

Analyse und Visualisierung von Effekten in genomweiten Expressionsdaten

Olivia Prazeres da Costa



Dissertation an der Medizinischen Fakultät der Ludwig-Maximilians-Universität

**Aus dem Institut für Medizinische Informationsverarbeitung, Biometrie und
Epidemiologie**

**Aus dem Institut für Medizinische Informationsverarbeitung, Biometrie und
Epidemiologie der Universität München
Vorstand: Prof. Dr. rer. nat. Ulrich Mansmann**

**Analyse und Visualisierung von Effekten in genomweiten
Expressionsdaten**

Dissertation

**Zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München**

vorgelegt von

Olivia Prazeres da Costa

aus München

2014

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatter: Prof. Dr. rer. nat. Ulrich Mansmann

Mitberichterstatter: Prof. Dr. Karl-Hans Englmeier

Mitbetreuung durch den
Promovierten Mitarbeiter:

Dekan: Prof. Dr. med. Dr. h. c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 16.12.2014

Abstract

Einleitung: Modell-basierte Vorhersagen für molekulare Netzwerke und zelluläre Interaktionen können durch zwei verschiedene Strategien der Systembiologie getroffen werden, die top-down und bottom-up Strategien. Die bottom-up Strategie beginnt bei *a priori* Wissen über einzelne Grundelemente und fügt diese zu größeren Einheiten wie Signalwegen oder ganzen Systemen zusammen. Top-down Strategien setzen bei Datensätzen eines Systems an und versuchen Netzwerke, Interaktionen oder Komponenten zu identifizieren, die für das Systemverhalten (z.B. Phänotyp) verantwortlich sind. Im Folgenden werden beide Strategien auf unterschiedliche Transkriptionsdaten angewendet und die Ergebnisse visualisiert. Beide Strategien können auf linearen Regressionsmodellen basieren. In dieser Arbeit werden lineare Regressionsmodelle höherer Ordnung mittels eines neuen visuellen Hilfsmittels, des Eruptionsdiagramms, verglichen.

Methodik: Eruptionsdiagramme werden durch die Überlagerung zweier Vulkandiagramme erstellt. Beide Vulkandiagramme werden von derselben Datengrundlage generiert, stammen jedoch von zwei verschiedenen Modellen. Jedes Gen wird von einem Pfeil repräsentiert, welcher bei dem Punkt des Vulkandiagramms von Modell 1 startet und bei dem Punkt des Vulkandiagramms aus Modell 2 endet. Im Rahmen der Modellselektion können Eruptionsdiagramme als visuelles Hilfsmittel verwendet werden, um (ir)relevante Kovariaten, Störfaktoren und Effektmodifikation aufzudecken.

Ergebnisse: Es werden zwei verschiedene Transkriptionsdatensätze analysiert: ein Maus-Infektionsdatensatz und ein humaner Asthmadatensatz. Für die Analyse des Infektionsdatensatzes werden verschiedene lineare Regressionsmodelle miteinander verglichen. Durch eine rückwärts-gewandte Modellselektionsstrategie wird gezeigt, dass durch die Infektionskovariaten erster Ordnung zusätzliche erklärende Kraft gewonnen wird. Durch das Eruptionsdiagramm werden Effekte zweiter Ordnung aufgedeckt. Ein Modellvergleich identifiziert die Kovariaten dritter Ordnung als Störfaktoren. Das Modell zweiter Ordnung, welches am besten zu den Daten passt, wird für die weiterführende Analyse verwendet. Die Ergebnisse der Interaktionskovariate werden in aggravating und alleviating Effekte unterteilt. Ein Interaktionseffekt ist alleviating (aggravating, neutral), falls der Effekt der kombinierten Kovariaten schwächer (stärker, identisch) als die Summe der individuellen Effekte dieser Kovariaten ist. Bei der bottom-up Analyse des Asthmadatensatzes werden die Daten nicht auf Einzelgenebene sondern auf Gengruppenebene analysiert. Zunächst wird das passende Regressionsmodell mit Hilfe des Eruptionsdiagramms aufgestellt. Der Einfluss der einzelnen Gene auf das globale Testergebnis der Gengruppen wird in diagnostischen Balkendiagrammen genauer untersucht. Eine Signalweganalyse der Gengruppen zeigt neue Biomarker und Signalwege für die Charakterisierung von allergischem und nicht-allergischem Asthma auf.

Diskussion: Die Ergebnisse der Transkriptionsanalyse werden durch Anreicherungsanalysen auf ihre funktionelle Relevanz hin untersucht. Die Ergebnisse zeigten unterschiedliche funktionelle Eigenschaften der aggravating und alleviating Gene auf. Die Anreicherungsanalyse des Asthmadatensatzes der Gene, die von Störfaktoren beeinflusst werden und durch Effektmodifikation gekennzeichnet sind, weisen jedoch keine funktionellen Unterschiede auf.

Abstract

Introduction: Model-based prediction of molecular networks and cellular interactions can be identified by two different strategies of systems biology, top-down and bottom-up strategies. The bottom-up strategy starts at *a priori* knowledge about single elements and merges into more complex units like signalling pathways or whole systems. Top-down strategies explore datasets of a system and try to identify networks, interactions or components responsible for the system behaviour (i.e. phenotype). In this thesis both strategies are applied to different transcription datasets and the corresponding results are visualized. The strategies can be based on linear regression models. In this work higher order regression models are compared using a new visual tool, the eruption plot.

Methods: Eruption plots are generated by overlapping two volcano plots. Both volcano plots are based on the same data, but originate from two different models. Each gene is symbolized by an arrow, which connects the point from the volcano plot of the first model to the point from the volcano plot of the second model. The eruption plot is a visual supporting tool in model selection revealing (ir)relevant covariates, confounding factors and effect modification.

Results: Two different transcription datasets are analysed in this work, a mouse-infection dataset and a human asthma dataset. For the analysis of the infection dataset two different linear regression models are being compared. As part of a backward driven model selection strategy the infection covariates provide additional explanatory power. The eruption plot highlights second order effects. A model comparison reveals third order covariates as confounding factors. The second order model that matches the data best is applied in the subsequent analysis. The results of the interaction covariates are divided into aggravating and alleviating effects. An interaction effect is alleviating (aggravating, neutral) if the effect of both covariates is lower (higher, identical) than the sum of both individual effects of these covariates. Within the bottom-up analysis of the asthma dataset the genes are not analysed on the single gene level but on the level of gene groups. By application of an eruption plot the adequate regression model is selected. Diagnostic bar plots help to further investigate the influence of the single gene on the global test result of the gene group. A pathway analysis of the gene groups shows new biomarkers and signalling pathways to characterize allergic and non-allergic asthma.

Discussion: The functional relevance of both types of analysis is explored in detail through enrichment analysis. The results of the aggravating and alleviating genes show distinct functional properties. Genes of the asthma dataset are divided into genes influenced by confounding factors and effect modification. The enrichment analysis of these two groups, however, reveals no functional differences.

Inhaltsverzeichnis

1	Einleitung.....	1
1.1.	Strategien für die Identifizierung von Interaktionen.....	1
1.1.1	Top-down Strategie	1
1.1.2	Bottom-up Strategie.....	2
1.2	Lineare Regressionsmodelle höherer Ordnung in Genexpressionsdaten	2
1.3	Infektionsdatensatz	4
1.4	Asthma.....	5
1.5	Forschungsziel	6
2	Material und Methoden.....	7
2.1	Eruptionsdiagramme	7
2.2	Globaler Test.....	11
2.3	Cluster, Signalweg und Transkriptionsfaktor-Bindestellenanalyse	12
2.4	Experimenteller Aufbau der <i>Yersinien</i> Infektionsdaten.....	13
2.5	Experimenteller Aufbau der Asthmadaten.....	13
2.6	Microarray Experimente sowie Normalisierung und Filterung.....	14
2.7	Anreicherungsanalyse	15
3	Ergebnisse	16
3.1	Analyse der Infektionsdaten.....	16
3.1.1	Kovariaten des Infektionsdatensatzes.....	16
3.1.2	Transkriptionelle Effekte durch Interaktionsmodelle.....	16
3.1.3	Top-down Analyse von geschachtelten Modellen durch Eruptionsdiagramme	17
3.1.4	Bestimmung von beschreibender und erklärender Kraft der Kovariaten.....	19
3.1.5	Aufdeckung funktionaler Gencluster durch Interaktionsanalyse.....	21
3.1.6	Aufdeckung biologisch relevanter Merkmale durch Interaktionsanalyse	23
3.2	Analyse der Asthmadaten	25
3.2.1	Batch Effekte	25
3.2.2	Auswahl des Modells basierend auf dem Eruptionsdiagramm	27
3.2.3	Top-down Analyse: Differentiell exprimierte Gene zwischen AA, NAA und GK.....	29

3.2.4	Bottom-up Analyse: Differentiell regulierte Signalwege zwischen AA, NAA und GK	30
3.2.4.1	Differentiell regulierte Signalwege.....	35
3.2.4.2	Differentielle Genregulation von T-Zellgruppen	41
3.2.5	Differentielle Genregulation in CD4 ⁺ -Zellen.....	43
3.2.5.1	Top-down Analyse der CD4 ⁺ -Zellen.....	43
3.2.5.2	Bottom-up Analyse der CD4 ⁺ -Zellen.....	43
4	Diskussion.....	46
4.1	Verfahren zur Bestimmung von differentieller Genexpression: Evidenzbasierte Verfahren und Erstellung von Genranglisten.....	46
4.2	Modellselektionskriterien in Verbindung mit dem Eruptionsdiagramm	47
4.3	Pipelines zur Analyse differentieller Genexpression	48
4.3.1	Explorative bzw. top-down Verfahren	48
4.3.1.1	Gen-weise Analyse von differentieller Expression.....	49
4.3.1.2	Klassifikation von Genen	49
4.3.2	Konfirmatorische Gengruppenanalysen bzw. bottom-up Verfahren.....	50
4.4	Interpretation des Eruptionsdiagramms	51
4.5	Anreicherungsanalysen der Ergebnisse der Infektionsdaten	52
4.5.1	Signifikante Gene der Interaktionskovariate.....	52
4.5.2	Abweichung der zu erwartenden Effekte.....	54
4.5.2.1	Aggravating Effekte.....	55
4.5.2.2	Alleviating Effekte.....	56
4.5.3	Vergleich der Ergebnisse: unterschiedliche Ergebnisse der alleviating und aggravating Datensätze.....	58
4.6	Anreicherungsanalyse der Gene des Asthmadatensatzes.....	61
4.6.1	Alle signifikanten Gene der Kovariate AA	61
4.6.2	Von einem Störfaktor beeinflusste Gene	62
4.6.3	Gene, die bei dem Modellvergleich keine zusätzliche erklärende Kraft der Interaktionskovariate zeigen	63

4.6.4	Vergleich der Ergebnisse der Anreicherungsanalysen der Asthmadaten.....	64
4.7	Biologische Relevanz der Ergebnisse der Infektionsdatensatzanalyse	65
4.8	Biologische Interpretation der Ergebnisse der linearen Regressionsanalyse der Asthmadaten	66
5	Zusammenfassung und Ausblick	68
6	Anhang	70
6.1	Abbildung der Funktionsanalyse der Cluster von Modell 3.....	70
6.2	Tabelle der alleviating und aggravating Interaktionen.....	71
6.3	Anreicherungsanalyse der Interaktionskovariate	80
6.4	Abbildung des Eruptionsdiagramms von B.....	84
6.5	Legende der Ingenuity Signalwege	85
7	Quellenverzeichnis.....	86
8	Abkürzungsverzeichnis	91
9	Abbildungsverzeichnis.....	93
10	Tabellenverzeichnis	95
11	Danksagung.....	96
12	Lebenslauf.....	97

1 Einleitung

1.1. Strategien für die Identifizierung von Interaktionen

Die Systembiologie zielt darauf ab, ein ganzheitliches Bild von biologischen Funktionen zu geben. Dabei kommen zwei grundlegende Strategien der Systembiologie zum Einsatz, um zelluläre Interaktionen zu identifizieren und Modell-basierte Vorhersagen für molekulare Netzwerke zu treffen (Schneider and Klabunde 2013), einerseits die top-down Strategien und andererseits die bottom-up Strategien. Die bottom-up Strategie startet bei *a priori* Wissen über einzelne Grundelemente und fügt diese zu größeren Einheiten wie Signalwegen oder ganzen Systemen zusammen. Top-down Strategien starten bei Datensätzen eines Systems und versuchen Netzwerke, Interaktionen oder Komponenten zu identifizieren, die für das Systemverhalten (z.B. den Phänotyp) verantwortlich sind. In dieser Arbeit werden beide Strategien auf unterschiedliche Transkriptionsdaten angewendet.

1.1.1 Top-down Strategie

Hochdurchsatz-Technologien führen in der Molekularbiologie zu einer Zunahme von Informationen auf Systemebene. Dabei handelt es sich um Informationen über Genome, Transkriptome, Proteome, Metabolome und Transkriptionsfaktor-Bindestellen (TFBS) Profile (Lanza, Crook et al. 2012).

In einem top-down Ansatz wird ein System aufgebrochen, um die Zusammensetzung der Subsysteme zu verstehen (Abbildung 1). Jedes Subsystem wird wiederum detaillierter zerlegt, bis das Ausgangssystem in seinen Grundelementen dargestellt wird. Dieses Vorgehen ist hauptsächlich Computer-basiert und ist nicht abhängig von bestehendem Wissen über molekulare Funktionen und deren Komponenten. Dabei werden Netzwerkbeziehungen von konsistent korrelierten Veränderungen zwischen Komponenten abgeleitet. Z.B. werden die molekularen Profile von Krebszellen mit entsprechenden Kontrollen verglichen. Prinzipiell kann dieser Ansatz Netzwerke aufbauen. Die Beziehungen der Netzwerke bestehen aus rückgefolgerten Interaktionen zwischen zellulären Komponenten, und zwar unabhängig davon, ob sie auf der molekularen Ebene beschreibbar sind oder nicht.

Top-down Methoden der Systembiologie beginnen mit Datensätzen, die ein biologisches System repräsentieren (Microarrays) und versuchen Subsysteme, Signalwege oder Komponenten zu identifizieren, die verantwortlich für das Systemverhalten sind (z.B. den Phänotyp; Interaktion von Mausstamm und Interferon-gamma- (IFN- γ) Stimulation). In dieser Arbeit werden mit Hilfe eines Infektionsdatensatzes Regressionsmodelle höherer Ordnung angewandt, um Gen-weise relevante Interaktionen zwischen biologischen Komponenten zu identifizieren.

