis and obesity. *Dermatology*, 232(6):633–639, 2016.
- H. Joe and Jr. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

BIBLIOGRAPHY

- Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.
- 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.
- Robert J Johnston, Amanda C Poholek, Daniel DiToro, Isharat Yusuf, Danelle Eto, Burton Barnett, Alexander L Dent, Joe Craft, and Shane Crotty. Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science*, 325(5943):1006–1010, 2009.
- Ian Jolliffe. *Principal Component Analysis*, pages 1094–1096. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_455. URL https://doi.org/10.1007/978-3-642-04898-2_455.
- Shinji Kagami, Hidehisa Saeki, Yuichiro Tsunemi, Koichiro Nakamura, Yoshihiro Kuwano, Mayumi Komine, Takashi Nakayama, Osamu Yoshie, and Kunihiko Tamaki. CCL27-transgenic mice show enhanced contact hypersensitivity to Th2, but not Th1 stimuli. *European journal of immunology*, 38(3):647–657, 2008.
- Takashi Kakinuma, Hidehisa Saeki, Yuichiro Tsunemi, Hideki Fujita, Noriko Asano, Hiroshi Mitsui, Yayoi Tada, Motoshi Wakugawa, Takahiro Watanabe, Hideshi Torii, et al. Increased serum cutaneous T cell-attracting chemokine (CCL27) levels in patients with atopic dermatitis and psoriasis vulgaris. *Journal of allergy and clinical immunology*, 111(3):592–597, 2003.
- Hee-Kap Kang, Diane Ecklund, Michael Liu, and Syamal K Datta. Apigenin, a non-mutagenic dietary flavonoid, suppresses lupus by inhibiting autoantigen presentation for expansion of autoreactive Th1 and Th17 cells. *Arthritis research & therapy*, 11(2):R59, 2009.
- Aditi Kanhere, Arnulf Hertweck, Urvashi Bhatia, M Refik Gökmen, Esperanza Perucha, Ian Jackson, Graham M Lord, and Richard G Jenner. T-bet and GATA3 orchestrate Th1 and Th2 differentiation through lineage-specific targeting of distal regulatory elements. *Nature communications*, 3:1268, 2012.
- Jagat R Kanwar, Rupinder K Kanwar, Hannah Burrow, and Sara Baratchi. Recent advances on the roles of NO in cancer and chronic inflammatory disorders. *Current medicinal chemistry*, 16(19):2373–2394, 2009.
- Daniel H Kaplan, Botond Z Igyártó, and Anthony A Gaspari. Early immune events in the induction of allergic contact dermatitis. *Nature Reviews Immunology*, 12(2):114, 2012.
- Audrey Kauffmann, Robert Gentleman, and Wolfgang Huber. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25(3):415–6, 2009.
- Leonard Kaufman and Peter Rousseeuw. *Clustering by means of medoids*. North-Holland, 1987.
- Michael G Kenward and James H Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pages 983–997, 1997.

BIBLIOGRAPHY

- Paul A Khavari. Modelling cancer in human skin tissue. *Nature Reviews Cancer*, 6(4):270, 2006.
- Muin J Khoury, Marta Gwinn, Paula W Yoon, Nicole Dowling, Cynthia A Moore, and Linda Bradley. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genetics in Medicine*, 9(10):665, 2007.
- Ville Kiiski, Oskar Karlsson, Anita Remitz, and Sakari Reitamo. High serum total IgE predicts poor long-term outcome in atopic dermatitis. *Acta dermato-venereologica*, 95(8):943–947, 2015.
- Thomas Korn, Mohamed Oukka, Vijay Kuchroo, and Estelle Bettelli. Th17 cells: effector T cells with inflammatory properties. In *Seminars in immunology*, volume 19, pages 362–371. Elsevier, 2007.
- Sandra Koseoglu, Zhuomei Lu, Chandra Kumar, Paul Kirschmeier, and Jun Zou. AKT1, AKT2 and AKT3-dependent cell survival is cell line-specific and knockdown of all three isoforms selectively induces apoptosis in 20 human tumor cell lines. *Cancer biology & therapy*, 6(5):755–762, 2007.
- SB Kotsiantis, Dimitris Kanellopoulos, and PE Pintelas. Data preprocessing for supervised learning. *International Journal of Computer Science*, 1(2):111–117, 2006.
- Linda Krause, Vagkan Mourantchanian, Knut Brockow, Fabian J Theis, Carsten B Schmidt-Weber, Bettina Knapp, Nikola S Mueller, and Stefanie Eyerich. A computational model to predict severity of atopic eczema from 30 serum proteins. *Journal of Allergy and Clinical Immunology*, 138(4):1207–1210, 2016.
- Linda Krause, Nikola S Mueller, and Stefanie Eyerich. Multiplex platform technology and bioinformatics are essential for development of biomarkers in atopic dermatitis Reply. *Journal of Allergy and Clinical Immunology*, 139(3):1065–1066, 2017.
- Jan Krumsiek, Karsten Suhre, Thomas Illig, Jerzy Adamski, and Fabian J Theis. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC systems biology*, 5(1):21, 2011.
- Joseph B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
- Joseph B Kruskal. An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM review*, 25(2):201–237, 1983.
- William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.
- Dong-Ming Kuang, Xiao Xiao, Qiyi Zhao, Min-Min Chen, Xue-Feng Li, Rui-Xian Liu, Yuan Wei, Fang-Zhu Ouyang, Dong-Ping Chen, Yan Wu, et al. B7-H1-expressing antigen-presenting cells

BIBLIOGRAPHY

- mediate polarization of protumorigenic Th22 subsets. *The Journal of clinical investigation*, 124(10):4657–4667, 2014.
- Matthew A Kunicki, Laura C Amaya Hernandez, Kara L Davis, Rosa Bacchetta, and Maria-Grazia Roncarolo. Identity and diversity of human peripheral Th and T regulatory cells defined by single-cell mass cytometry. *The Journal of Immunology*, 200(1):336–346, 2018.
- B Kunz, AP Oranje, L Labreze, J-F Stalder, J Ring, and A Taieb. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. *Dermatology*, 195(1):10–19, 1997.
- Miron B Kursa, Witold R Rudnicki, et al. Feature selection with the Boruta package. *J Stat Softw*, 36(11):1–13, 2010.
- Godfrey N Lance and William Thomas Williams. A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The computer journal*, 9(4):373–380, 1967.
- Richard G Langley and Charles N Ellis. Evaluating psoriasis with psoriasis area and severity index, psoriasis global assessment, and lattice system physician’s global assessment. *Journal of the American Academy of Dermatology*, 51(4):563–569, 2004.
- Richard G. Langley, Boni E. Elewski, Mark Lebwohl, Kristian Reich, Christopher E.M. Griffiths, Kim Papp, Lluís Puig, Hidemi Nakagawa, Lynda Spelman, B Sigurgeirsson, Enrique Rivas, Tsen-Fang Tsai, Norman Wasel, Stephen Tyring, Thomas Salko, Isabelle Hampele, Marianne Notter, Alexander Karpov, Silvia Helou, and Charis Papavassilis. Secukinumab in Plaque Psoriasis — Results of Two Phase 3 Trials. *New England Journal of Medicine*, 371(4):326–338, jul 2014. doi: 10.1056/nejmoa1314258. URL <https://doi.org/10.1056/nejmoa1314258>.
- Richard GB Langley, Steven R Feldman, Judit Nyirady, Peter van de Kerkhof, and Charis Papavassilis. The 5-point Investigator’s Global Assessment (IGA) Scale: a modified tool for evaluating plaque psoriasis severity in clinical trials. *Journal of Dermatological Treatment*, 26(1): 23–31, 2015.
- Felix Lauffer, Manja Jargosch, Linda Krause, Natalie Garzorz-Stark, Regina Franz, Sophie Roenneberg, Alexander Böhner, Nikola S Mueller, Fabian J Theis, Carsten B Schmidt-Weber, et al. Type I Immune Response Induces Keratinocyte Necroptosis and Is Associated with Interface Dermatitis. *Journal of Investigative Dermatology*, 2018.
- Cosmin Lazar, Laurent Gatto, Myriam Ferro, Christophe Bruley, and Thomas Burger. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *Journal of proteome research*, 15(4):1116–1125, 2016.
- F Le Duff, E Fontas, D Giacchero, L Sillard, J-P Lacour, J-P Ortonne, and T Passeron. 308-nm excimer lamp vs. 308-nm excimer laser for treating vitiligo: a randomized study. *British Journal of dermatology*, 163(1):188–192, 2010.