1.1.2 Bottom-up Strategie

Bottom-up Strategien (McDonald 2011) verbinden einzelne Systeme zu einem übergreifenden, komplexeren System (Abbildung 1). Die individuellen Grundelemente werden zuerst detailliert untersucht. Diese Elemente werden zusammengeführt um größere Subsysteme zu erstellen, die wiederum zusammengeführt werden, bis ein komplettes System geformt ist. Interaktive Netzwerke werden von *a priori* Wissen über molekulare Interaktionen der Komponenten erstellt. Der Fokus liegt auf bekannten Einheiten, basierend auf bekannten und experimentell verifizierten Strukturen. Z.B. sind *Wnt*, *Notch*, *Hedgehog* Signalwege identifiziert worden durch die interagierenden Gene bei der *Drosophila* Entwicklung.

Bottom-up Methoden starten oft bei detailliertem Wissen über einzelne Komponenten und fügen diese zu größeren Einheiten zusammen. So dient z.B. GlobalAncova (Hummel, Meister et al. 2008) als ein Beispiel einer bottom-up Methode. GlobalAncova ist ein globaler Test auf differentielle Genexpression von Gengruppen. Dabei wird das Vorwissen über Gengruppen auf Transkriptionsdaten angewandt, um Interaktionen zwischen Phänotypen zu untersuchen. In der vorliegenden Arbeit wird die Methode GlobalAncova auf einen Asthma-Transkriptionsdatensatz angewendet. Hier sind Gengruppen zum Teil vorgegeben und neue Gengruppen von Immunzellen werden auf den globalen Test angewendet.

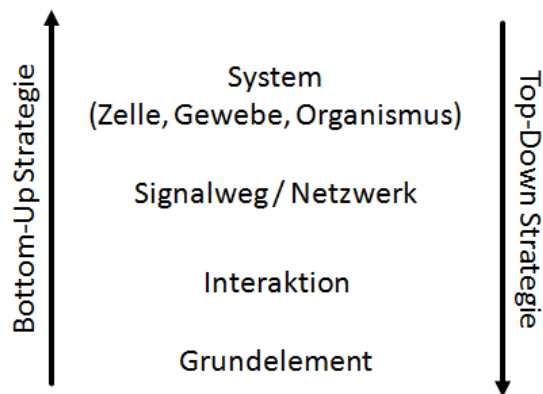


Abbildung 1: Entgegengesetzte Richtungen der Strategien in der Systembiologie.

1.2 Lineare Regressionsmodelle höherer Ordnung in Genexpressionsdaten

Genexpression ist das Ergebnis einer Vielzahl von verschiedenen Mechanismen, deren Effekte sich nicht einfach addieren lassen, da sie komplexe Interaktionen zeigen. Um die Biologie der sich verändernden, zu Grunde liegenden Genexpression zu verstehen, bedarf es passender Methoden. In diesem Kapitel werden einfache Werkzeuge präsentiert, um die Herausforderungen anzugehen. Dies wird anhand eines Beispieldatensatzes dargestellt, bei dem die transkriptionelle Antwort von zwei Mausstämmen auf IFN- γ -Stimulation untersucht wird.

Interaktionen höherer Ordnung und komplexe Effektmodifikationen können typischerweise in veränderter Aktivität von Transkriptionsfaktoren (TF) resultieren. In der Regel wird die Analyse transkriptioneller Regulation auf dem Level individueller TF-Ziel-Gene durchgeführt. Durch Genom-weite Transkriptionsmessungen mittels Microarrays kann ein umfassenderes Bild der Signalprozesse wiedergegeben werden. Der am häufigsten verwendete Standard für die Analyse von Transkriptionsdaten ist die Anwendung von linearer Regression. Dies ist zum Beispiel im Bioconductor Paket Limma (Smyth 2004) implementiert. Durch lineare Regression werden einzelne Effekte bestimmter Faktoren, sogenannter Kovariaten, auf die Genexpression bestimmt. Beispiele für Kovariaten sind Gendeletionen, Umwelteinflüsse, Stress oder Zytokin-Stimulationen. Normalerweise wird angenommen, dass die Kovariaten unabhängig (z.B. additiv) zur Genexpression beitragen. Diese Art der Analyse ist sehr erfolgreich, obwohl sie eine ungerechtfertigte Vereinfachung konstituiert und die Annahme der Additivität oft falsch ist. Die extremsten Beispiele solcher Verletzungen sind die so genannten synthetisch lethalen Interaktionen, bei dem zwei separate Gendeletionen zu keinem Phänotypen führen. Die doppelte Gendeletion ist jedoch lethal (Pan, Ye et al. 2006; Collins, Miller et al. 2007). Nicht-Additivität kann sowohl auf der Phänotyp-Ebene als auch auf Einzel-Genexpressionsebene auftreten. Biologisch deutet dies auf eine Interaktion der Kovariaten auf der Ebene des transkriptionellen Netzwerkes hin. Aber wie lässt sich ein solches Zusammenspiel zwischen Kovariaten schnell identifizieren? Klassische Methoden, wie R^2 und das *Akaike information criterion* (AIC), vergleichen nicht die Effektschätzer zwischen verschiedenen Modellen einer spezifischen Kovariate. Daher wird an dieser Stelle das Eruptionsdiagramm (Prazeres da Costa, Hoffman et al. 2014) eingeführt, ein einfaches Werkzeug zur Identifizierung und Visualisierung von Interaktionen. Es ist dabei behilflich, zu Grunde liegende nicht-additive biologische Mechanismen aufzudecken. Durch die Hinzunahme von Interaktionstermen zum linearen Modell können deren Größe und Signifikanz bestimmt werden. Dadurch ist erkennbar, ob die Genexpression von einem simplen additiven Effekt, der durch ein Modell erster Ordnung (additives Modell) von Haupteffekten errechnet wird, abweicht. Ein Interaktionseffekt abweichend von Null ergibt, dass ein additives Modell bestehend aus Haupteffekten unvollständig ist. Interaktionen können entsprechend der Ergebnisse der Haupt- und Interaktionseffekte in eine der zwei folgenden Gruppen eingeteilt werden (Abbildung 2) (Beyer, Bandyopadhyay et al. 2007): Ein Interaktionseffekt zwischen zwei Kovariaten ist alleviating (aggravating, neutral), falls der Effekt der kombinierten Kovariaten schwächer (stärker, identisch) als die Summe der individuellen Effekte dieser Kovariaten ist.

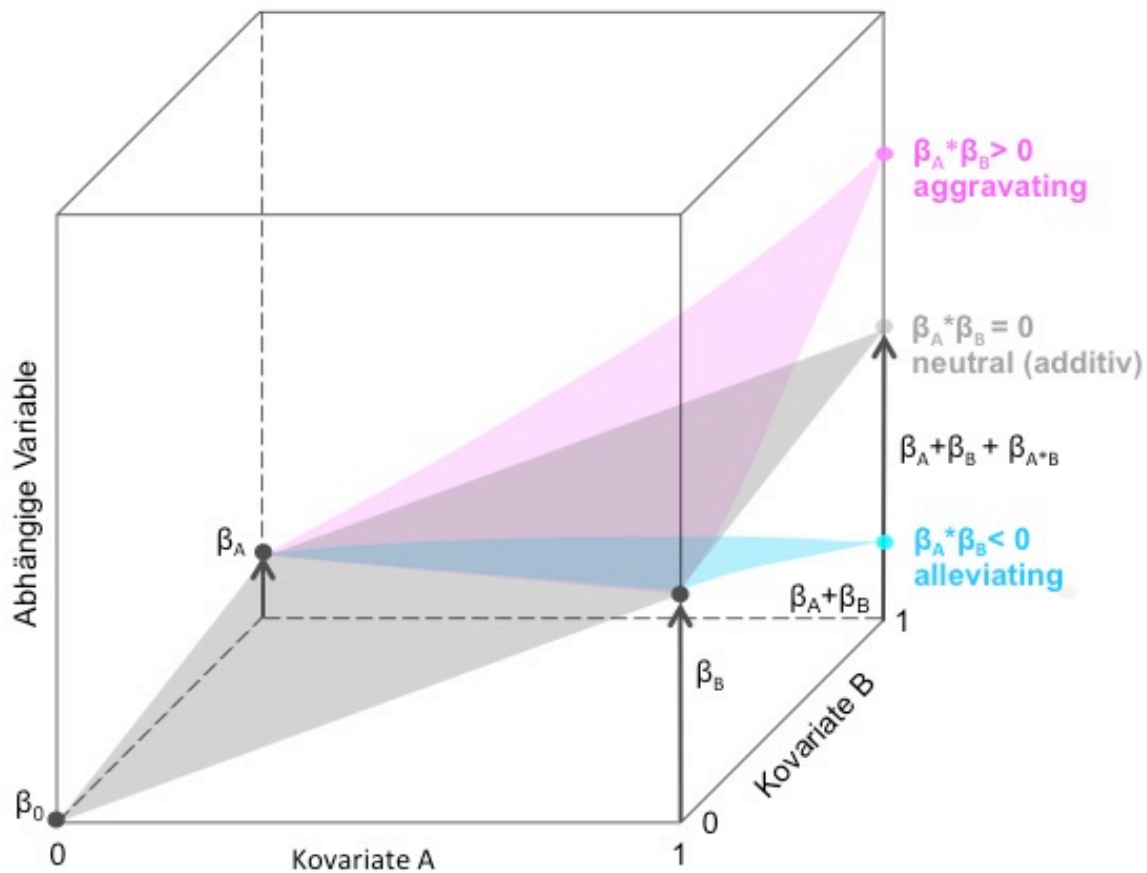


Abbildung 2: Schematische Darstellung von aggravating und alleviating Effekten.

Die Effektschätzer β können entweder in einem additiven (Kovariaten A und B) oder multiplikativen (Interaktionskovariate A*B) Modell errechnet werden. Die Abweichung vom additiven (grau) Modell kann positiv (pink) oder negativ (türkis) sein.

Einfache Interaktionsmodelle werden zur Analyse von Effekten kombinierter Gendeletionen (Capaldi, Kaplan et al. 2008; Spitzer, Griffiths et al. 2011) und zur Untersuchung von Arznei-Arznei- und Arznei-Gen-Interaktionen (Warringer, Anevski et al. 2008; Jaimovich, Rinott et al. 2010; Zhang, Vaga et al. 2011) verwendet. Dies sind Beispiele experimentellen Aufbaus, die einen spezifischen Faktor verwenden, z.B. doppelte Gendeletionen oder zweifache Arznei-Behandlung.

1.3 Infektionsdatensatz

Für die Entwicklung und zum Test unserer Methode wird ein dreifach faktorieller Datensatz verwendet. Der Datensatz ist öffentlich verfügbar (GEO accession Nr.: GSE 9273). In dieser Transkriptionsstudie werden die Effekte einer *in vitro* Infektion von Maus-Makrophagen in zwei verschiedenen C57BL/6 und BALB/c Mausstämmen untersucht (van Erp, Dach et al. 2006). Zwei

verschiedene infektiöse Stämme des intrazellulären Bakteriums *Yersinia enterocolitica* werden in An- und Abwesenheit des Zytokins IFN- γ verwendet (Tabelle 1).

	Control strain WA(pTTS, pP60)	Virulent strain WA(pYV)	Mock
C57BL/6	3	3	3
BALB/c	3	3	4
C57BL/6 + IFN- γ	3	3	3
BALB/c + IFN- γ	3	3	4

Tabelle 1: Experimenteller Aufbau und Anzahl der Replikate des Infektionsdatensatzes.

Die drei zu untersuchenden Kovariaten sind folglich der genetische Hintergrund (C57BL/6 und BALB/c), die Zytokin-Stimulation (+/-IFN- γ) und die bakterielle Infektion der Stämme WA(pTTS, p60) und WA(pYV). Es wird eine Analysestrategie bereitgestellt, um relevante Interaktionen zwischen Kovariaten aufzuzeigen. Diese Analyse zeigt die Relevanz der Aufdeckung von Effektmodifikation und Störfaktoren, welche generell nicht bei Modellen erster Ordnung untersucht werden. Daher wird eine falsche Interpretation von transkriptionellen Effekten vermieden.

1.4 Asthma

Asthma ist eine komplexe, chronisch inflammatorische Atemwegserkrankung und ist klinisch durch Keuchen, variablen Lungenfluss, Bronchialobstruktion, dauerhaft bestehende Überempfindlichkeit der Atemwege sowie Verengung der Atemwege, Verkrampfung der Bronchialmuskulatur und Bildung von Ödemen der Bronchialschleimhaut (Hogate, 2007) charakterisiert. Asthma beginnt meist in der Kindheit, ausgelöst durch genetische und Umweltfaktoren oder durch ein Zusammenspiel der beiden Faktoren (Vercelli D 2010). Kindheitsasthma hat in den letzten Jahrzehnten stetig zugenommen, besonders in westlichen Ländern (Weinberg 2000).

Asthma wird in allergisches Asthma (AA) und nicht-allergisches Asthma (NAA) unterteilt (Roanet-Manet 2002). Die Unterscheidung basiert auf dem Vorkommen von allergischen Symptomen und von spezifischer Sensibilisierung auf Allergene. Bisher ist die zugrundeliegende Immunregulation zwischen den zwei Phänotypen und gesunden Kindern unklar (Rädler 2014).

Die Heterogenität der Asthmasymptome, die wahrscheinlich auf einer Dysregulation des komplexen Zusammenspiels von Signalwegen basiert, ist eine Herausforderung für die

Identifizierung der zugrundeliegenden Immunpathologie. In Kooperation mit der Kinderklinik und Kinderpoliklinik im Dr. Haunerschen Kinderspital im Rahmen des SFB-TR22 unter der Leitung von Bianca Schaub sollen mit Hilfe der Microarray Technologie neue Gene und Signalwege identifiziert werden, um AA und NAA zu unterscheiden (Raedler, Prazeres da Costa, Ballenberger et al., in submission). Die beiden Phänotypen haben gemeinsame Merkmale, wie z.B. Keuchen, erhöhte Überempfindlichkeit der Atemwege und Luftnot, weisen aber auch unterschiedliche Merkmale auf wie allergische Inflammation. Wahrscheinlich haben die zu Grunde liegenden Immunmechanismen gemeinsame und unterschiedliche Signalwege. Durch bottom-up Methoden sollen differentiell exprimierte Gene und Signalwege in AA und NAA bestimmt werden.

1.5 Forschungsziel

Das Forschungsziel dieser Arbeit ist die Visualisierung komplexer Analysen von hochdimensionalen Transkriptionsdatensätzen.