BIBLIOGRAPHY

- Graham Le Gros, Shlomo Z Ben-Sasson, Robert Seder, FD Finkelman, and WE Paul. Generation of interleukin 4 (IL-4)-producing cells in vivo and in vitro: IL-2 and IL-4 are required for in vitro generation of IL-4-producing cells. *Journal of Experimental Medicine*, 172(3):921–929, 1990.
- Etienne P LeBel, Randy J McCarthy, Brian D Earp, Malte Elson, and Wolf Vanpaemel. A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3):389–402, 2018.
- Mark G Lebwohl, Arthur Kavanaugh, April W Armstrong, and Abby S Van Voorhees. US perspectives in the management of psoriasis and psoriatic arthritis: patient and physician results from the population-based Multinational Assessment of Psoriasis and Psoriatic Arthritis (MAPP) survey. *American journal of clinical dermatology*, 17(1):87–97, 2016.
- June-Yong Lee, Cara N Skon, You Jeong Lee, Soohwan Oh, Justin J Taylor, Deepali Malhotra, Marc K Jenkins, M Geoffrey Rosenfeld, Kristin A Hogquist, and Stephen C Jameson. The transcription factor KLF2 restrains CD4+ T follicular helper cell differentiation. *Immunity*, 42(2):252–264, 2015.
- Jeffrey T Leek and John D Storey. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics*, 3(9):e161, 2007.
- Jeffrey T Leek and John D Storey. A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences*, 105(48):18718–18723, 2008.
- Jeffrey T Leek, Robert B Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733, 2010.
- Jeffrey T Leek, W Evan Johnson, Hilary S Parker, Andrew E Jaffe, and John D Storey. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6):882–883, 2012.
- Adrien Marie Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot, 1805.
- Marco L Leung, Alexander Davis, Ruli Gao, Anna Casasent, Yong Wang, Emi Sei, Eduardo Sanchez, Dipen Maru, Scott Kopetz, and Nicholas E Navin. Single cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome research*, pages gr–209973, 2017.
- Stewart Leung, Xuebin Liu, Lei Fang, Xi Chen, Taylor Guo, and Jingwu Zhang. The cytokine milieu in the interplay of pathogenic Th1/Th17 cells and regulatory T cells in autoimmune disease. *Cellular & molecular immunology*, 7(3):182, 2010.

BIBLIOGRAPHY

- Andre A Lighvani, David M Frucht, Dragana Jankovic, Hidehiro Yamane, Julio Aliberti, Bruce D Hisson, Bai V Nguyen, Massimo Gadina, Alan Sher, William E Paul, et al. T-bet is rapidly induced by interferon- γ in lymphoid and myeloid cells. *Proceedings of the National Academy of Sciences*, 98(26):15137–15142, 2001.
- Zhixiang Lin, Can Yang, Ying Zhu, John Duchi, Yao Fu, Yong Wang, Bai Jiang, Mahdi Zamanighomi, Xuming Xu, Mingfeng Li, et al. Simultaneous dimension reduction and adjustment for confounding variation. *Proceedings of the National Academy of Sciences*, 113(51):14662–14667, 2016.
- Tiantian Liu, Hongjin Wu, Shixiu Wu, and Charles Wang. Single-cell sequencing technologies for cardiac stem cell studies. *Stem cells and development*, 26(21):1540–1551, 2017.
- Yang Liu, Yan-Ling Guo, Shi-Jie Zhou, Fei Liu, Feng-Jiao Du, Xiao-Jing Zheng, Hong-Yan Jia, and Zong-De Zhang. CREB is a positive transcriptional regulator of IFN- γ in latent, but not active, tuberculosis infections. *Clinical and Vaccine Immunology*, 2010.
- Michelle A Lowes, Toyoko Kikuchi, Judilyn Fuentes-Duculan, Irma Cardinale, Lisa C Zaba, Asifa S Haider, Edward P Bowman, and James G Krueger. Psoriasis vulgaris lesions contain discrete populations of Th1 and Th17 T cells. *Journal of Investigative Dermatology*, 128(5):1207–1211, 2008.
- Thomas X Lu, Ariel Munitz, and Marc E Rothenberg. MicroRNA-21 is up-regulated in allergic airway inflammation and regulates IL-12p35 expression. *The Journal of Immunology*, 182(8):4994–5002, 2009.
- Thomas Lumley. *leaps: Regression Subset Selection*, 2017. URL <https://CRAN.R-project.org/package=leaps>. R package version 3.0.
- Shuangge Ma, Xiao Song, and Jian Huang. Supervised group lasso with applications to microarray data analysis. *BMC bioinformatics*, 8(1):60, 2007.
- Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- Iain C Macaulay, Chris P Ponting, and Thierry Voet. Single-cell multiomics: multiple measurements from single cells. *Trends in Genetics*, 33(2):155–168, 2017.
- James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- Bidesh Mahata, Xiuwei Zhang, Aleksandra A Kolodziejczyk, Valentina Proserpio, Liora Haim-Vilmovsky, Angela E Taylor, Daniel Hebenstreit, Felix A Dingler, Victoria Moignard, Berthold Göttgens, et al. Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell reports*, 7(4):1130–1142, 2014.

BIBLIOGRAPHY

- David Mangnall, Nigel C Bird, and Ali W Majeed. The molecular physiology of liver regeneration following partial hepatectomy. *Liver international*, 23(2):124–138, 2003.
- CJ Mann. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency medicine journal*, 20(1):54–60, 2003.
- Teri A Manolio, Francis S Collins, Nancy J Cox, David B Goldstein, Lucia A Hindorff, David J Hunter, Mark I McCarthy, Erin M Ramos, Lon R Cardon, Aravinda Chakravarti, et al. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747, 2009.
- Kirk J Mantione, Richard M Kream, Hana Kuzelova, Radek Ptacek, Jiri Raboch, Joshua M Samuel, and George B Stefano. Comparing bioinformatic gene expression profiling methods: microarray and rna-seq. *Medical science monitor basic research*, 20:138, 2014.
- Gustavo J Martinez, Zhengmao Zhang, Yeonseok Chung, Joseph M Reynolds, Xia Lin, Anton M Jetten, Xin-Hua Feng, and Chen Dong. Smad3 differentially regulates the induction of regulatory and inflammatory T cell differentiation. *Journal of Biological Chemistry*, pages jbc-C109, 2009.
- Thomas Z Mayer, François A Simard, Alexandre Cloutier, Harsh Vardhan, Claire M Dubois, and Patrick P McDonald. The p38-MSK1 signaling cascade influences cytokine production through CREB and C/EBP factors in human neutrophils. *The Journal of Immunology*, page 1301117, 2013.
- Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- John McKinlay, Carol Link, Sara Arber, Lisa Marceau, Amy O'donnell, and Ann Adams. How do doctors in different countries manage the same patient? Results of a factorial experiment. *Health Services Research*, 41(6):2182–2200, 2006.
- John B McKinlay, Deborah A Potter, and Henry A Feldman. Non-medical influences on medical decision-making. *Social science & medicine*, 42(5):769–776, 1996.
- Cyrus R Mehta and Nitin R Patel. A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *Journal of the American Statistical Association*, 78(382):427–434, 1983.
- Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.
- Javier Mestas and Christopher CW Hughes. Of mice and not men: differences between mouse and human immunology. *The Journal of Immunology*, 172(5):2731–2738, 2004.
- Taciano L Milfont and Richard A Klein. Replication and reproducibility in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, 49(5):735–750, 2018.

BIBLIOGRAPHY

- Katrin Milger, Jeremias Götschke, Linda Krause, Petra Nathan, Francesca Alessandrini, Amanda Tufman, Rainald Fischer, Sabine Bartel, Fabian J Theis, Jürgen Behr, et al. Identification of a plasma miRNA biomarker signature for allergic asthma: A translational approach. *Allergy*, 72(12):1962–1971, 2017.
- Pierre Milpied, Amédée Renand, Julie Bruneau, Daniella A Mendes-Da-Cruz, Sébastien Jacquelin, Vahid Asnafi, Marie-Thérèse Rubio, Elizabeth MacIntyre, Yves Lepelletier, and Olivier Hermine. Neuropilin-1 is not a marker of human Foxp3+ Treg. *European journal of immunology*, 39(6):1466–1471, 2009.
- Kent T Miner and Michael Croft. Generation, persistence, and modulation of Th0 effector cells: role of autocrine IL-4 and IFN- γ . *The Journal of Immunology*, 160(11):5280–5287, 1998.
- Meng-Hsuan Mo, Liang Chen, Yebo Fu, Wendy Wang, and Sidney W Fu. Cell-free circulating miRNA biomarkers in cancer. *Journal of Cancer*, 3:432, 2012.
- Fazilat F Mohammed and Rama Khokha. Thinking outside the cell: proteases regulate hepatocyte division. *Trends in cell biology*, 15(10):555–563, 2005.
- Gemma Moncunill, John J Aponte, Augusto J Nhabomba, and Carlota Dobaño. Performance of multiplex commercial kits to quantify cytokine and chemokine responses in culture supernatants from Plasmodium falciparum stimulations. *PloS one*, 8(1):e52587, 2013.
- Pablo Montero, José A Vilar, et al. Tsclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1):1–43, 2014.
- Takeshi Morimoto, Mio Sakuma, Kunihiko Matsui, Nobuo Kuramoto, Jinichi Toshiro, Junji Murakami, Tsuguya Fukui, Mayuko Saito, Atsushi Hiraide, and David W Bates. Incidence of adverse drug events and medication errors in Japan: the JADE study. *Journal of general internal medicine*, 26(2):148–153, 2011.
- Timothy R Mosmann, Holly Cherwinski, Martha W Bond, Martin A Giedlin, and Robert L Coffman. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *The Journal of immunology*, 136(7):2348–2357, 1986.
- Kary Mullis, Fred Faloona, Stephen Scharf, RK Saiki, GT Horn, and H Erlich. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor symposia on quantitative biology*, volume 51, pages 263–273. Cold Spring Harbor Laboratory Press, 1986.
- Ugrappa Nagalakshmi, Zhong Wang, Karl Waern, Chong Shou, Debasish Raha, Mark Gerstein, and Michael Snyder. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881):1344–1349, 2008.
- NAKO e. V. NAKO Gesundheitsstudie, 2019. URL <https://nako.de/>. [Online; accessed 25-February-2019].

BIBLIOGRAPHY

- National Institutes of Health. PubMed - US National Library of Medicine, 2019. URL <https://www.ncbi.nlm.nih.gov/pubmed/>. [Online; accessed 13-February-2019].
- JA Nelder and RWM Wedderburn. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, pages 370–384, 1972.
- Frank O. Nestle, Daniel H. Kaplan, and Jonathan Barker. Psoriasis. *New England Journal of Medicine*, 361(5):496–509, 2009. doi: 10.1056/NEJMra0804595. URL <https://doi.org/10.1056/NEJMra0804595>. PMID: 19641206.
- Elena Niccolai, Antonio Taddei, Federica Ricci, Simona Rolla, Mario Milco D’Elios, Marisa Benagiano, Paolo Bechi, Lapo Bencini, Maria Novella Ringressi, Alessandro Pini, et al. Intra-tumoral IFN- γ -producing Th22 cells correlate with TNM staging and the worst outcomes in pancreatic cancer. *Clinical Science*, 130(4):247–258, 2016.
- Jo Nishino, Hidenori Ochi, Yuta Kochi, Tatsuhiko Tsunoda, and Shigeyuki Matsui. Sample size for successful genome-wide association study of major depressive disorder. *Frontiers in genetics*, 9: 227, 2018.
- Furong Niu, Diane C Wang, Jiapei Lu, Wei Wu, and Xiangdong Wang. Potentials of single-cell biology in identification and validation of disease biomarkers. *Journal of cellular and molecular medicine*, 20(9):1789–1795, 2016.
- Shinji Noda, James G Krueger, and Emma Guttman-Yassky. The translational revolution and use of biologics in patients with inflammatory skin diseases. *Journal of Allergy and Clinical Immunology*, 135(2):324–336, 2015.
- Ichiro Nomura, Bifeng Gao, Mark Boguniewicz, Marc A Darst, Jeffrey B Travers, and Donald YM Leung. Distinct patterns of gene expression in the skin lesions of atopic dermatitis and psoriasis: a gene microarray analysis. *Journal of Allergy and Clinical Immunology*, 112(6):1195–1202, 2003.
- Roza I Nurieva, Yeonseok Chung, Gustavo J Martinez, Xuexian O Yang, Shinya Tanaka, Tatyana D Matskevitch, Yi-Hong Wang, and Chen Dong. Bcl6 mediates the development of T follicular helper cells. *Science*, 325(5943):1001–1005, 2009.
- Duncan T Odom, Nora Zizlsperger, D Benjamin Gordon, George W Bell, Nicola J Rinaldi, Heather L Murray, Tom L Volkert, Jörg Schreiber, P Alexander Rolfe, David K Gifford, et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science*, 303(5662):1378–1381, 2004.
- Ulf Olsson, Fritz Drasgow, and Neil J Dorans. The polyserial correlation coefficient. *Psychometrika*, 47(3):337–347, 1982.
- Oliver Ommen, Britta Ullrich, Christian Janßen, and Holger Pfaff. Die ambulante-stationäre Schnittstelle in der medizinischen Versorgung. *Medizinische Klinik*, 102(11):913–917, 2007.