Für die Analyse von hochdimensionalen Transkriptionsdatensätzen können zwei verschiedene Strategien angewendet werden (Abbildung 1), die top-down und bottom-up Strategie. Ein Schritt dieser Strategien ist die Modellselektion. Es soll gezeigt werden, wie das Eruptionsdiagramm die Ergebnisse des globalen Tests GlobalAncova ergänzen kann. GlobalAncova berechnet einen p-Wert für den Vergleich von zwei Modellen. Der p-Wert ist sehr niedrig, wenn das komplexere Modell die hochdimensionalen Daten erklärt. Das Eruptionsdiagramm dagegen visualisiert die Gene, für welche das komplexere Modell besonders wichtig ist. Das Eruptionsdiagramm wird also als neues visuelles Hilfsmittel für Modellselektion eingeführt. Die Anwendung des Eruptionsdiagramms wird am Beispiel von zwei verschiedenen Transkriptionsdatensätzen demonstriert und die Ergebnisse der Transkriptionsanalysen mittels bekannter Diagramme dargestellt.

2 Material und Methoden

2.1 Eruptionsdiagramme

Um die Effektgröße (z.B. Genexpression) und Signifikanz (p-Werte einer verwandten Teststatistik) einer bestimmten Variable A für eine hohe Anzahl von Objekten (z.B. Genen) darzustellen, werden normalerweise Vulkandiagramme verwendet. Dabei wird jedes Objekt durch einen Punkt repräsentiert. Die Effektgröße wird auf der x-Achse (z.B. Expression auf einer \log_2 -Skala) und die Signifikanz auf der y-Achse (p-Werte auf einer \log_{10} -Skala) angetragen (Li 2012). Das Eruptionsdiagramm (Prazeres da Costa, Hoffman et al. 2014) ist eine Überlagerung zweier Vulkandiagramme der Variable A . Beide Vulkandiagramme werden von derselben Datengrundlage generiert, stammen jedoch von zwei verschiedenen Modellen, Modell 1 und Modell 2 (Abbildung 3). Jedes Objekt wird von einem Pfeil repräsentiert, welcher bei dem Punkt des Vulkandiagramms von Modell 1 startet und bei dem Punkt des Vulkandiagramms aus Modell 2 endet. Die blauen Linien in dem Diagramm entsprechen dem Grenzwert der BH-korrigierten p-Werte von 0.05 (in $-\log_{10}$) und dem Effektschätzer von 1.5 (in \log_2). Die drei Linien teilen das Diagramm in sechs Felder ein. Die oberen rechten und linken Felder sind die Regionen von Interesse (ROI), da sie die Gene von statistischer Signifikanz und hohen Effektschätzern enthalten. Im Folgenden wird dargestellt, wie Eruptionsdiagramme verwendet werden, um im Rahmen von Modellselektion (ir)relevante Kovariaten, Störfaktoren und Effektmodifikation aufzudecken.

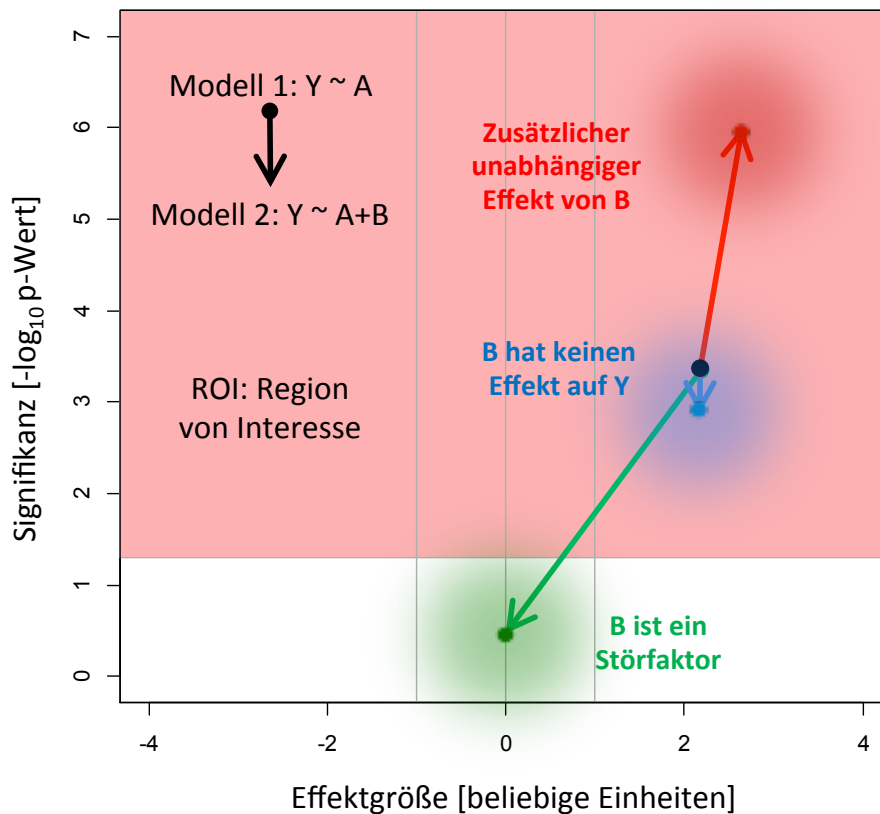


Abbildung 3: Schematische Darstellung der Interpretationsmöglichkeiten (rot, blau, grün) des Eruptionsdiagramms.

Die Ergebnisse von zwei Modellen können anhand des Eruptionsdiagramms interpretiert und dargestellt werden. Dabei können die Pfeile des Eruptionsdiagramms unterschiedliche Längen und Richtungen aufweisen. Dies hilft, um den Vergleich der beiden Modelle zu interpretieren (farblich neben den Pfeilen).

Es soll angenommen werden, dass Modell 1 ein lineares Regressionsmodell der abhängigen Variable Y gegen die Kovariate A ist, abgekürzt: $Y \sim A$. Eine Variable B , die einen zusätzlichen Effekt unabhängig von A hat, erhöht die erklärende Kraft des erweiterten Modells 2, $Y \sim A+B$, im Vergleich zu Modell 1, indem es z.B. die unerklärte Varianz („Noise“) reduziert. Daher wird die Signifikanz eines potentiellen Effekts in A erhöht, während die Effektgröße von A nahezu nicht betroffen ist. Das Eruptionsdiagramm von A stellt daher einen langen Pfeil dar, der nahezu gerade nach oben zeigt (Abbildung 3). Variable B ist somit eine relevante Kovariate. Wenn andererseits B keinen zusätzlichen Effekt hat, wird die Effektgröße von A reduziert und daher auch die Signifikanz von A . In diesem Fall zeigt die Richtung des Pfeils nach unten und leicht gegen die y-Achse zeigen (Abbildung 3), also ist B eine irrelevante Kovariate. Die Darstellung einer Anwendung auf mehrere Objekte findet sich in Abbildung 4.

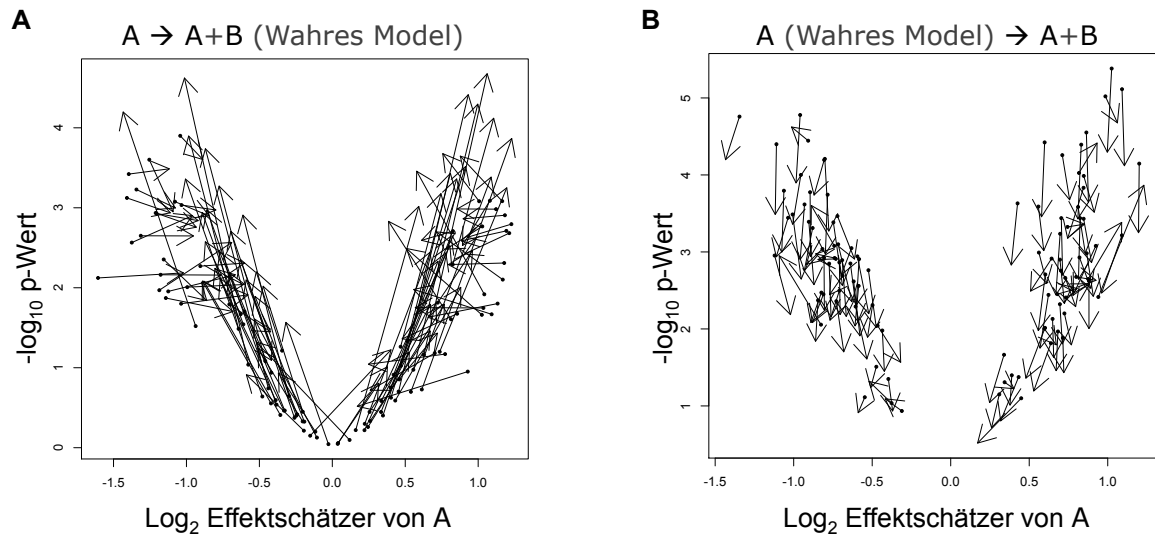


Abbildung 4: Eruptiondiagramme angewendet auf einen Simulationsdatensatz. A: Komponente B ist relevant; B: Komponente B ist irrelevant.

Ein Störfaktor beschreibt fälschlicherweise eine Assoziation zwischen abhängiger und unabhängiger Variable (Breslow and Day 1980). Dies wird durch die Assoziation einer versteckten Variablen (des Störfaktors) mit der abhängigen und unabhängigen Variablen verursacht. Der Störfaktor muss nicht auf dem kausalen Weg der abhängigen und unabhängigen Variablen liegen. Abbildung 5A stellt das Konzept eines Störfaktors anhand eines Eruptiondiagramms dar.

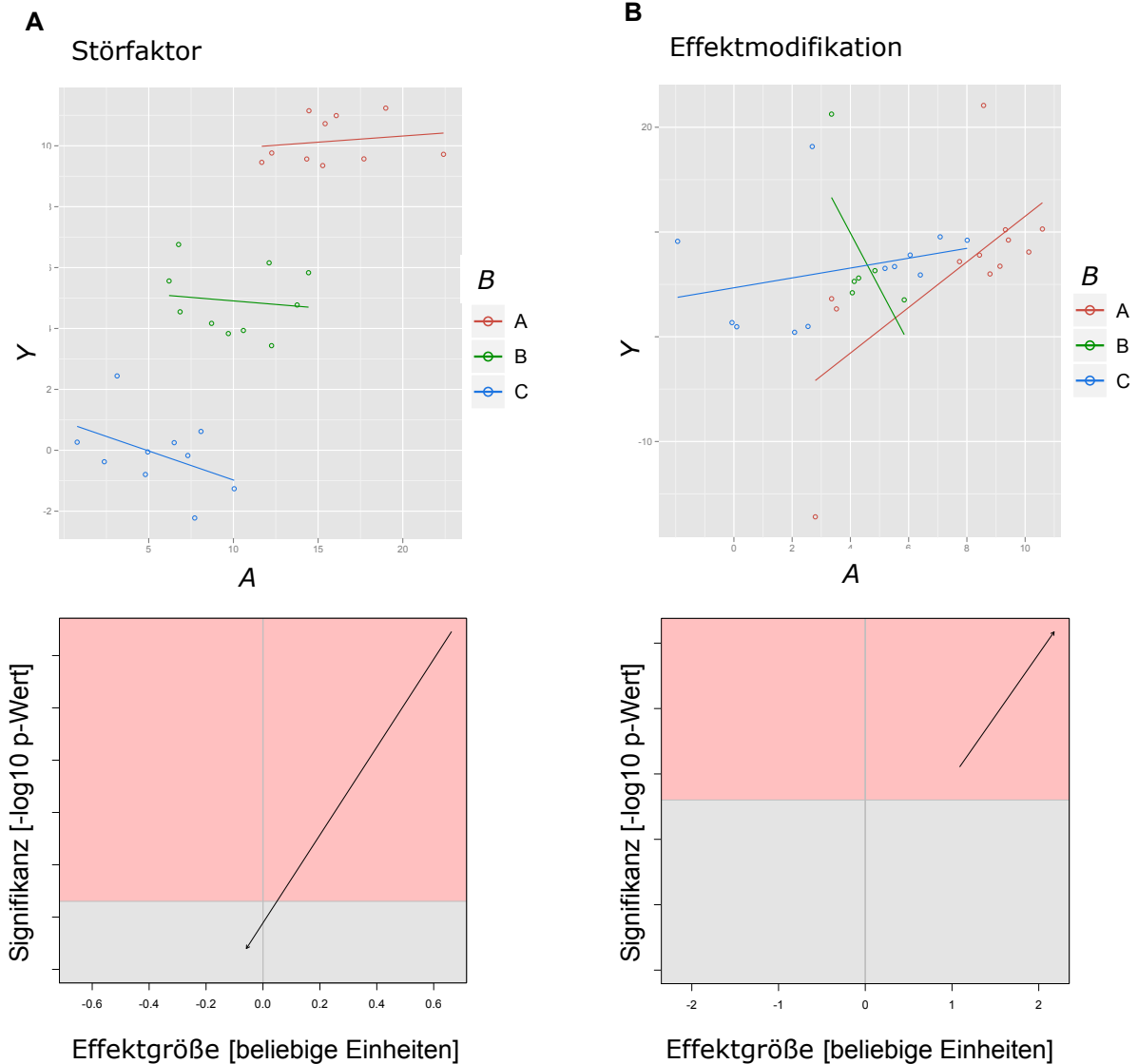


Abbildung 5: Eruptionsdiagramme angewendet auf Simulationsdatensätze zur Identifikation von Störfaktoren und Effektmodifikation.

A: Das obere Diagramm zeigt ein Punktdiagramm von willkürlich aufsteigenden Daten. Die x-Achse zeigt die unabhängige Variable A und die y-Achse die abhängige Variable Y . Die Daten werden den Gruppen A (rot), B (grün) und C (blau) zugeteilt. Der Pfeil des unteren Diagramms zeigt den Vergleich von Modell $Y \sim A+B$ zu Modell $Y \sim A$. **B:** Das obere Diagramm zeigt ein Punktdiagramm von willkürlich verteilten Daten. Die x-Achse zeigt die unabhängige Variable A und die y-Achse die abhängige Variable Y . Die Daten werden den Gruppen A (rot), B (grün) und C (blau) zugeteilt. Der Pfeil des unteren Diagramms zeigt den Vergleich der zwei Modelle $Y \sim A+B$ und $Y \sim A+B+A:B$.

Die Punkte des Datensatzes steigen an und werden durch eine dritte Variable B gruppiert. Der Effekt ist in allen Untergruppen gleich. Folglich ist keine Effektmodifikation vorhanden. Hier ist Y unabhängig von A . Jedoch sind Y und A positiv mit einem Störfaktor B korreliert. Der Pfeil

beginnt bei Modell 1, $Y \sim A$, und endet bei Modell 2, $Y \sim A+B$. Wenn nun B im Modell 2 enthalten ist, verdrängt dies alle Effekte, die fälschlicherweise A in Modell 1 zugeordnet sind. Daher zeigt das Eruptionsdiagramm von A einen kurzen Pfeil.

Effektmodifikation (auch Interaktion genannt) tritt auf, wenn die Effekte von diskreten Variablen A und B nicht additiv sind. Z.B. sind die B -gruppenspezifischen Schätzer von A signifikant unterschiedlich (VanderWeele 2009). Das Eruptionsdiagramm (Abbildung 5B) kann verwendet werden, um Interaktionen aufzudecken. Der Datensatz enthält zufällig generierte Punkte. Die Datenpunkte sind willkürlich einer Gruppenvariable B zugeordnet. Nun wird ein zweites lineares Modell 2 berechnet, das die Interaktionsvariable enthält: $Y \sim A+B+A:B$. Die Regressionslinien entsprechen der Anpassung des linearen Modells 2. Die Gruppen-spezifischen Schätzer unterscheiden sich untereinander deutlich und folglich ist Effektmodifikation präsent. Bei Effektmodifikation wird durch die Interaktionsvariable $A:B$ die erklärende Kraft erhöht. Das untere Diagramm zeigt das Eruptionsdiagramm der beiden Modelle. Der Pfeil beginnt bei dem Effekt und der Signifikanz des ersten Modells und endet bei den Ergebnissen des zweiten Modells.

In Kombination mit globalen statistischen Tests wie GlobalAncova (Hummel, Meister et al. 2008) zeigen die Eruptionsdiagramme auf, ob eine Variable zusätzliche erklärende Kraft hat oder nicht. Daher können Eruptionsdiagramme als visuelles Hilfsmittel für Modellselektion verwendet werden. Eine Variable ohne zusätzliche erklärende Kraft wird aus dem Modell ausgeschlossen, da ein schlankeres Modell nach *Occam's razor* (Breiman 2001) bevorzugt wird. Der iterative Ausschluss (Einschluss) von Variablen führt zu einer rückwärts (vorwärts) gerichteten Modellselektion. Es wird eine rückwärts gerichtete Strategie angewendet, bei der schrittweise Kovariaten eliminiert werden und das Ergebnis der Modelle nach Deletion durch das Eruptionsdiagramm verglichen wird.

2.2 Globaler Test

GlobalAncova (Hummel, Meister et al. 2008) bietet eine generelle Methodik an, um die Expressionsstruktur innerhalb einer Gengruppe hinsichtlich Designaspekten einer Studie zu untersuchen. Gen-weise lineare Modelle werden benutzt, um die Beziehung zwischen Genexpression und phänotypischen oder genomischen Kovariaten zu untersuchen. Eine ANOVA-basierte Quadratsumme der individuellen Gen-weisen linearen Modelle wird berechnet und zu einer Gruppensaussage zusammengefasst. Anhand eines Permutationstests und einer asymptotischen Verteilung der Teststatistik unter Annahme der Nullhypothese werden die p -Werte berechnet. GlobalAncova umfasst eine umfangreiche Auswahl an Modellen. Im Folgenden wird GlobalAncova angewendet, um zwei lineare Regressionsmodelle zu vergleichen (1000

Permutationen je p-Wert Berechnung). Die Ergebnisse des globalen Tests werden mit den Ergebnissen des Eruptionsdiagramms verglichen.

Der Einfluss der einzelnen Gene der Gruppe auf das Testergebnis wird durch ein Balkendiagramm veranschaulicht. Dabei wird für jedes Gen die Reduktion der Quadratsumme in einem Balken aufgetragen. Beide Gruppen des Vergleichs werden mit je einer Farbe gekennzeichnet. Die schwarze Referenzlinie gibt den mittleren Residuen-Quadratfehler wieder und entspricht der erwarteten Höhe der Balken unter der Nullhypothese. Diese besagt, dass kein Gen mit dem Phänotyp assoziiert ist.

Die differentiell exprimierten Gengruppen werden zusätzlich mittels IPA (Ingenuity® Systems, www.ingenuity.com) analysiert. Die Software ordnet die Ergebnissen der Microarray Analyse der Datenbank zu. Die Software wurde verwendet, um überrepräsentierte Signalwege in den Ergebnissen zu identifizieren. Die Berechnung des p-Wertes basiert auf einem rechts-seitigen Fisher Exakt Test.

2.3 Cluster, Signalweg und Transkriptionsfaktor-Bindestellenanalyse

Für jedes Gen wird der geschätzte Effekt der Interaktionskovariate verwendet, um entweder einen alleviating oder aggravating Effekt zu bestimmen. Ein positiver Wert wird funktionell als aggravating und ein negativer Wert als alleviating für die Interaktion interpretiert. Um einen Gencluster zu bestimmen, werden die geschätzten Effekte der Haupt-Kovariaten und der Interaktionskovariate verwendet. Die geschätzten Effekte der drei Kovariaten werden hierarchisch geclustert und das Ergebnis in einer Heatmap dargestellt. Nur die Gene, die einen signifikanten globalen Effekt zeigen (Filterung: p-Wert der F-Statistik < 0.05 nach einer „false discovery rate“ (FDR) -Korrektur nach Benjamini-Hochberg (BH), wobei wenigstens eine der Kovariaten einen Effektschätzer von $> \pm 1.5$ haben muss), werden in der Cluster Analyse dargestellt. Das Dendrogramm dieses Clusterings wird verwendet, um die p-Werte in einer zweiten Heatmap entsprechend darzustellen.

Jede der drei Kovariaten kann entweder positive oder negative Werte enthalten. Daher erhält man eine Kombination von acht verschiedenen Clustern. Diese Gencluster werden für eine Gen-Ontologie Analyse benutzt. Dabei wird die DAVID bioinformatics suite (Huang da, Sherman et al. 2009) angewandt. DAVID stellt funktionale Interpretationen von langen Genlisten zur Verfügung. Biologisch angereicherte Kategorien (z.B. GO Kategorien) werden identifiziert und funktional-verwandte Gengruppen werden aufgedeckt.

Die Gen-Cluster, die für die TFBS Analyse verwendet werden, werden zuerst nach Effektgröße der Interaktionskovariaten $> \pm 0.5$ gefiltert. Dann werden die Promotor-Sequenzen (-500 bis +100 bp relative zum Transkriptionsstart Site, die mit einer EnSEMBL ID mappen) der resultierenden Gene für die TFBS Analyse mit Hilfe des „Regulatory Sequence Analysis Tool“

(Thomas-Chollier, Defrance et al. 2011) zusammengefügt. Für jedes Gencluster werden die überrepräsentierten TFBS durch den „Transcription Factor Matrix Explorer“ (Defrance and Touzet 2006) vorhergesagt. Die putativen TFBS stammen aus der TRANSFAC Datenbank (Matys, Kel-Margoulis et al. 2006). Alle Einstellungen und Grenzwerte werden wie bei Marcinowski et al. (Marcinowski, Lidschreiber et al. 2012) verwendet.

2.4 Experimenteller Aufbau der *Yersinien* Infektionsdaten

Die publizierten Transkriptionsdaten werden aus Knochenmarks-Makrophagen der Maus generiert (van Erp, Dach et al. 2006). Der Microarray Datensatz ist verfügbar (GEO accession Nr.: GSE 9273). Dieser umfasst drei experimentelle Faktoren (Tabelle 1): *H*, der genetische Hintergrund der Maus-Makrophagen (zwei verschiedene Mausstämme C57BL/6 und BALB/c); *I*, IFN- γ -Stimulation; *I*, Infektion von zwei verschiedenen bakteriellen *Yersinien* Stämmen, WA(pTTS, p60) der nicht-virulente Kontrollstamm generiert durch Trulzsch et al. (Trulzsch, Roggenkamp et al. 2003) und der virulente Stamm WA(pYV). Tabelle 1 zeigt die Anzahl der Replikate und die Kombinationen der experimentellen Faktoren, die in den einzelnen Microarrays vorliegen.

2.5 Experimenteller Aufbau der Asthmadaten

Die klinischen Daten wurden von der Kinderklinik und Kinderpoliklinik im Dr. Haunerschen Kinderspital im Rahmen des SFB-TR22 unter Leitung der PI Bianca Schaub erhoben. Die Studienpopulation umfasste vier bis 15 Jahre alte allergische Asthmatiker (AA), nicht-allergische Asthmatiker (NAA) und gesunde Kinder (GK) und ist eine Teilmenge der CLARA Studienpopulation (Raedler *et al.* submitted). Es werden asthmatische Kinder klinisch untersucht, ein Lungenfunktionstest und eine Brust Radiographie durchgeführt, ein Blutbild erstellt und totales und spezifisches IgE durch den RAST Test gemessen. Die Einteilung nach AA/NAA findet nach klassischen Asthma Symptomen, Arzt Diagnose, medizinischer Behandlung und Lungenfunktion statt. Die Unterscheidung zwischen AA und NAA basierte auf positiver oder negativer Allergisierung. Die Gruppe GK wird durch Ausschluss von Allergie oder chronischer Krankheit und derselben Altersgruppe bestimmt. Die Blutkörperchen (PBMCs) einer Untergruppe von 14 AA, 8 NAA und 14 GK werden für Microarray Experimente ausgewählt (Tabelle 2). Von diesen werden zusätzlich die CD4 positiven (CD4⁺) T-Zellen von fünf AA und drei GK untersucht (Tabelle 3).

Die Immunantworten der stimulierten Zellen mit Medium (M) spiegeln den Grundstatus wider. Der TLR Agonist LpA als innater Stimulus wird wegen seiner Relevanz in der Protektion gegen allergische Krankheiten als zweite Stimulationsbedingung (LpA) gewählt. Die dritte

Stimulationsbedingung CD328 setzt sich aus CD3 (anfänglicher Signalaktivator) und dem Ko-Stimulator CD28 (nötig für Proliferation) zusammen (CD328).

PBMCs		Stimulationsbedingung		
		LpA	CD328	M
Phänotyp	AA	13	14	14
	NAA	8	8	8
	GK	14	13	13

Tabelle 2: Anzahl der PBMCs Microarrays

CD4+		Stimulationsbedingung M
Phänotyp	AA	3
	GK	5

Tabelle 3: Anzahl der CD4+-Microarrays

2.6 Microarray Experimente sowie Normalisierung und Filterung

Die RNA der PBMCs und CD4⁺-Zellen wird isoliert und auf Qualität und Quantität untersucht. Die totale RNA wird gefärbt und auf Affymetrix GeneChip® Human Gene 1.0 ST Arrays nach Anweisung des Herstellers hybridisiert. Die Qualität der gescannten Arrays wird durch MvA-, Dichte- und RNA-Degradierungsdiagramme untersucht. Alle Schritte werden mit R und Bioconductor (Gentleman, Carey et al. 2004) durchgeführt. Die Daten sind bei Gene Expression Omnibus verfügbar (GSE40889) (Edgar, Domrachev et al. 2002).

Hintergrundkorrektur und Normalisierung werden mit RMA (Irizarry, Hobbs et al. 2003) durchgeführt. Da das Färben und die Hybridisierung zu zwei verschiedenen Zeitpunkten stattfinden, wird wegen einem potentiellen Batch Effekt korrigiert, indem eine Batch Kovariate B in das Modell eingeführt wird. Zusätzlich wird für jedes Gen der durchschnittliche Unterschied je Stimulationsgruppe zwischen den Batch-betroffenen und nicht-betroffenen Proben berechnet. Dieser Unterschied wird von den Batch-betroffenen Proben (GSM1004208-GSM1004220) abgezogen.

Gene, die eine niedrige Variabilität über alle Proben zeigten, werden mittels eines Intensitätsfilters (Intensität > 100 in 25% der Proben) entfernt. Außerdem wird ein Varianzfilter benutzt: Der Interquartilsabstand der \log_2 Intensitäten sollte mindestens 0.5 betragen. Die gefilterten Gene werden auf differenzielle Genexpression zwischen AA, NAA und GK mit Hilfe einer moderaten Bayes'sche T-Statistik untersucht (Smyth 2004). Für multiples Testen wird durch BH korrigiert (Benjamini and Hochberg 1995). Akzeptanz oder Verwerfung

jeder Hypothese wird mit einer Serie von verwandten T-Statistiken geprüft. Die Gene werden durch einen Grenzwert (p-Wert: 0.05) als signifikant hoch-oder runter reguliert klassifiziert.

2.7 Anreicherungsanalyse

Zwei verschiedene Anreicherungsanalysen werden auf die Ergebnisse der Genexpressionsanalyse angewendet.

Zum einen werden die Ergebnisse mit einer „gene set enrichment analysis“ GSEA von Subramanian et al. (Subramanian, Tamayo et al. 2005; Subramanian, Kuehn et al. 2007) analysiert. Dieses Verfahren besteht aus zwei Schritten. Zunächst werden die Gene durch eine Gen-weise Analyse geordnet, und für jedes Gen wird eine Teststatistik ermittelt. Im zweiten Schritt werden die Gen-Ränge in einer Gruppe G untersucht. G ist angereichert, wenn hohe Ränge vertreten sind. Die Ränge Verteilung von G wird mit der Gruppe aller anderen Gene verglichen. Es handelt sich dabei um ein kompetitives Verfahren. Die Analyse wird auf der Webseite gestartet (<http://www.broadinstitute.org/gsea/msigdb/annotate.jsp>). Überschneidungen mit allen gegebenen Gengruppen (C1-C7) werden für die Analyse ausgewählt und nur die obersten zehn Gengruppen mit einem q-Wert unter 0.05 berücksichtigt.

Zum anderen wird das Anreicherungsverfahren „DAVID“ (Huang da, Sherman et al. 2009) gewählt. Signifikante Ergebnisse werden durch den p-Wert bestimmt, der durch den modifizierten Fisher Test berechnet wird. Je kleiner dieser ist, desto angereicherter ist die Kategorie.

Durch die Anwendung von zwei verschiedenen Verfahren wird die Datengrundlage der Kategorien erweitert. Da beide Verfahren auf unterschiedlichen Methoden beruhen, wird die Auswahl der Ergebnisse erweitert und die Ergebnisse können sich ergänzen. Dadurch können die Ergebnisse nicht direkt miteinander verglichen werden, sie werden separat analysiert.

3 Ergebnisse

3.1 Analyse der Infektionsdaten

Top-down Methoden der Systembiologie identifizieren molekulare Interaktionen von Systemkomponenten auf Basis von genomweiten „omics“-Experimenten (Bruggeman and Westerhoff 2007). In der vorliegenden Arbeit werden Interaktionen zunächst auf der Ebene von Experimentbedingungen einer Genexpressionsstudie mit Hilfe von linearen Regressionsmodellen untersucht (Prazeres da Costa, Hoffman et al. 2014), die häufig auf großangelegte Genexpressions-Studien angewendet werden. Als Standard fungieren Modelle erster Ordnung, welche das Expressionslevel eines Gens als gewichtete Summe der Kovariaten zu erklären versucht. Der Einfachheit halber enthalten diese linearen Modelle keine Kovariaten höherer Ordnung, also Interaktionen zwischen verschiedenen Kovariaten. Der Vorteil von Modellen höherer Ordnung wurde bisher noch nicht auf erhöhte statistische Signifikanz und Spezifität hin untersucht. Schrittweise werden Kovariaten aus den einzelnen Modellen entfernt. Anschließend werden die Modelle mit Hilfe des Eruptionsdiagramms verglichen.