BIBLIOGRAPHY

- Nooshin Omranian, Jeanne MO Eloundou-Mbebi, Bernd Mueller-Roeber, and Zoran Nikoloski. Gene regulatory network inference using fused lasso on multiple data sets. *Scientific reports*, 6: 20533, 2016.
- Lucille ML Ong, Johanna CJM De Haes, Alaysia M Hoos, and Frits B Lammes. Doctor-patient communication: a review of the literature. *Social science & medicine*, 40(7):903–918, 1995.
- Chiaki Ono, Zhiqian Yu, Yoshiyuki Kasahara, Yoshie Kikuchi, Naoto Ishii, and Hiroaki Tomita. Fluorescently activated cell sorting followed by microarray profiling of helper T cell subtypes from human peripheral blood. *PloS one*, 9(11):e111405, 2014.
- Sandra Orchard, Samuel Kerrien, Sara Abbani, Bruno Aranda, Jignesh Bhate, Shelby Bidwell, Alan Bridge, Leonardo Briganti, Fiona SL Brinkman, Gianni Cesareni, et al. Protein interaction data curation: the international molecular exchange (imex) consortium. *Nature methods*, 9(4): 345, 2012.
- AD Ormerod, R Weller, P Copeland, N Benjamin, SH Ralston, P Grabowksi, and Richard Herriot. Detection of nitric oxide and nitric oxide synthases in psoriasis. *Archives of Dermatological Research*, 290(1-2):3–8, 1998.
- H Ott, J Wilke, JM Baron, PH Höger, and R Fölster-Holst. Soluble immune receptor serum levels are associated with age, but not with clinical phenotype or disease severity in childhood atopic dermatitis. *Journal of the European Academy of Dermatology and Venereology*, 24(4):395–402, 2010.
- Hervé Pagès, Marc Carlson, Seth Falcon, and Nianhua Li. *AnnotationDbi: Annotation Database Interface*, 2018. R package version 1.43.1.
- Ronaldo P Panganiban, Yanli Wang, Judie Howrylak, Vernon M Chinchilli, Timothy J Craig, Avery August, and Faoud T Ishmael. Circulating microRNAs as biomarkers in patients with allergic rhinitis and asthma. *Journal of Allergy and Clinical Immunology*, 137(5):1423–1432, 2016.
- Rosa Parisi, Deborah P.M. Symmons, Christopher E.M. Griffiths, and Darren M. Ashcroft. Global Epidemiology of Psoriasis: A Systematic Review of Incidence and Prevalence. *Journal of Investigative Dermatology*, 133(2):377–385, feb 2013. doi: 10.1038/jid.2012.339. URL <https://doi.org/10.1038/jid.2012.339>.
- Jacqueline Parkin and Bryony Cohen. An overview of the immune system. *The Lancet*, 357(9270): 1777–1789, 2001.
- Anoop P Patel, Itay Tirosh, John J Trombetta, Alex K Shalek, Shawn M Gillespie, Hiroaki Wakimoto, Daniel P Cahill, Brian V Nahed, William T Curry, Robert L Martuza, et al. Single-cell rna-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190): 1396–1401, 2014.

BIBLIOGRAPHY

- James G Patton, Jeffrey L Franklin, Alissa M Weaver, Kasey Vickers, Bing Zhang, Robert J Coffey, K Mark Ansel, Robert Blelloch, Andrei Goga, Bo Huang, et al. Biogenesis, delivery, and function of extracellular RNA. *Journal of extracellular vesicles*, 4(1):27494, 2015.
- Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.
- Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.
- Gayathri K. Perera, Paola Di Meglio, and Frank O. Nestle. Psoriasis. *Annual Review of Pathology: Mechanisms of Disease*, 7(1):385–422, feb 2012. doi: 10.1146/annurev-pathol-011811-132448. URL <https://doi.org/10.1146/annurev-pathol-011811-132448>.
- Alessandra Pernis, Sanjay Gupta, Kenneth J Gollob, Evan Garfein, Robert L Coffman, Chris Schindler, and Paul Rothman. Lack of interferon gamma receptor beta chain and the prevention of interferon gamma signaling in TH1 cells. *Science*, 269(5221):245–247, 1995.
- Vanessa M Peterson, Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C Wilson, Renee Moore, Terrill K McClanahan, Svetlana Sadekova, and Joel A Klappenbach. Multiplexed quantification of proteins and transcripts in single cells. *Nature biotechnology*, 35(10):936, 2017.
- Emanuel F Petricoin, Kathryn C Zoon, Elise C Kohn, J Carl Barrett, and Lance A Liotta. Clinical proteomics: translating benchside promise into bedside reality. *Nature reviews Drug discovery*, 1(9):683, 2002.
- Maximilian W Plank, Gerard E Kaiko, Steven Maltby, Jessica Weaver, Hock L Tay, Wei Shen, Mark S Wilson, Scott K Durum, and Paul S Foster. Th22 cells form a distinct th lineage from th17 cells in vitro with unique transcriptional properties and tbet-dependent th1 plasticity. *The Journal of Immunology*, 198(5):2182–2190, 2017.
- Y Poumay, F Dupont, S Marcoux, M Leclercq-Smekens, M Herin, and A Coquette. A simple reconstructed human epidermis: preparation of the culture model and utilization in in vitro studies. *Archives of dermatological research*, 296(5):203–211, 2004.
- Maria Quaranta, Stefanie Eyerich, Bettina Knapp, Francesca Nasorri, Claudia Scarponi, Martina Mattii, Natalie Garzorz, Anna T Harlfinger, Teresa Jaeger, Martine Grosber, et al. Allergic contact dermatitis in psoriasis patients: typical, delayed, and non-interacting. *PloS one*, 9(7): e101814, 2014a.
- Maria Quaranta, Bettina Knapp, Natalie Garzorz, Martina Mattii, Venu Pullabhatla, Davide Pennino, Christian Andres, Claudia Traidl-Hoffmann, Andrea Cavani, Fabian J Theis, et al. Intraindividual genome expression analysis reveals a specific molecular signature of psoriasis and eczema. *Science translational medicine*, 6(244):244ra90–244ra90, 2014b.

BIBLIOGRAPHY

- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <https://www.R-project.org/>.
- Trivellone E Raghunathan, James M Lepkowski, John Van Hoewyk, and Peter Solenberger. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology*, 27(1):85–96, 2001.
- Nuh N Rahbari, O James Garden, Robert Padbury, Mark Brooke-Smith, Michael Crawford, Rene Adam, Moritz Koch, Masatoshi Makuuchi, Ronald P Dematteo, Christopher Christophi, et al. Posthepatectomy liver failure: a definition and grading by the International Study Group of Liver Surgery (ISGLS). *Surgery*, 149(5):713–724, 2011.
- Adaikalavan Ramasamy, Adrian Mondry, Chris C Holmes, and Douglas G Altman. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS medicine*, 5(9):e184, 2008.
- Jean-Marie Ramirez, Nicolò C Brembilla, Olivier Sorg, Rachel Chicheportiche, Thomas Matthes, Jean-Michel Dayer, Jean-Hilaire Saurat, Eddy Roosnek, and Carlo Chizzolini. Activation of the aryl hydrocarbon receptor reveals distinct requirements for IL-22 and IL-17 production by human T helper cells. *European journal of immunology*, 40(9):2450–2459, 2010.
- Daniel Ramsköld, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R Faridani, Gregory A Daniels, Irina Khrebtukova, Jeanne F Loring, Louise C Laurent, et al. Full-length mrna-seq from single-cell levels of rna and individual circulating tumor cells. *Nature biotechnology*, 30(8):777, 2012.
- Petra Rattay, Hans Butschalowsky, Alexander Rommel, Franziska Prütz, Susanne Jordan, Enno Nowossadeck, Olga Domanska, and Panagiotis Kamtsiuris. Inanspruchnahme der ambulanten und stationären medizinischen Versorgung in Deutschland. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 56(5-6):832–844, 2013.
- Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- Aviv Regev, Sarah A Teichmann, Eric S Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna Clatworthy, et al. Science forum: the human cell atlas. *Elife*, 6:e27041, 2017.
- Jette L Riis, Claus Johansen, Christian Vestergaard, Rikke Bech, Knud Kragballe, and Lars Iversen. Kinetics and differential expression of the skin-related chemokines CCL27 and CCL17 in psoriasis, atopic dermatitis and allergic contact dermatitis. *Experimental dermatology*, 20(10):789–794, 2011.