3.1.1 Kovariaten des Infektionsdatensatzes

Die Untersuchung wird am Beispiel eines Transkriptionsdatensatzes ausgeführt. Dieser basiert auf Knochenmark-Mausmakrophagen (van Erp, Dach et al. 2006) und enthält vier verschiedene experimentelle Faktoren, die weiterhin Kovariaten genannt werden. *H*: Die Makrophagen stammen entweder von BALB/c- oder C57BL/6-Mäusen; *I*: Die Makrophagen wurden mit IFN- γ stimuliert oder nicht; *L*: Die Makrophagen wurden mit dem virulenten *Y. enterocolitica* Stamm WA(pYV) oder dem *Y. enterocolitica* Kontrollstamm WA(pTTS, p60) infiziert oder nicht (Tabelle 1). Des Weiteren sind Kombinationen dieser Kovariaten in dem Datensatz vorhanden, wie die IFN- γ -Stimulation mit einem der beiden *Y. enterocolitica* Stämmen jeweils in beiden Mausstämmen. Daher eignet sich dieser Datensatz zur Untersuchung von Interaktionen. In dieser Arbeit liegt der Fokus auf der transkriptionellen Antwort der beiden Mausstämmen und dem Effekt der IFN- γ -Stimulation, um die Genexpressionsunterschiede zwischen den Stämmen durch IFN- γ -Stimulation zu untersuchen, also die Interaktion der Kovariaten *H* und *I*.

3.1.2 Transkriptionelle Effekte durch Interaktionsmodelle

Die C57BL/6-Mäuse sind fähig, die *Yersinien* Infektion zu kontrollieren und zu eliminieren. Im Gegensatz dazu können die BALB/c-Mäuse dies nicht kontrollieren, was auf ein niedriges Niveau von IFN- γ zurückzuführen ist. Die Resistenz der C57BL/6-Mäuse gegen die *Yersinien* Infektion

korreliert hingegen mit einem starken Anstieg der IFN- γ Produktion am Anfang der Infektion (Hancock, Schaedler et al. 1986; Autenrieth, Beer et al. 1994). Daher ergibt sich hinsichtlich des Überlebens der Mäuse eine Interaktion zwischen den Mausstämmen und IFN- γ . Die transkriptionelle Antwort, die dieser Interaktion unterliegt, und der Unterschied der IFN- γ Produktion sind unklar. In dieser Arbeit werden die zu erwartenden Interaktionen auf molekularer (transkriptioneller) Ebene mit einem linearen Modell untersucht, das die Genexpression Y durch die Hauptkovariate und alle möglichen Interaktionskovariaten (Modell 1, Tabelle 4) erklärt. Die Interaktionskovariate $H:\Gamma$ wirkt signifikant bei der Erklärung der Daten mit. Dies wird aus dem Vulkandiagramm (Jin, Riley et al. 2001) (Abbildung 6) ersichtlich.

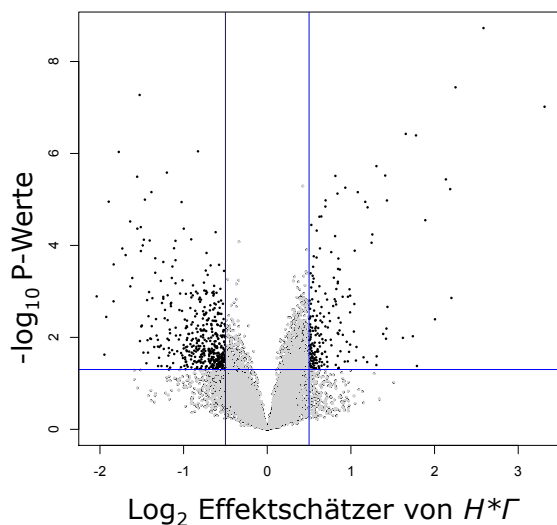


Abbildung 6: Vulkandiagramm der Interaktionskovariate von Modell 1, angewendet auf den Infektionsdatensatz.

Daher muss die Interaktionskovariate $H:\Gamma$ bei der Analyse durch lineare Regression berücksichtigt werden, um differentiell exprimierte Gene der beiden Mausstämmen durch IFN- γ -Stimulation zu analysieren.

3.1.3 Top-down Analyse von geschachtelten Modellen durch Eruptionsdiagramme

Das Modell 1 wird im Folgenden sukzessive reduziert (Tabelle 4), zunächst durch Eliminierung der Kovariate dritter Ordnung $H:\Gamma:I$ (Modell 2), dann durch Interaktionen von I mit H und Γ (Modell 2) und schließlich durch Eliminierung von I (Modell 4).

Modellname	Lineares Regressionsmodell
Modell 1	$Y \sim H + \Gamma + H:\Gamma + I + H:I + \Gamma:I + \mathbf{H:I:\Gamma}$

Modell 2	$Y \sim H + \Gamma + H:\Gamma + I + H:I + \Gamma:I$
Modell 3	$Y \sim H + \Gamma + H:\Gamma + I$
Modell 4	$Y \sim H + \Gamma + H:\Gamma$

Tabelle 4: Die linearen Regressionsmodelle des Infektionsdatensatzes.

Die hierarchische Ordnung der Modelle ermöglicht die Anwendung einer rückwärts-gewandten Modellselektionsstrategie. Dabei werden, beginnend bei Modell 1, sukzessive Kovariaten gestrichen, bis eine Verbesserung der Selektionskriterien erreicht wird. Das Hauptziel ist die Untersuchung der Effektschätzer von $H:\Gamma$. Dazu wird das Eruptionsdiagramm verwendet, um die Interaktionskovariate $H:\Gamma$ zwischen zwei Modellen zu vergleichen.

Im Folgenden wird getestet, ob die Kovariate dritter Ordnung $H:\Gamma:I$ den Effekt der Interaktionskovariate $H:\Gamma$ stört. Dazu werden die Ergebnisse von Modell 1 und Modell 2 in einem Eruptionsdiagramm verglichen (Abbildung 7A). Die Pfeile starten bei Modell 1 und enden bei Modell 2. Die Effektgröße wird auf der x-Achse aufgetragen und die Signifikanz anhand von \log_{10} p-Werten auf der y-Achse. Schwarze und graue Pfeile repräsentieren signifikante und nicht-signifikante Gene. Rote und blaue Pfeile stellen Gene dar, die innerhalb des ROI liegen und nur in Modell 1 oder Modell 2 signifikant sind. Die Richtung der Pfeile weist auf erhöhte Signifikanz und Effektgröße von Modell 2 hin. Die Kovariate $H:\Gamma:I$ ist folglich eine irrelevante Kovariate. Außerdem werden die p-Werte der Interaktionskovariate $H:\Gamma$ beider Modelle in einem Dichtediagramm gemessen (Abbildung 7B). Dieses verdeutlicht die höhere Signifikanz von Modell 2. Die Resultate stimmen mit den Ergebnissen des globalen Tests (F-Wert = 1,07) überein. Die Kovariate $H:\Gamma:I$ liefert keine zusätzliche erklärende Kraft.

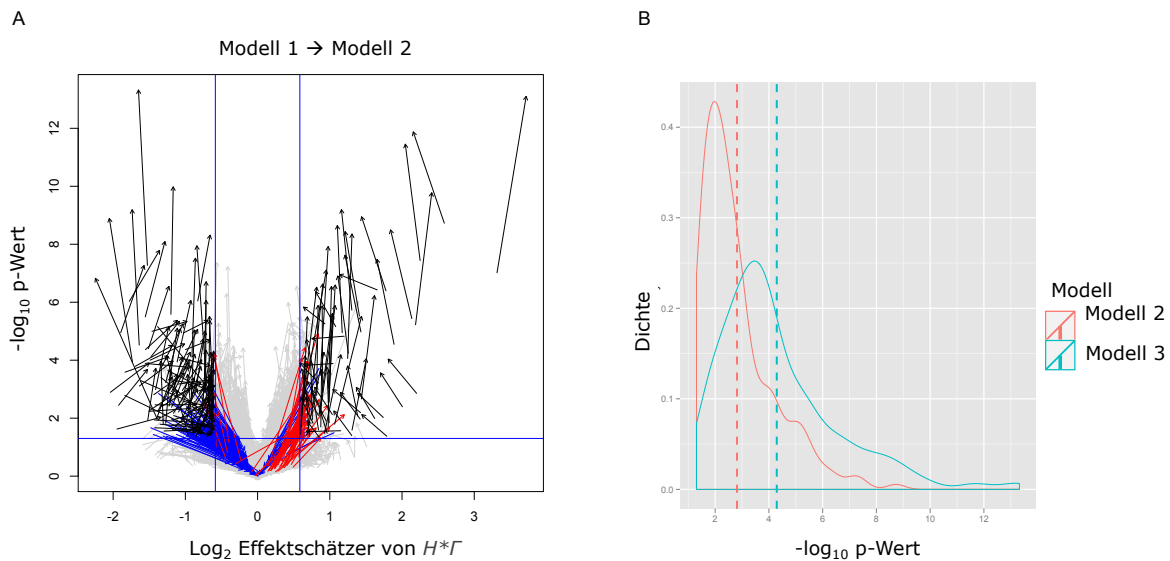


Abbildung 7: A: Eruptionsdiagramm von Modell 1 und Modell 2. B: Dichtediagramm der -log₁₀ p-Werte von Modell 1 und Modell 2.

3.1.4 Bestimmung von beschreibender und erklärender Kraft der Kovariaten

Im nächsten Schritt wird die erklärende Kraft der Kovariaten zweiter Ordnung $H:I$ und $\Gamma:I$ zusätzlich zur Interaktionskovariate $H:\Gamma$ untersucht. Es wird also das Modell 3 (Tabelle 2) analysiert, das die vier Haupt-Kovariaten und die Interaktionskovariate $H:\Gamma$ enthält. Wenn die Kovariaten zweiter Ordnung gestrichen werden, kann man testen, ob man erklärende Kraft gewinnt oder verliert. Dazu wird das Modell 2 mit dem Modell 3 in einem Eruptionsdiagramm verglichen (Abbildung 8). Das Diagramm zeigt keine erheblichen Unterschiede in Signifikanz und Effektgröße zwischen den beiden Modellen. Das Dichtediagramm unterstreicht dieses Ergebnis. Dementsprechend liefert der globale Test keine hohe Signifikanz der Kovariaten zweiter Ordnung (F-Wert 1,39). Folglich verbessern die Kovariaten zweiter Ordnung nicht die Anpassungsgüte. Aufgrund von generellen Modellselektionskriterien nach *Occam's razor* wird das schlankere Modell 3 für die weiterführende Analyse verwendet.

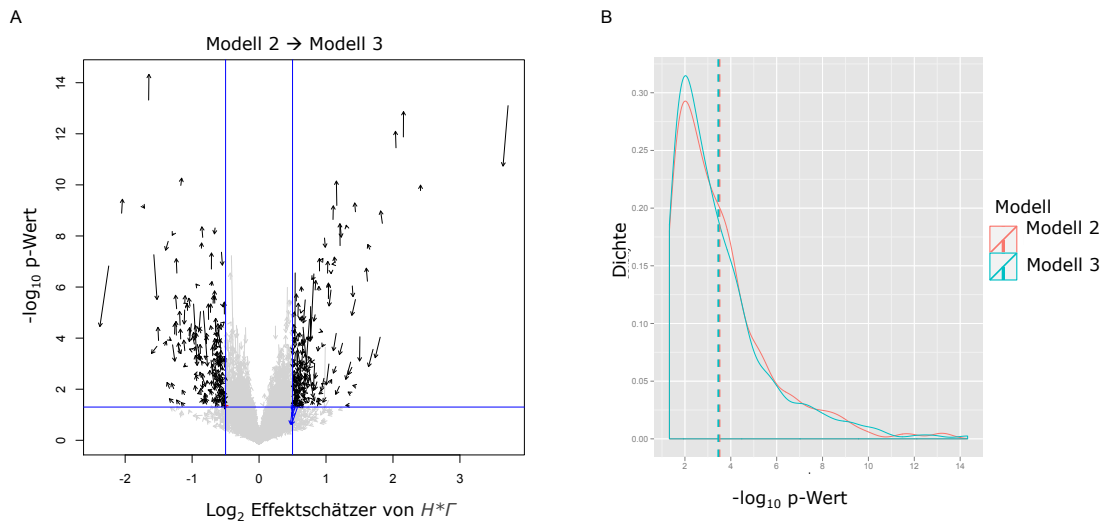


Abbildung 8: A: Eruptionsdiagramm von Modell 2 und Modell 3. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 2 und Modell 3.

Der Datensatz enthält außerdem Proben, die mit verschiedenen *Yersinien* Stämmen infiziert worden sind (Tabelle 1). Zwar liegt der Fokus der Arbeit auf der Untersuchung differenzieller Ko-Expression von H und Γ ; dennoch werden alle Microarray Daten einschließlich der *Yersinien* Infektion für die Analyse verwendet. Im Folgenden wird getestet ob die Kovariate I zusätzliche erklärende Kraft auf die Interaktionskovariate $H:\Gamma$ hat. Daher wird ein Schritt weiter nach unten in der Modellhierarchie gegangen (Tabelle 4) und die Kovariate I gestrichen (Modell 4). Dieses Modell 4 enthält also die zwei Hauptkovariaten H und Γ und deren Interaktionskovariate. In einem Eruptionsdiagramm sind die Unterschiede zwischen Modell 3 und Modell 4 bezüglich der Interaktionskovariate $H:\Gamma$ aufgezeigt (Abbildung 9).

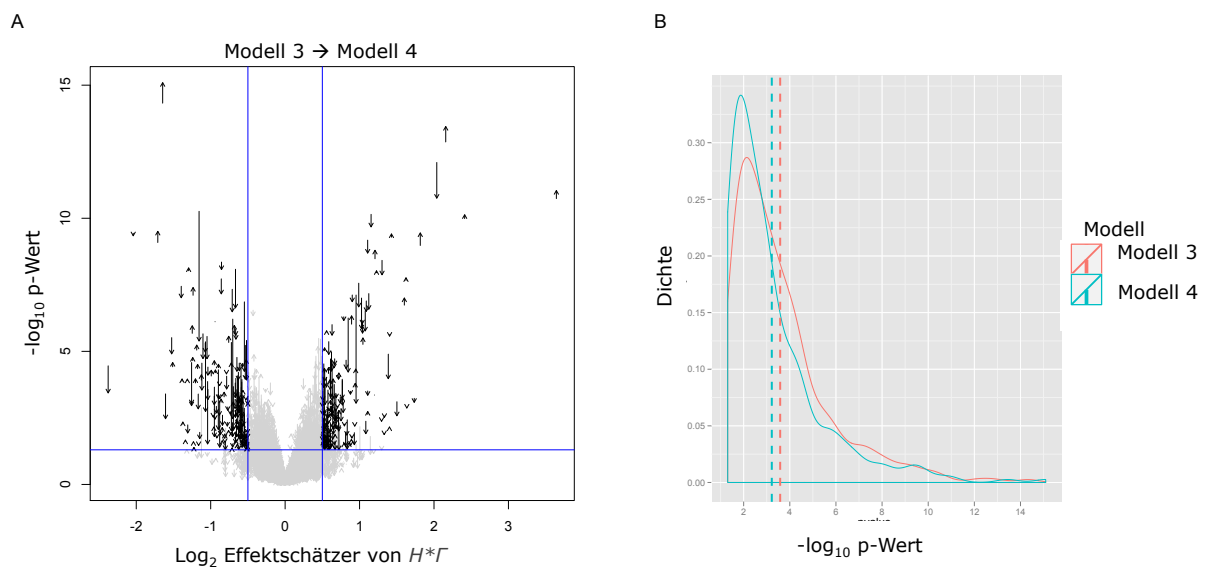


Abbildung 9: A: Eruptionsdiagramm von Modell 3 und Modell 4. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 3 und Modell 4.

Die Pfeile des Eruptionsdiagramms beginnen bei dem komplexeren Modell 3 und enden bei dem Modell 4. Die Richtung der Pfeile weist eine Veränderung der p-Werte auf, während sich die Effektschätzer zwischen den beiden Modellen nicht ändern. Das Dichtediagramm verdeutlicht den Unterschied der beiden Modelle bezüglich der erhöhten statistischen Kraft von Modell 3. Der globale Test zeigt dasselbe Ergebnis: Die Effektschätzer der Kovariate I sind signifikant (F-Wert = 11,28). Folglich gewinnt man erklärende Kraft durch die Haupt-Kovariate I , so dass das Modell 3 für die weiterführende Analyse ausgewählt wird.

Alle Vergleiche durch das Eruptionsdiagramm zeigen, dass die Daten Effekte zweiter Ordnung beinhalten. Modell 3 zeigt die höchste Spezifität und eignet sich daher für eine detaillierte Analyse der Daten. Es beinhaltet die Kovariate I und schließt die irrelevante Kovariate $H:\Gamma:I$ aus. Daher ist das Modell 3 für die weitere Analyse und biologische Interpretation der Ergebnisse geeignet.

3.1.5 Aufdeckung funktionaler Gencluster durch Interaktionsanalyse

Nachdem das passende lineare Regressionsmodell für den Datensatz bestimmt worden ist, werden die Gene analysiert, die einen Interaktionseffekt zwischen H und Γ zeigen. Dazu werden die Ergebnisse der Haupteffekte von H und Γ und deren Interaktionseffekte hierarchisch geclustert (siehe Methoden) und mit Hilfe einer Heatmap dargestellt (Abbildung 10). Die erste Spalte zeigt die Effektschätzer der Kovariate H , die dem Genexpressionsunterschied zwischen den Mausstämmen BALB/c und C57BL/6 ohne IFN- γ -Stimulation entsprechen (rot: hochreguliert in BALB/c; grün: runter reguliert in BALB/c). Die zweite Spalte Γ stellt die Unterschiede in der Genexpression durch IFN- γ -Stimulation in BALB/c Mäusen, die dritte Spalte die Effektschätzer der Interaktionskovariate dar. Positive Werte sind pink gefärbt und werden als aggravating interpretiert, während negative Werte in türkis als alleviating interpretiert werden. Die dargestellten Gene sind gefiltert nach einem p-Wert der F-Statistik ≤ 0.05 (FDR-Korrektur nach BH) und dem Effektschätzer von ± 1.5 . Die p-Werte der Gene sind entsprechend des Baumdiagramms der Effektschätzer in einer zweiten Heatmap dargestellt. Die Werte gehen von weiß nach blau. Aus den beiden Heatmaps wird ersichtlich, dass hohe Effektschätzer auch signifikante p-Werte aufweisen.

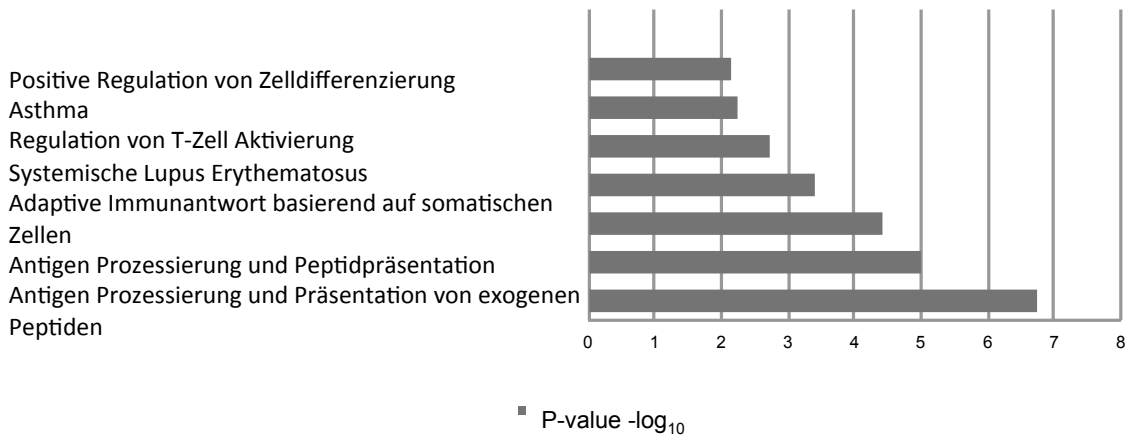


Abbildung 10: Clusteranalyse der Effekte und P-Werte von Modell 3.

Jede der drei Kovariaten dieser Analyse hat entweder positive oder negative Werte. Daher können Gene in acht verschiedene Cluster fallen. Nun werden die funktionellen Eigenschaften der acht Cluster durch eine Anreicherungsanalyse hinsichtlich Gen-Ontologie (GO) Terme (biologischer Prozess) und KEGG Signalwege (Kanehisa 2002) untersucht (Anhang 6.1). Die signifikantesten Kategorien ($p \leq 0.01$) werden in einem Balkendiagramm dargestellt. Diese sind von unten (am signifikantesten) nach oben (nicht signifikant) sortiert. Ähnliche Kategorien werden von der signifikantesten und spezifischsten Kategorie repräsentiert. Die Mehrzahl der Cluster zeigt „Immunantwort“ als die am meisten angereicherte Kategorie.

Interessanterweise sind in Cluster 3 „Zellmigration“ und „Zelltod“ als angereicherte Kategorien ausgewiesen. Wundantwort und Abwehrantwort werden mit dem Cluster 7 assoziiert. Cluster 6 weist „Antigen Prozessieren und Präsentation durch MHC Klasse II“ auf (Abbildung 11). Cluster 6, welcher Antigen-präsentierende Gene der Klasse II Aa, Ab und Eb enthält, wird auf TFBS durch eine *in silico* Promotor Analyse untersucht (Abbildung 11). Die Analyse durch TRANSFAC ergibt eine gemeinsame Bindestelle für NFY und die CAAT Box (Abbildung 11). Andere Gene des Clusters 6 sind in die IFN- γ induzierte Immunantwort involviert, wie TRIM30d, IIGP1 und CXCL9. Der apoptotische Regulator Cflar, auch als Flip bekannt, ist ebenfalls in Cluster 6 vorhanden.

Cluster 6
Positive Haupteffekte H und I , negative Interaktion



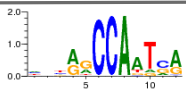
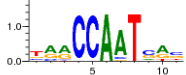
Motif	Logo	Promotoren mit Bindestellen	p-Wert
CAAT box		11/22	4.64e-09
NF-Y		11/22	2.16e-08

Abbildung 11: GO- und TFBS Analyse von Cluster 6 der Clusteranalyse.

Die TFBS NFY und CAAT Box liegen nahe dem Transkriptionsstart dieser Gene und unterstreichen damit die funktionelle Relevanz dieser Analyse. Dies wird weiterhin unterstützt durch die Tatsache, dass die NFY Bindestelle Teil des MHC-Klasse-II-Moleküls ist. Möglicherweise sind diese Gene C2TA Ziel-Gene und werden daher durch MHC-Klasse-II ko-reguliert.

3.1.6 Aufdeckung biologisch relevanter Merkmale durch Interaktionsanalyse

Um die biologische Relevanz der Interaktionsanalyse zu zeigen, wird das Gen H2-Ea diskutiert, welches zu einem MHC-Klasse-II-Gen wird, das von IFN- γ durch ein C2TA Transaktivatorprotein reguliert wird. H2-Ea ist in BALB/c ein aktives Gen, dessen Produkt die Oberfläche des Peptid-präsentierenden H2-Heterodimer formt. In C57BL/6-Mäusen ist es ein Pseudogen H2-Ea-ps auf Grund einer großen genomischen Deletion, die den Core-Promotor und den Transkriptionsstart umfasst. Daher gibt es in C57BL/6 fast kein Transkript von H2-Ea-ps. Dieses Gen fällt in das Cluster 8. Die detaillierte Analyse des Expressionsmusters dieses Gens wird in einem Punktdiagramm (Abbildung 12) veranschaulicht und deckt auf: a) Das Transkriptionslevel ist in C57BL/6 stark reduziert im Vergleich zu BALB/c-Mäusen; b) IFN- γ -Stimulation zeigt keinen Einfluss auf die Expression des Pseudogens in C57BL/6-Makrophagen.

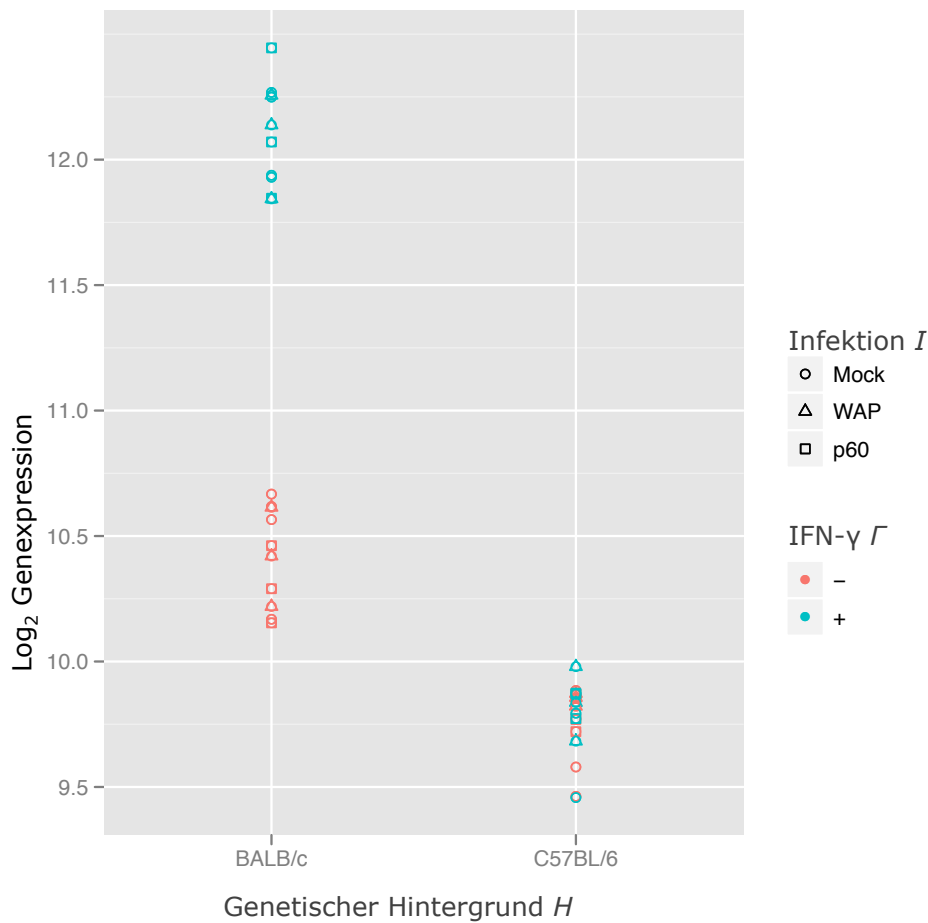


Abbildung 12: Genexpression von H2-Ea-ps des genetischen Hintergrund *H*, BALB/c und C57BL/6.

Jeder Punkt entspricht einem Wert eines Microarray. Die Kovariate *I* ist mit verschiedenen Symbolen und die Kovariate *Γ* ist mit zwei verschiedenen Farben gekennzeichnet.

Die Genexpression von H2-Ea-ps in BALB/c ist hochreguliert durch IFN-γ-Stimulation. Weiterhin hat es ein niedriges Expressionslevel in C57BL/6-Mäusen, mit und ohne IFN-γ Stimulation. Die Einfügung der Kovariate *I* bringt in C57BL/6 Mäusen keine zusätzliche erklärende Kraft. H2-Ea ist ein spezieller Fall durch seine Natur als Pseudogen in einem der analysierten Mausstämmen. Die funktionellen klassischen MHC Klasse II Gene sind Teil von Cluster 6. Die MHC Klasse II Gene sind in C57BL/6 und durch IFN-γ hochreguliert.

3.2 Analyse der Asthmadaten

Die Asthmadaten werden durch eine bottom-up Strategie analysiert. Bei dieser Strategie werden jene Mechanismen untersucht, durch welche funktionelle Eigenschaften in den Interaktionen von bekannten Komponenten (Bruggeman and Westerhoff 2007) entstehen. Die klinischen Daten wurden an der Kinderklinik und Kinderpoliklinik im Dr. Haunerschen Kinderspital im Rahmen des SFB-TR22 erhoben und in Kooperation analysiert (Raedler, Prazeres da Costa, Ballenberger et al., submitted).

3.2.1 Batch Effekte

Dabei führte die Autorin dieser Arbeit die Qualitätskontrolle der RMA (robust multi-array average) -normalisierten Transkriptionsdaten mittels PCA (principal component analysis) Analyse (Abbildung 13) und Boxplot (Abbildung 14) durchgeführt.

Die erste Komponente dieser PCA Analyse stellt den Phänotyp (x-Achse), die zweite Komponente die Stimulationsbedingung (y-Achse) und die dritte Komponente das Scan Datum der Microarrays (z-Achse) dar. Zusätzlich wird jedes Scan-Datum farblich gekennzeichnet. Dadurch wird ersichtlich, dass ein Batch Effekt durch die Durchführung der Microarrays an zwei verschiedenen Zeitpunkten (Jahr 2010 und 2011) entsteht.