BIBLIOGRAPHY

- 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, 2007.
- 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.
- Julia M Rohrer, Boris Egloff, and Stefan C Schmukle. Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological science*, 28(12):1821–1832, 2017.
- Sergio Romagnani. Regulation of the development of type 2 T-helper cells in allergy. *Current opinion in immunology*, 6(6):838–846, 1994.
- Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- Lothar Sachs and Jürgen Hederich. *Angewandte Statistik: Methodensammlung mit R*. Springer-Verlag, 2009.
- Shimon Sakaguchi, Tomoyuki Yamaguchi, Takashi Nomura, and Masahiro Ono. Regulatory T cells and immune tolerance. *Cell*, 133(5):775–787, 2008.
- Mio Sakuma, Takeshi Morimoto, Kunihiro Matsui, Susumu Seki, Nobuo Kuramoto, Jinichi Toshiro, Junji Murakami, Tsuguya Fukui, Mayuko Saito, Atsushi Hiraide, et al. Epidemiology of potentially inappropriate medication use in elderly patients in Japanese acute care hospitals. *Pharmacoepidemiology and drug safety*, 20(4):386–392, 2011.
- Buka Samten, Susan T Howard, Steven E Weis, Shiping Wu, Homayoun Shams, James C Townsend, Hassan Safi, and Peter F Barnes. Cyclic AMP response element-binding protein positively regulates production of IFN- γ by T cells in response to a microbial pathogen. *The Journal of Immunology*, 174(10):6357–6363, 2005.
- Steffen Sass, Florian Buettner, Nikola S Mueller, and Fabian J Theis. A modular framework for gene set analysis integrating multilevel omics data. *Nucleic acids research*, 41(21):9622–9633, 2013.
- Toshiro Sato, Daniel E Stange, Marc Ferrante, Robert GJ Vries, Johan H Van Es, Stieneke Van Den Brink, Winan J Van Houdt, Apollo Pronk, Joost Van Gorp, Peter D Siersema, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett’s epithelium. *Gastroenterology*, 141(5):1762–1772, 2011.
- Joseph L Schafer and John W Graham. Missing data: our view of the state of the art. *Psychological methods*, 7(2):147, 2002.

BIBLIOGRAPHY

- Juliane Schäfer and Korbinian Strimmer. Learning large-scale graphical gaussian models from genomic data. In *AIP Conference Proceedings*, volume 776, pages 263–276. AIP, 2005.
- KU Schallreuter, Ch Levenig, J Berger, J Umbert, RK Winkelmann, L Wegener, O Correia, O Chosidow, Ph Saiag, S Bastuji-Garin, et al. Severity scoring of atopic dermatitis: the SCORAD index. *Dermatology*, 186(1):23–31, 1993.
- Mark Schena, Dari Shalon, Ronald W Davis, and Patrick O Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467–470, 1995.
- Nathalie Schmitt and Hideki Ueno. Regulation of human helper T cell subset differentiation by cytokines. *Current opinion in immunology*, 34:130–136, 2015.
- Nathalie Schmitt, Yang Liu, Salah-Eddine Bentebibel, Indira Munagala, Laure Bourdery, K Venuprasad, Jacques Banchereau, and Hideki Ueno. The cytokine TGF- β co-opts signaling via STAT3-STAT4 to promote the differentiation of human T FH cells. *Nature immunology*, 15(9):856, 2014.
- Michael P Schön. The plot thickens while the scope broadens: a holistic view on IL-17 in psoriasis and other inflammatory disorders. *Experimental dermatology*, 23(11):804–806, 2014.
- Luca Scrucca, Michael Fop, Thomas Brendan Murphy, and Adrian E. Raftery. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1):205–233, 2017. URL <https://journal.r-project.org/archive/2017/RJ-2017-008/RJ-2017-008.pdf>.
- Stephen F Sells, Sumathi Muthukumar, Vikas P Sukhatme, Scott A Crist, and Vivek M Rangnekar. The zinc finger transcription factor EGR-1 impedes interleukin-1-inducible tumor growth arrest. *Molecular and cellular biology*, 15(2):682–692, 1995.
- Joan Seoane, Hong-Van Le, Lijian Shen, Stewart A Anderson, and Joan Massagué. Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell*, 117(2):211–223, 2004.
- Rajen D Shah and Richard J Samworth. Variable selection with error control: another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):55–80, 2013.
- Alex K Shalek, Rahul Satija, Xian Adiconis, Rona S Gertner, Jellert T Gaublomme, Raktima Raychowdhury, Schraga Schwartz, Nir Yosef, Christine Malboeuf, Diana Lu, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, 498(7453):236, 2013.

BIBLIOGRAPHY

- Hui Shao, Dwight H Kono, Ling-Yu Chen, Elyssa M Rubin, and Jonathan Kaye. Induction of the early growth response (Egr) family of transcription factors during thymic selection. *Journal of Experimental Medicine*, 185(4):731–744, 1997.
- Hyun-Jin Shin, Jee-Boong Lee, Sung-Hwan Park, Jun Chang, and Chang-Woo Lee. T-bet expression is regulated by EGR1-mediated signaling in activated T cells. *Clinical immunology*, 131(3): 385–394, 2009.
- Uri Simonsohn, Joseph P Simmons, and Leif D Nelson. Specification curve: Descriptive and inferential statistics on all reasonable specifications. 2015.
- Tej Pratap Singh, Howard H Zhang, Izabela Borek, Peter Wolf, Michael N Hedrick, Satya P Singh, Brian L Kelsall, Bjorn E Clausen, and Joshua M Farber. Monocyte-derived inflammatory Langerhans cells and dermal dendritic cells mediate psoriasis-like inflammation. *Nature communications*, 7:13581, 2016.
- Henrik Singmann, Ben Bolker, Jake Westfall, and Frederik Aust. *afex: Analysis of Factorial Experiments*, 2018. URL <https://CRAN.R-project.org/package=afex>. R package version 0.21-2.
- D Skiljevic, B Bonaci-Nikolic, D Brasanac, and M Nikolic. Apoptosis of keratinocytes and serum dnase i activity in patients with cutaneous lupus erythematosus: relationship with clinical and immunoserological parameters. *Journal of the European Academy of Dermatology and Venereology*, 31(3):523–529, 2017.
- Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Melius, J., Cirillo, E., Coort, S. L., Digles, D., Ehrhart, F., Giesbertz, P., Kalafati, M., Martens, M., Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L. M. T., Evelo, C. T., Pico, A. R., Willighagen, and E. L. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, 46(D1):D661–D667, Jan 2018.
- Catherine H Smith, Alexander Vincent Anstey, JNWN Barker, AD Burden, RJG Chalmers, DA Chandler, Andrew Yule Finlay, CEM Griffiths, K Jackson, Neil J McHugh, et al. British Association of Dermatologists’ guidelines for biologic interventions for psoriasis 2009. *British Journal of Dermatology*, 161(5):987–1019, 2009.
- Jamie Snider, Max Kotlyar, Punit Saraon, Zhong Yao, Igor Jurisica, and Igor Stagljar. Fundamentals of protein interaction network mapping. *Molecular systems biology*, 11(12):848, 2015.
- Society for Judgment and Decision Making. Publication guidelines: Data and Materials, 2019. URL <http://journal.sjdm.org/>. [Online; accessed 20-February-2019].
- Jaemog Soh, Robert J Donnelly, Serguei Kotenko, Thomas M Mariano, Jeffrey R Cook, Ning Wang, Stuart Emanuel, Barbara Schwartz, Toru Miki, and Sidney Pestka. Identification and sequence

BIBLIOGRAPHY

- of an accessory factor required for activation of the human interferon γ receptor. *Cell*, 76(5):793–802, 1994.
- Owen D Solberg, Edwin J Ostrin, Michael I Love, Jeffrey C Peng, Nirav R Bhakta, Lydia Hou, Christine Nguyen, Margaret Solon, Cindy Nguyen, Andrea J Barczak, et al. Airway epithelial miRNA expression is altered in asthma. *American journal of respiratory and critical care medicine*, 186(10):965–974, 2012.
- Jae W Song and Kevin C Chung. Observational studies: cohort and case-control studies. *Plastic and reconstructive surgery*, 126(6):2234, 2010.
- Richard D Sontheimer. Lichenoid tissue reaction/interface dermatitis: clinical and histological perspectives. *Journal of Investigative Dermatology*, 129(5):1088–1099, 2009.
- Charles Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.
- Hugo Steinhaus. Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci*, 1(804):801, 1956.
- Jonathan AC Sterne, Ian R White, John B Carlin, Michael Spratt, Patrick Royston, Michael G Kenward, Angela M Wood, and James R Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, 338:b2393, 2009.
- Stephen M Stigler. Gauss and the invention of least squares. *The Annals of Statistics*, pages 465–474, 1981.
- Stephen M Stigler. Francis Galton’s account of the invention of correlation. *Statistical Science*, pages 73–79, 1989.
- Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 14(9):865, 2017.
- Mari Strengell, Sampsa Matikainen, Jukka Sirén, Anne Lehtonen, Don Foster, Ilkka Julkunen, and Timo Sareneva. IL-21 in synergy with IL-15 or IL-18 enhances IFN- γ production in human NK and T cells. *The Journal of Immunology*, 170(11):5464–5469, 2003.
- Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):25, 2007.
- Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.

BIBLIOGRAPHY

- Jiuru Sun, Catherina H Bird, Vivien Sutton, Lisa McDonald, Paul B Coughlin, Tanya A De Jong, Joseph A Trapani, and Phillip I Bird. A cytosolic granzyme B inhibitor related to the viral apoptotic regulator cytokine response modifier A is present in cytotoxic lymphocytes. *Journal of Biological Chemistry*, 271(44):27802–27809, 1996.
- Nancy S Sung, William F Crowley Jr, Myron Genel, Patricia Salber, Lewis Sandy, Louis M Sherwood, Stephen B Johnson, Veronica Catanese, Hugh Tilson, Kenneth Getz, et al. Central challenges facing the national clinical research enterprise. *Jama*, 289(10):1278–1287, 2003.
- William R Swindell, Andrew Johnston, Steve Carbajal, Gangwen Han, Christian Wohn, Jun Lu, Xianying Xing, Rajan P Nair, John J Voorhees, James T Elder, et al. Genome-wide expression profiling of five mouse models identifies similarities and differences with human psoriasis. *PLoS one*, 6(4):e18266, 2011.
- Susanne J Szabo, Sean T Kim, Gina L Costa, Xiankui Zhang, C Garrison Fathman, and Laurie H Glimcher. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell*, 100(6):655–669, 2000.
- Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, et al. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, page gkw937, 2016.
- Roba M Talaat, Sara F Mohamed, Iman H Bassyouni, and Ahmed A Raouf. Th1/Th2/Th17/Treg cytokine imbalance in systemic lupus erythematosus (SLE) patients: Correlation with disease activity. *Cytokine*, 72(2):146–153, 2015.
- Chao Tang, Xiaolong Shi, Wei Wang, Dequan Zhou, Jing Tu, Xueying Xie, Qinyu Ge, Peng Feng Xiao, Xiao Sun, and Zuhong Lu. Global analysis of in vivo EGR1-binding sites in erythroleukemia cell using chromatin immunoprecipitation and massively parallel sequencing. *Electrophoresis*, 31(17):2936–2943, 2010.
- Yachao Tao, Menglan Wang, Enqiang Chen, and Hong Tang. Liver regeneration: analysis of the main relevant signaling molecules. *Mediators of inflammation*, 2017, 2017.
- Shigeru Teshima, Hideki Nakanishi, Mikio Nishizawa, Katsuhiko Kitagawa, Masaki Kaibori, Masanori Yamada, Kozo Habara, A-Hon Kwon, Yasuo Kamiyama, Seiji Ito, et al. Up-regulation of IL-1 receptor through PI3K/Akt is essential for the induction of iNOS gene expression in hepatocytes. *Journal of hepatology*, 40(4):616–623, 2004.
- The National Academies of Science, Engineering and Medicine. Press Release: Institute of Medicine to Become National Academy of Medicine, 2015. URL <http://www.nationalacademies.org/hmd/Global/News%20Announcements/IOM-to-become-NAM-Press-Release.aspx>. [Online; accessed 30-October-2018].

BIBLIOGRAPHY

- Henri Theil. *Economic Policy and Forecasting*, 1961.
- Judith Thijs, Todor Krastev, Stephan Weidinger, Constantinus F Buckens, Marjolein de Bruin-Weller, Carla Bruijnzeel-Koomen, Carsten Flohr, and DirkJan Hijnen. Biomarkers for atopic dermatitis: a systematic review and meta-analysis. *Current opinion in allergy and clinical immunology*, 15(5):453–460, 2015a.
- Judith L Thijs, Stefan Nierkens, Athula Herath, CAF Bruijnzeel-Koomen, Edward F Knol, Barbara Giovannone, Marjolein S de Bruin-Weller, and D Hijnen. A panel of biomarkers for disease severity in atopic dermatitis. *Clinical & Experimental Allergy*, 45(3):698–701, 2015b.
- Judith L Thijs, Julia Drylewicz, Renée Fiechter, Ian Strickland, Matthew A Sleeman, Athula Herath, Richard D May, Catharina AFM Bruijnzeel-Koomen, Edward F Knol, Barbara Giovannone, et al. EASI p-EASI: Utilizing a combination of serum biomarkers offers an objective measurement tool for disease severity in atopic dermatitis patients. *Journal of Allergy and Clinical Immunology*, 140(6):1703–1705, 2017a.
- Judith L Thijs, Athula Herath, Marjolein S de Bruin-Weller, and DirkJan Hijnen. Multiplex platform technology and bioinformatics are essential for development of biomarkers in atopic dermatitis. *Journal of Allergy and Clinical Immunology*, 139(3):1065, 2017b.
- Suyan Tian, James G Krueger, Katherine Li, Ali Jabbari, Carrie Brodmerkel, Michelle A Lowes, and Mayte Suárez-Fariñas. Meta-analysis derived (MAD) transcriptome of psoriasis defines the “core” pathogenesis of disease. *PloS one*, 7(9):e44274, 2012.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- Itay Tirosh, Benjamin Izar, Sanjay M Prakadan, Marc H Wadsworth, Daniel Treacy, John J Trombetta, Asaf Rotem, Christopher Rodman, Christine Lian, George Murphy, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science*, 352(6282): 189–196, 2016.
- Harold Tjalsma, Albert Bolhuis, Jan DH Jongbloed, Sierd Bron, and Jan Maarten van Dijk. Signal peptide-dependent protein transport in bacillus subtilis: a genome-based survey of the secretome. *Microbiology and Molecular Biology Reviews*, 64(3):515–547, 2000.
- Vesna Todorovic. Gene expression: Single-cell rna-seq—now with protein. *Nature Methods*, 14(11): 1028, 2017.
- Yoshiki Tokura. Extrinsic and intrinsic types of atopic dermatitis. *Journal of dermatological science*, 58(1):1–7, 2010.
- Claudia Traidl, Silvia Sebastiani, Cristina Albanesi, Hans F Merk, Pietro Puddu, Giampiero Girolomoni, and Andrea Cavani. Disparate cytotoxic activity of nickel-specific CD8+ and CD4+ T cell subsets against keratinocytes. *The Journal of Immunology*, 165(6):3058–3064, 2000.