Der Effekt ist in allen Gruppen, Phänotypen sowie Stimulationsbedingungen präsent. Daher kann man den Batch Effekt korrigieren, indem eine zusätzliche Kovariate in das lineare Modell aufgenommen wird.

PCA Mapping (27.3%)

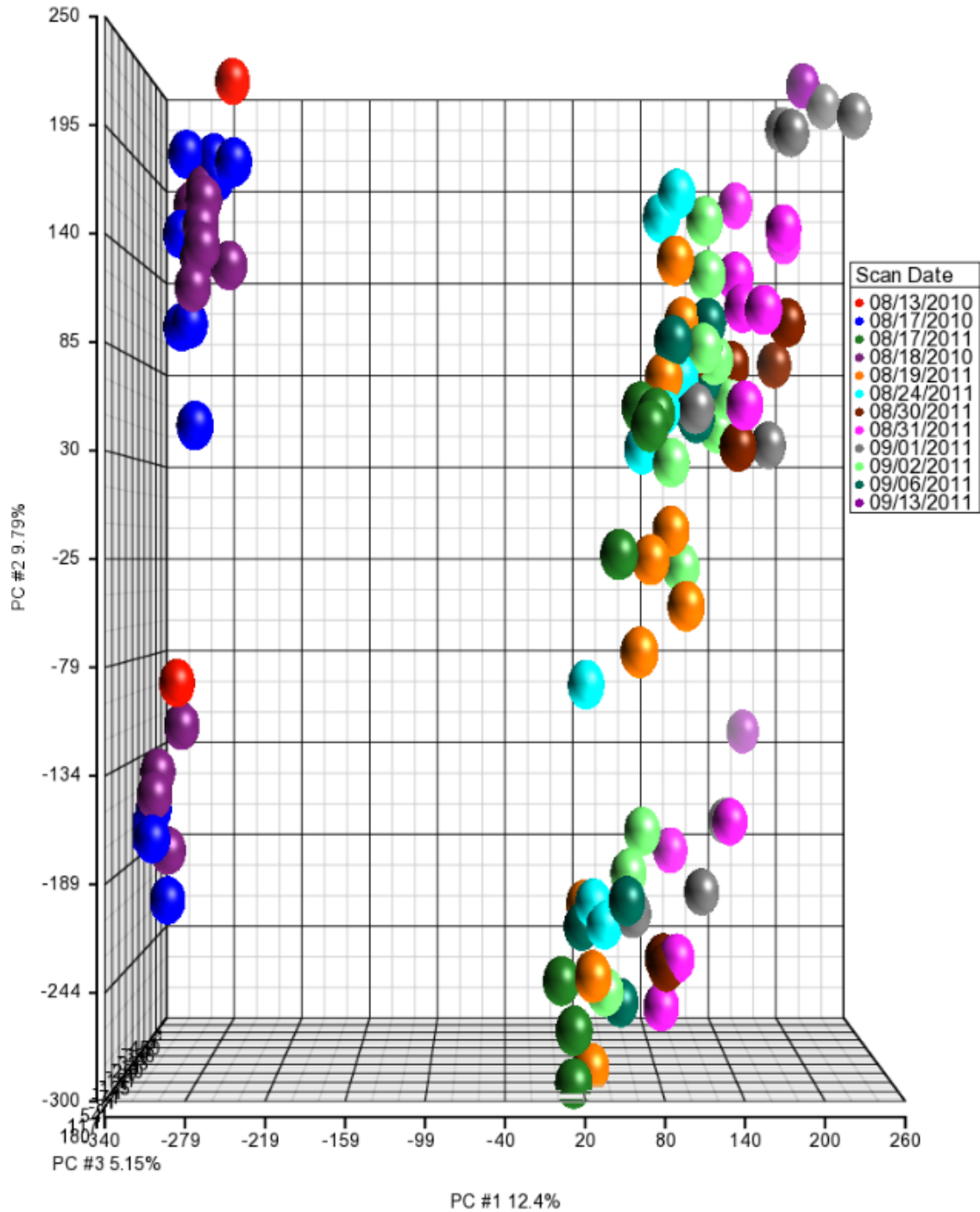


Abbildung 13: PCA Analyse der Asthmadaten. Jede Farbe kennzeichnet ein Scan Datum der Microarrays.

Der Boxplot (Abbildung 14) deckt einen weiteren Batch Effekt auf, der durch das Färben zu unterschiedlichen Zeitpunkten entstanden ist. Bis zu zwölf Microarrays können an einem Zeitpunkt gefärbt werden. Die betroffenen Microarrays sind mit einem Stern gekennzeichnet. In diesem Fall sind nicht alle Gruppen in dem Batch vertreten. Demzufolge kann nicht durch eine

Batch Kovariate korrigiert werden. Daher wird für jedes Gen der durchschnittliche Unterschied zwischen den Batch-betroffenen und nicht-betroffenen Proben berechnet und der Unterschied von den Batch-betroffenen Proben abgezogen.

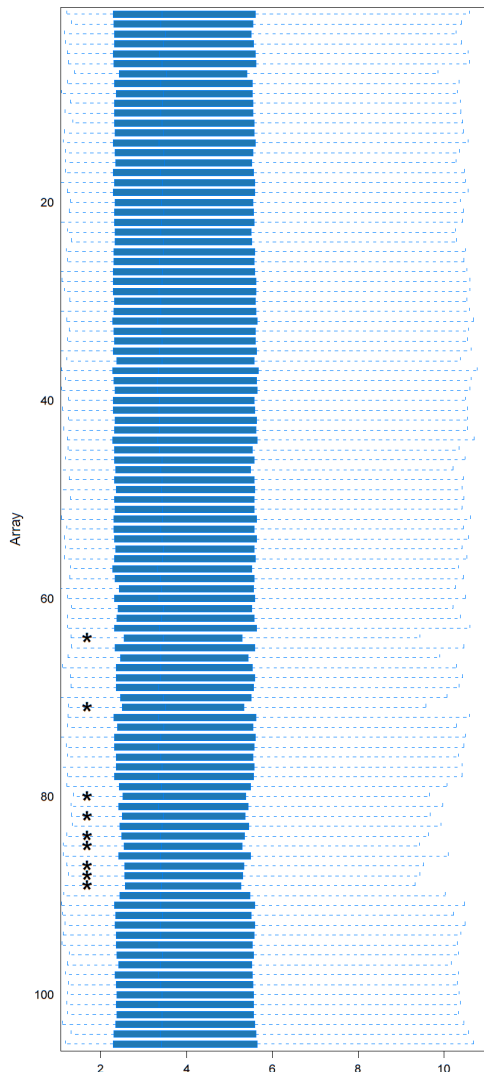


Abbildung 14: Boxplots der Asthmadaten.

Microarrays, die bei der Qualitätskontrolle auffallen, sind mit einem Stern gekennzeichnet.

3.2.2 Auswahl des Modells basierend auf dem Eruptionsdiagramm

Mittels linearer Regression untersuchte die Autorin des vorliegenden Manuskripts den Datensatz auf bekannte Interaktionen von Genen und auf differenzielle Regulation von einzelnen Genen untersucht. Dazu wird ein passendes Regressionsmodell erstellt. Das Eruptionsdiagramm wird verwendet, um das korrekte Modell für die Regressionsanalyse des Asthma-Transkriptionsdatensatzes zu identifizieren. Das lineare Modell erster Ordnung umfasst die Kovariaten Phänotyp P (AA, NAA und GK), Stimulationsbedingung S (LpA, CD328 und M) und die

Batch Kovariate B (Tabelle 5, Modell 1). Da die Microarrays Kombinationen der Kovariaten P und S enthalten (Tabelle 2), kommt ein Modell zweiter Ordnung, das die Interaktionskovariate $P:S$ enthält, für die Regressionsanalyse in Frage (Tabelle 5, Modell 2). Für die Berechnung der linearen Modelle werden die Kovariaten P und S in die Stellvertretervariablen AA (1: AA und 0: GK) und NAA (1: NAA und 0: GK) und LpA (1: LpA und 0: M) und $CD328$ (1: $CD328$ und 0: M) aufgeteilt.

Modellname	Lineares Regressionsmodell	Regressionsmodell mit Stellvertretervariable
Modell 1	$Y \sim P + S + B$	$Y \sim AA + NAA + LpA + CD328 + B$
Modell 2	$Y \sim P + S + P:S + B$	$Y \sim AA + NAA + LpA + CD328 + AA:LpA + AA:CD328 + NAA:LpA + NAA:CD328 + B$

Tabelle 5: Lineare Regressionsmodelle der Asthmadaten.

Mittels eines Eruptionsdiagramms (Prazeres da Costa, Hoffman et al. 2014) (Abbildung 15A) bestimmte die Autorin dieser Dissertation das passende Modell für den Datensatz.

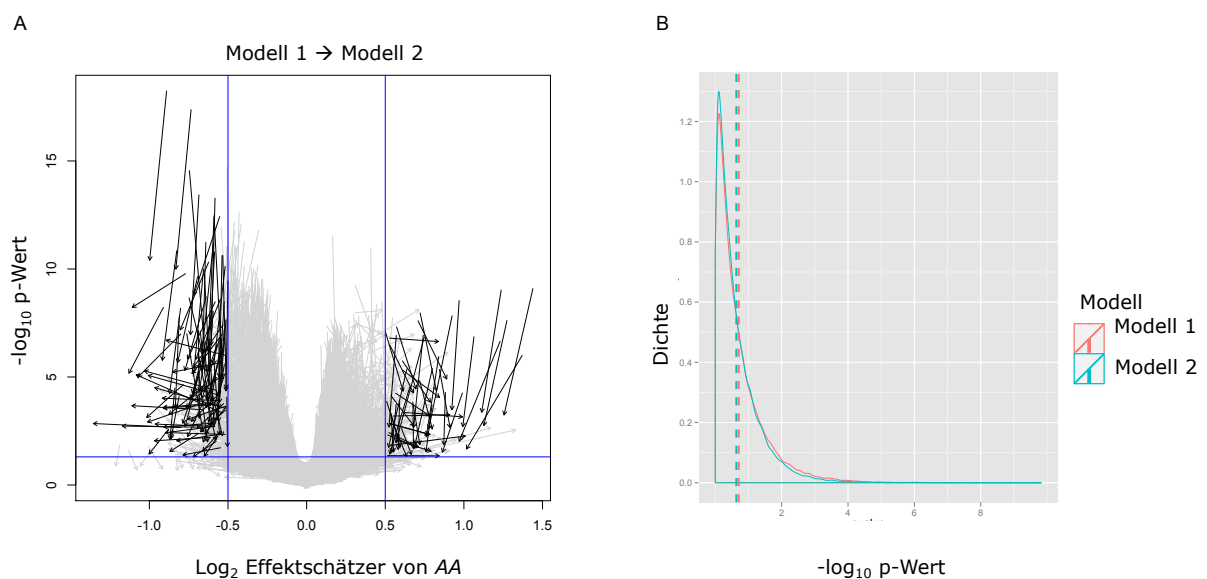


Abbildung 15: A: Eruptionsdiagramm von AA. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 1 und Modell 2.

Das Eruptionsdiagramm (Abbildung 15) zeigt den Vergleich zwischen Modell 1 und Modell 2 mittels Stellvertretervariable AA . Die p-Werte von Modell 1, das nicht $P:S$ enthält, sind etwas signifikanter und die Effektschätzer größer. Dies bestätigen die Ergebnisse des Dichtediagramms in Abbildung 15B und das Ergebnis des globalen Tests GlobalAncova (p-Wert: 0,55). Bei den Genen, deren p-Wert und Effektschätzer sich ändern, zeigt sich die

Interaktionskovariate $P:S$ als Störfaktor (Abbildung 3). Die Gene, die sich nur im p-Wert ändern, zeigen keine zusätzliche erklärende Kraft von $P:S$ auf.

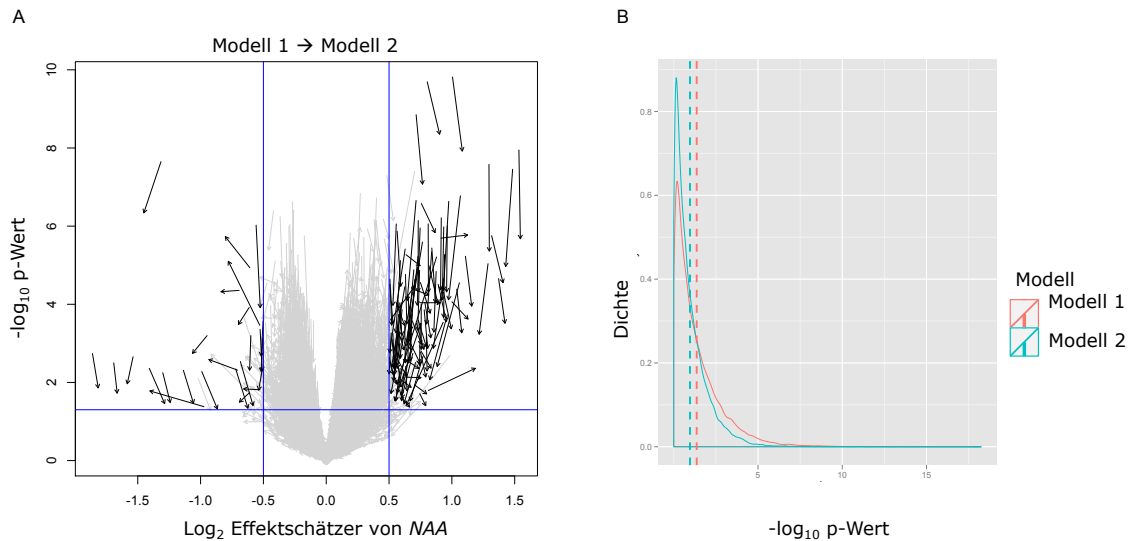


Abbildung 16: A: Eruptionsdiagramm von NAA. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 1 und Modell 2.

Die Eruptionsdiagramme der Stellvertretervariablen NAA und Kovariate B zeigen dieselben Ergebnisse, Modell 1 zeigt höhere p-Werte im Vergleich zu Modell 2 (Abbildung 16 A und B und Anhang 6.2). Daher wird für die weitere Analyse Modell 1 verwendet.