BIBLIOGRAPHY

- Sara Trifari, Charles D Kaplan, Elise H Tran, Natasha K Crellin, and Hergen Spits. Identification of a human helper T cell population that has abundant production of interleukin 22 and is distinct from T H-17, T H 1 and T H 2 cells. *Nature immunology*, 10(8):864, 2009.
- Robert Choate Tryon. *Cluster analysis: Correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality*. Edwards brother, Incorporated, lithoprinters and publishers, 1939.
- Anastasia Tsanaktsi, Elena E Solomou, and Stamatis-Nick C Liossis. Th1/17 cells, a subset of Th17 cells, are expanded in patients with active systemic lupus erythematosus. *Clinical Immunology*, 195:101–106, 2018.
- Lam C Tsoi, Sarah L Spain, Jo Knight, Eva Ellinghaus, Philip E Stuart, Francesca Capon, Jun Ding, Yanming Li, Trilokraj Tejasvi, Johann E Gudjonsson, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature genetics*, 44(12):1341, 2012.
- John W Tukey. Comparing individual means in the analysis of variance. *Biometrics*, pages 99–114, 1949.
- John Wilder Tukey. The problem of multiple comparisons. *Multiple Comparisons*, 1953.
- Victor LJ Tybulewicz. Vav-family proteins in T-cell signalling. *Current opinion in immunology*, 17(3):267–274, 2005.
- Rajni Upadhyay, Bhavyata Dua, Bhawna Sharma, Mohan Natrajan, Ajai Kumar Jain, Balaji Kithiganahalli Narayanaswamy, and Beenu Joshi. Transcription factors STAT-4, STAT-6 and CREB regulate Th1/Th2 response in leprosy patients: effect of *M. leprae* antigens. *BMC infectious diseases*, 19(1):52, 2019.
- Deepali V Sawant, Weiguo Yao, Zachary Wright, Cindy Sawyers, Robert S Tepper, Sandeep K Gupta, Mark H Kaplan, and Alexander L Dent. Serum microRNA-21 as a biomarker for allergic inflammatory disease in children. *Microrna*, 4(1):36–40, 2015.
- Stef Van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, 16(3):219–242, 2007.
- Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67, 2011. URL <https://www.jstatsoft.org/v45/i03/>.
- Ellen H Van Den Bogaard, Geuranne S Tjabringa, Irma Joosten, Mieke Vonk-Bergers, Esther Van Rijssen, Henk J Tijssen, Mirthe Erkens, Joost Schalkwijk, and Hans JPM Koenen. Crosstalk between keratinocytes and T cells in a 3D microenvironment: a model to study inflammatory skin diseases. *Journal of Investigative Dermatology*, 134(3):719–727, 2014.

BIBLIOGRAPHY

- Leslie van der Fits, Sabine Mourits, Jane SA Voerman, Marius Kant, Louis Boon, Jon D Laman, Ferry Cornelissen, Anne-Marie Mus, Edwin Florenca, Errol P Prens, et al. Imiquimod-induced psoriasis-like skin inflammation in mice is mediated via the IL-23/IL-17 axis. *The Journal of Immunology*, 182(9):5836–5845, 2009.
- BK Van Weemen and AHWM Schuurs. Immunoassay using antigen—enzyme conjugates. *FEBS letters*, 15(3):232–236, 1971.
- Marc Veldhoen, Richard J Hocking, Christopher J Atkins, Richard M Locksley, and Brigitta Stockinger. TGF β in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. *Immunity*, 24(2):179–189, 2006.
- Marc Veldhoen, Catherine Uyttenhove, Jacques Van Snick, Helena Helmbj, Astrid Westendorf, Jan Buer, Bruno Martin, Christoph Wilhelm, and Brigitta Stockinger. Transforming growth factor- β ‘reprograms’ the differentiation of T helper 2 cells and promotes an interleukin 9-producing subset. *Nature immunology*, 9(12):1341, 2008.
- Alexandra-Chloé Villani, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, Andrew Butler, Shiwei Zheng, Suzan Lazo, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, 356(6335):eaah4573, 2017.
- Lauren Vogel. Gut feelings a strong influence on physician decisions, 2018.
- Andrea von Berg, Sibylle Koletzko, Armin Grübl, Birgit Filipiak-Pittroff, H-Erich Wichmann, Carl Peter Bauer, Dietrich Reinhardt, and Dietrich Berdel. The effect of hydrolyzed cow’s milk formula for allergy prevention in the first year of life: the German Infant Nutritional Intervention Study, a randomized double-blind trial. *Journal of Allergy and Clinical Immunology*, 111(3): 533–540, 2003.
- Laura Von Kobyletzki, Åke Svensson, Christian Apfelbacher, and Jochen Schmitt. Factors that predict remission of infant atopic dermatitis: a systematic review. *Acta dermato-venereologica*, 95(4):389–394, 2015.
- Sholom Wacholder, Joseph K McLaughlin, Debra T Silverman, and Jack S Mandel. Selection of controls in case-control studies: I. Principles. *American journal of epidemiology*, 135(9): 1019–1028, 1992.
- Thomas Wallach, Katja Schellenberg, Bert Maier, Ravi Kiran Reddy Kalathur, Pablo Porras, Erich E Wanker, Matthias E Futschik, and Achim Kramer. Dynamic circadian protein–protein interaction networks predict temporal organization of cellular functions. *PLoS genetics*, 9(3): e1003398, 2013.

BIBLIOGRAPHY

- Runmin Wei, Jingye Wang, Erik Jia, Tianlu Chen, Yan Ni, and Wei Jia. GSimp: A Gibbs sampler based left-censored missing value imputation approach for metabolomics studies. *PLoS computational biology*, 14(1):e1005973, 2018.
- Stephan Weidinger and Natalija Novak. Atopic dermatitis. *The Lancet*, 387(10023):1109–1122, 2016.
- Bernard L Welch. The generalization of student’s’ problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, 1947.
- Sally E Wenzel. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nature medicine*, 18(5):716, 2012.
- Adeline R Whitney, Maximilian Diehn, Stephen J Popper, Ash A Alizadeh, Jennifer C Boldrick, David A Relman, and Patrick O Brown. Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences*, 100(4):1896–1901, 2003.
- H-E Wichmann, R Kaaks, Wolfgang Hoffmann, K-H Jöckel, KH Greiser, and J Linseisen. Die Nationale Kohorte. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 55(6-7): 781–789, 2012.
- Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.
- Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.
- Christopher L Williams, Marcia M Schilling, Sung Hoon Cho, Keunwook Lee, Mei Wei, Mark Boothby, et al. STAT4 and T-bet are required for the plasticity of IFN- γ expression across Th2 ontogeny and influence changes in Ifng promoter DNA methylation. *The Journal of Immunology*, page 1203360, 2013.
- Boris J Winterhoff, Makayla Maile, Amit Kumar Mitra, Attila Sebe, Martina Bazzaro, Melissa A Geller, Juan E Abrahante, Molly Klein, Raffaele Hellweg, Sally A Mullany, et al. Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecologic oncology*, 144(3):598–606, 2017.
- Michael Thomas Wong, David Eng Hui Ong, Frances Sheau Huei Lim, Karen Wei Weng Teng, Naomi McGovern, Sriram Narayanan, Wen Qi Ho, Daniela Cerny, Henry Kun Kiaang Tan, Rosslyn Anicete, et al. A high-dimensional atlas of human t cell diversity reveals tissue-specific trafficking and cytokine signatures. *Immunity*, 45(2):442–456, 2016.
- Steven H Woolf. The meaning of translational research and why it matters. *Jama*, 299(2):211–213, 2008.

BIBLIOGRAPHY

- World Health Organisation. The top 10 causes of death, 2018. URL <http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. [Online; accessed 05-November-2018].
- Baolin Wu. Differential gene expression detection and sample classification using penalized linear regression models. *Bioinformatics*, 22(4):472–476, 2005.
- Zhihao Wu, Thomas W Lauer, Anna Sick, Sean F Hackett, and Peter A Campochiaro. Oxidative stress modulates complement factor H expression in retinal pigmented epithelial cells by acetylation of FOXO3. *Journal of Biological Chemistry*, 282(31):22414–22425, 2007.
- Jianguo Xia, Erin E Gill, and Robert EW Hancock. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols*, 10(6):823, 2015.
- Bo Yan, Lei Liu, Shaoqiang Huang, Yan Ren, Huayi Wang, Zhenglin Yao, Lin Li, She Chen, Xiaodong Wang, and Zhiyuan Zhang. Discovery of a new class of highly potent necroptosis inhibitors targeting the mixed lineage kinase domain-like protein. *Chemical Communications*, 53(26):3637–3640, 2017.
- Ence Yang, Gang Wang, Jizhou Yang, Beiyan Zhou, Yanan Tian, and James J Cai. Epistasis and destabilizing mutations shape gene expression variability in humans via distinct modes of action. *Human molecular genetics*, 25(22):4911–4919, 2016.
- Zhao Yang, Chong Li, Zusen Fan, Hongjie Liu, Xiaolong Zhang, Zhiming Cai, Liqin Xu, Jian Luo, Yi Huang, Luyun He, et al. Single-cell sequencing reveals variants in ARID1A, GPRC5A and MLL2 driving self-renewal of human bladder cancer stem cells. *European urology*, 71(1):8–12, 2017.
- K Yoneda, T Demitsu, Y Matsuoka, T Moriue, K Nakai, Y Kusida, R Haba, and Y Kubota. Subcellular activation site of caspase-3 in apoptotic keratinocytes observed in lichenoid tissue reaction. *British Journal of Dermatology*, 158(5):1166–1168, 2008.
- Masato Yoneda, Kaori Suzuki, Shingo Kato, Koji Fujita, Yuichi Nozaki, Kunihiro Hosono, Satoru Saito, and Atsushi Nakajima. Nonalcoholic fatty liver disease: US-based acoustic radiation force impulse elastography. *Radiology*, 256(2):640–647, 2010.
- Ryutaro Yoshiki, Kenji Kabashima, Tetsuya Honda, Satoshi Nakamizo, Yu Sawada, Kazunari Sugita, Haruna Yoshioka, Shun Ohmori, Bernard Malissen, Yoshiki Tokura, et al. IL-23 from Langerhans cells is required for the development of imiquimod-induced psoriasis-like dermatitis by induction of IL-17A-producing $\gamma\delta$ T cells. *Journal of Investigative Dermatology*, 134(7):1912–1921, 2014.
- Di Yu, Sudha Rao, Louis M Tsai, Sau K Lee, Yiqing He, Elissa L Sutcliffe, Monika Srivastava, Michelle Linterman, Lei Zheng, Nicholas Simpson, et al. The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity*, 31(3):457–468, 2009.

BIBLIOGRAPHY

- Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287, 2012. doi: 10.1089/omi.2011.0118.
- Feng-xia Zhan, Juan Li, Min Fang, Juan Ding, and Qian Wang. Importance of Th22 Cell Disequilibrium in Immune Thrombocytopenic Purpura. *Medical science monitor: international medical journal of experimental and clinical research*, 24:8767, 2018.
- Ai-hua Zhang, Hui Sun, Guang-li Yan, Ying Han, and Xi-jun Wang. Serum proteomics in biomedical research: a systematic review. *Applied biochemistry and biotechnology*, 170(4):774–786, 2013.
- Chunhong Zheng, Liangtao Zheng, Jae-Kwang Yoo, Huahu Guo, Yuanyuan Zhang, Xinyi Guo, Boxi Kang, Ruozhen Hu, Julie Y Huang, Qiming Zhang, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell*, 169(7):1342–1356, 2017.
- Wei-ping Zheng and Richard A Flavell. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell*, 89(4):587–596, 1997.
- Jingyi Zhou, Wenjuan Liao, Jing Yang, Ke Ma, Xue Li, Yachen Wang, Donglai Wang, Lina Wang, Yu Zhang, Yuxin Yin, et al. FOXO3 induces FOXO1-dependent autophagy by activating the AKT1 signaling pathway. *Autophagy*, 8(12):1712–1723, 2012.
- Jinfang Zhu, Hidehiro Yamane, and William E Paul. Differentiation of effector CD4 T cell populations. *Annual review of immunology*, 28:445–489, 2009.
- Zhitu Zhu, Diane C Wang, Laurențiu M Popescu, and Xiangdong Wang. Single-cell transcriptome in the identification of disease biomarkers: opportunities and challenges, 2014.
- Christina Zielinski, Fabian Theis, and Subarna Palit. Meeting the challenges of high-dimensional data analysis in immunology. *bioRxiv*, page 473215, 2018.
- Christina E Zielinski. Human T cell immune surveillance: phenotypic, functional and migratory heterogeneity for tailored immune responses. *Immunology letters*, 190:125–129, 2017.
- Donald W Zimmerman. A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1):173–181, 2004.
- Ulrich M Zissler, Moritz Ulrich, Constanze A Jakwerth, Sandra Rothkirch, Ferdinand Guerth, Markus Weckmann, Matthias Schiemann, Bernhard Haller, Carsten B Schmidt-Weber, and Adam M Chaker. Biomatrix for upper and lower airway biomarkers in patients with allergic asthma. *Journal of Allergy and Clinical Immunology*, 142(6):1980–1983, 2018.
- Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

BIBLIOGRAPHY

Joseph Zubin. A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology*, 33(4):508, 1938.

Yiming Zuo, Yi Cui, Guoqiang Yu, Ruijiang Li, and Habtom W Resson. Incorporating prior biological knowledge for network-based differential gene expression analysis using differentially weighted graphical lasso. *BMC bioinformatics*, 18(1):99, 2017.