3.2.3 Top-down Analyse: Differenziell exprimierte Gene zwischen AA, NAA und GK

Die Autorin dieser Arbeit filterte zunächst die normalisierten Genexpressionsdaten basierend auf Intensität und Varianz. Nach der Filterung werden 3536 Gene hinsichtlich differentieller Genexpression untersucht. Differenziell regulierte Gene zwischen den Phänotypen AA, NAA und GK werden durch Modellanpassung von Modell 1 bestimmt. Dabei erklärt die Genexpression Y die Kovariaten Phänotyp P , Stimulationsbedingung S und die Batch Kovariate B in einem additiven Modell erster Ordnung: $Y \sim P + S + B$. Die moderate Bayes'sche T-Statistik wird benutzt, um die statistische Signifikanz und differentielle Expression zu berechnen. Nach Filterung von p-Werten und Effektgröße werden 1523 Gene als signifikant unterschiedlich zwischen den möglichen Vergleichen der Phänotypen (Abbildung 17A) identifiziert. Diese Gene wurden in Kooperation (Rädler 2014) hinsichtlich differentieller Regulation untersucht. Die drei Vulkandiagramme visualisieren die Ergebnisse der differentiellen Regulation: AA gegen GK, NAA gegen GK und NAA gegen AA. Gene, die über $\pm 0,5$ Expressionsunterschied zeigen und deren p-Wert $\geq -\log_{10}(0,05)$ ist, werden als signifikant unterschiedlich exprimiert bezeichnet. Das Venndiagramm (Abbildung 17B) stellt das Ergebnis des Vergleichs der drei Gruppen dar. Jeder Gruppenvergleich wird durch einen Kreis symbolisiert. Die Anzahl der differentiell exprimierten Gene ist innerhalb der Kreise angegeben, während die Anzahl der differentiell regulierten Gene,

die in zwei oder drei Vergleichen vorkommen, in den sich überschneidenden Kreisen dargestellt wird. Abbildung 17C repräsentiert das Clusterergebnis der 1523 Gene als Heatmap. Hochregulierte Gene sind in gelb und runter-regulierte Gene in blau dargestellt.

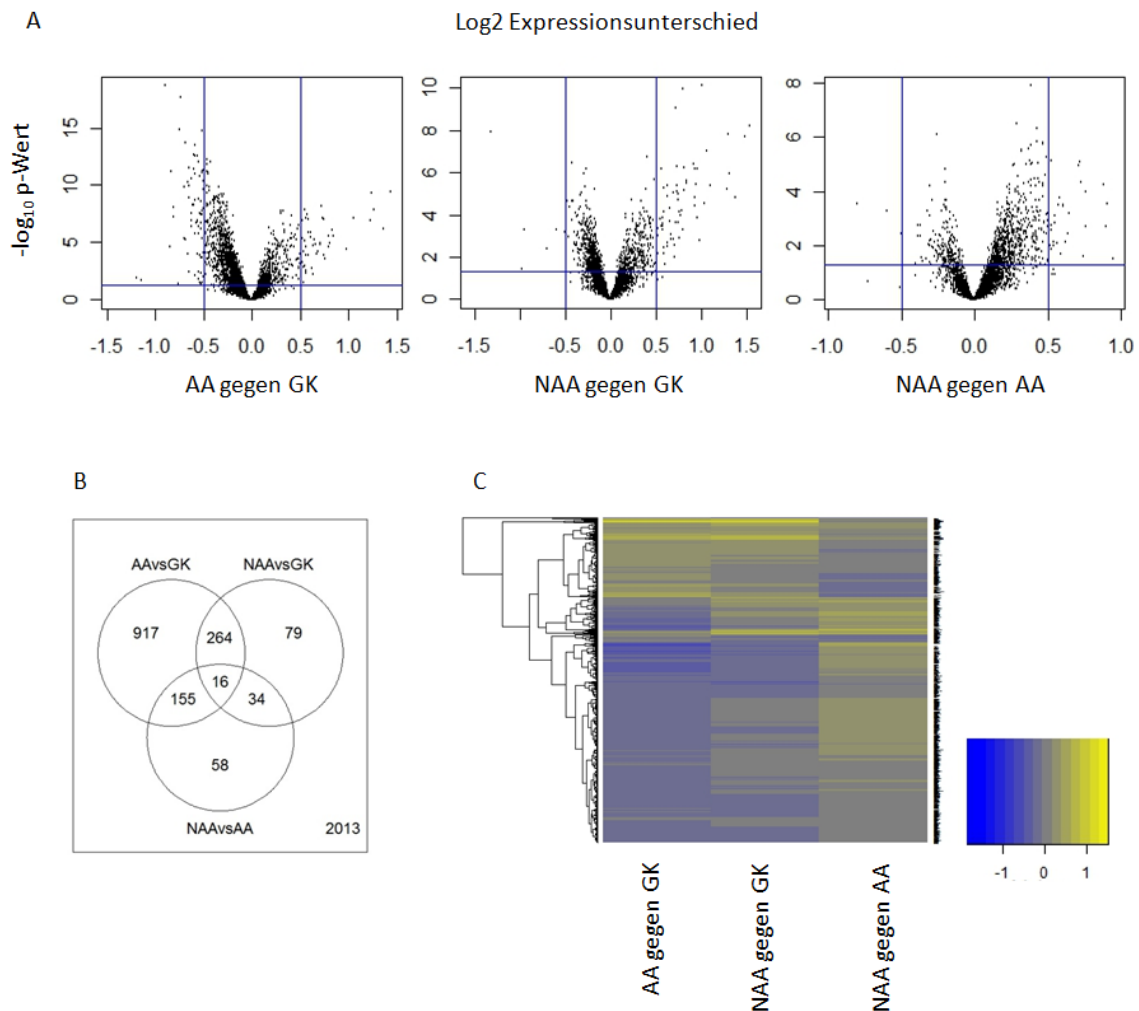


Abbildung 17: Differenziell exprimierte Gene der Asthma-Microarray-Analyse.

A: Vulkandiagramme der drei Vergleiche der Phänotypen. B: Venndiagramm. Die Anzahl der signifikant unterschiedlich regulierten Gene ist in den Kreisen angegeben. Die Gene, die in Vergleichen gemeinsam sind, sind in den überschneidenden Kreisen angegeben. C: Heatmap der Expressionsunterschiede der signifikant unterschiedlich regulierten Gene.

3.2.4 Bottom-up Analyse: Differenziell regulierte Signalwege zwischen AA, NAA und GK

Zusätzlich zur top-down Analyse auf der Ebene einzelner Gene wird eine Analyse funktioneller Gengruppen durchgeführt. Diese enthalten das Vorwissen der Struktur der Daten auf der Ebene von Geninteraktionen. Mittels eines globalen Tests GlobalAncova wurden von der Autorin dieses Manuskripts die Zusammenhänge zwischen der Expression einer bestimmten Gruppe von Genen

mit den Kovariaten *P*, *S* und *B* quantifiziert. Die Nullhypothese lautet: ‚Kein Gen der Gengruppe ist differentiell exprimiert‘.

Die Gengruppen basieren einerseits auf KEGG-Signalwegen (Kanehisa 2002). Andererseits wurden in Kooperation (Rädler 2014) Gengruppen von Zelltypen erstellt: Verschiedene T-Zell Gruppen (TC), Makrophagen (Mac), natürliche Killerzellen (NK) und dendritischen (DC) und B-Zellen (BC) (jeweils TFen, Zytokine, Chemokine und Oberflächenmarker) (Tabelle 6, Literaturquellen sind in der Tabelle angegeben).

nTh ^E (Sallusto)		Tfh ^E (Fazilleau)		NK ^E (Akbari, Stock)		Th1 ^E (Szabo, Kim et)		Th2 ^E (Cosmi)		Th9 ^E (Ma, Tangye)	
Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID
CD62L	793191	CD200	808165	KLRB1	796105	STAT4	805777	GATA3	796436	PU.1	794786
CXCR4	792221	TNFSF	816362	NKG2	796118	TBX21	800802	STAT6	792610	IL9	811425
CCR7	805546	CD40L	817018	IL4	810797	Hlx	790989	CCR3	807938	IRF4	811655
IL2	801503	CD44	793934	IFNG	796478	IRF1	811401	CCR4	807844	STAT6	792610
CD45R	810269	IL6R	790578	NCR1	803138	CXCR6	807937	CCR8	807888	GATA	796436
		BTLA	808951	NCR2	811943	CD195	809329	PTGDR	797436		
		PDCD1	806029	NCR3	817977 817851 817973	IL12RB1	803538	IL4	810797		
		ICOS	804770	NLRK1	796115	IL12RB2	790220	IL5	811402		
		IL21R	799429	KLRC1	796118	IL18R1	804403	IL6	813180		
		CD57	795288	IL22	796480	IFNG	796478	IL13	810797		
		CXCR5	794433	IFNG	796478	LTA	817797 811813 817925	IL9	811425		
		OX40	791141	IL10	792390	PRF1	793416				
				IL21R	799429	GNLY	804323				
				IL1R1	804399						
Th17 ^E (Annunziato)		Th22 ^E (Zhang, Pan)		Treg ^E (Hill, Feuerer)		B ^E (Gibson, Dong et)		DC ^E (Fanger)		Mac ^E (Murray)	
Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID	Gene	Aff. ID
RORC	792008	AHR	813161	FOXP3	817263	CD19	799448	HLA-	817948	CD14	811461
IL-23R	790218	IL22	796480	CD25	793191	CD20	794028	CD40	806315	CD40	806315
CCR6	812336	IL26	796479	CTLA-4	804769	CD22	802783	CD86	808203	CD11b	799509
CCR4	807844	IL13	810797	GITR	791140	CD40	806315	CD83	811698	EMR1	802510
KLRB1	796105	CCR4	807844	GARP	795055	OCT2	803712	CD11c	799512	CD68	800451
IL23R	790218	CCR6	812336	CCR7	801503	Pax5	816121	LAMP3	809234	CSF1R	811507
IL17A	812021	CCR10	801568	CCR4	807844	CD79A	802913	IFNA	815462	IL4R	799428
IL17F	812702	STAT3	801560	TNFRSF	802505	BLK	814462	MHCII	811514		
IL21	810270			TNFRSF	791141	POU2AF	795159	ICAM-1	802560		
IL22	796480			PDCD1	806029			CD123	816575 817632		
IL26	796479			RUNX1	807019			CD64	790504		
CCL20	804886			IKZF2	805867			CD32	790675		
IL12RB	790220			ENTPD1	792951						
				GPR83	795104						

Tabelle 6: Charakteristische Gene von verschiedenen T-Zellen.

Die Gengruppen (Tabelle 6) wurden von der Autorin der vorliegenden Arbeit mittels GlobalAncova je Stimulation analysiert worden und die Ergebnisse in Tabelle 7 zusammengefasst. Signifikante Ergebnisse (permutierter p-Wert < 0.05) sind in dicken Lettern dargestellt. Bei den KEGG-Signalwegen gibt es in fast jeder Gruppe eine signifikante differentielle Regulation, außer in den Gruppen „p53“ und „Composite“. Die Signalwege TGF-β, TLR, Kalzium und mTOR sind am signifikantesten unterschiedlich reguliert (grau schattiert in

Tabelle 7). Auffällig sind gehäufte signifikante Ergebnisse zwischen den Gruppen unter LpA-Stimulation. Bei den Zelltypen gibt es signifikante Unterschiede bei Th2-Zellen, Treg-Zellen, B-Zellen und Makrophagen (grau schattiert in Tabelle 7).

Analyse	Anzahl	PBMCs								
		AA vs GK (14 vs 14)			NAA vs GK (8 vs 14)			NAA vs AA (8 vs 14)		
KEGG Signalwege	Gene	Nicht- stimuliert	CD3/28	LpA	Nicht- stimuliert	CD3/28	LpA	Nicht- stimuliert	CD3/28	LpA
MAPK	282	0.085	0.471	0.011	0.111	0.226	0.055	0.102	0.096	0.028
TGF-β	91	0.108	0.312	0.000	0.017	0.214	0.115	0.225	0.395	0.051
Apoptose	88	0.035	0.169	0.015	0.142	0.212	0.129	0.097	0.195	0.239
NK cyto	142	0.006	0.445	0.123	0.090	0.276	0.227	0.071	0.380	0.131
Cyto-cyto	270	0.282	0.653	0.049	0.394	0.398	0.072	0.404	0.164	0.153
p53	69	0.175	0.438	0.385	0.059	0.245	0.723	0.490	0.092	0.619
TLR	102	0.149	0.286	0.024	0.181	0.251	0.015	0.057	0.043	0.040
Comp	72	0.393	0.583	0.141	0.383	0.200	0.088	0.135	0.093	0.102
TCR	111	0.015	0.196	0.042	0.079	0.156	0.065	0.141	0.673	0.061
NOD	59	0.057	0.473	0.033	0.055	0.271	0.018	0.088	0.245	0.193
Chemo	199	0.214	0.821	0.067	0.542	0.565	0.063	0.444	0.339	0.050
JAK	159	0.037	0.284	0.023	0.122	0.102	0.095	0.185	0.103	0.193
Kalzium	185	0.017	0.270	0.017	0.128	0.150	0.017	0.037	0.283	0.005
mTOR	58	0.015	0.006	0.034	0.405	0.148	0.089	0.202	0.003	0.188

Zelltypen										
nTh	5	0.079	0.586	0.092	0.689	0.738	0.303	0.267	0.223	0.844
Tfh	12	0.430	0.549	0.290	0.576	0.293	0.162	0.733	0.232	0.176
NK	14	0.527	0.366	0.337	0.149	0.563	0.187	0.183	0.619	0.487
Th1	13	0.153	0.429	0.282	0.098	0.538	0.491	0.535	0.916	0.422
Th2	11	0.244	0.492	0.047	0.310	0.162	0.026	0.339	0.108	0.728
Th9	5	0.643	0.241	0.637	0.093	0.702	0.170	0.135	0.400	0.271
Th17	13	0.156	0.634	0.425	0.088	0.163	0.515	0.148	0.069	0.547
Th22	8	0.198	0.304	0.278	0.066	0.410	0.353	0.082	0.051	0.094
Treg	14	0.034	0.273	0.099	0.428	0.319	0.507	0.036	0.857	0.798
B	9	0.152	0.020	0.250	0.086	0.157	0.103	0.271	0.109	0.343
DC	11	0.337	0.768	0.327	0.113	0.163	0.138	0.223	0.117	0.082
Mac	7	0.291	0.826	0.12	0.006	0.237	0.027	0.050	0.234	0.006

Tabelle 7: Regulation der KEGG Signalwege und T-Zell Gruppen-spezifische Gene.

Zusammenfassung der Unterschiede in der Regulation von KEGG-Signalwegen und Expression von T-Zell spezifischen Genen. Jeder Phänotyp wird für jede Stimulationsbedingung separat verglichen. Die Zellen geben den permutierten p-Wert des GlobalAncova Tests wider.

3.2.4.1 Differenziell regulierte Signalwege

In den nächsten Abschnitten werden die KEGG-Signalwege in Kooperation (Rädler 2014) genauer untersucht und mittels IPA (Ingenuity) dargestellt. Zunächst werden die Effektgrößen des Vergleichs der Phänotypen AA gegen GK unter LpA-Stimulation des TGF- β Signalweges dargestellt (Abbildung 18).

TGF-β Signaling

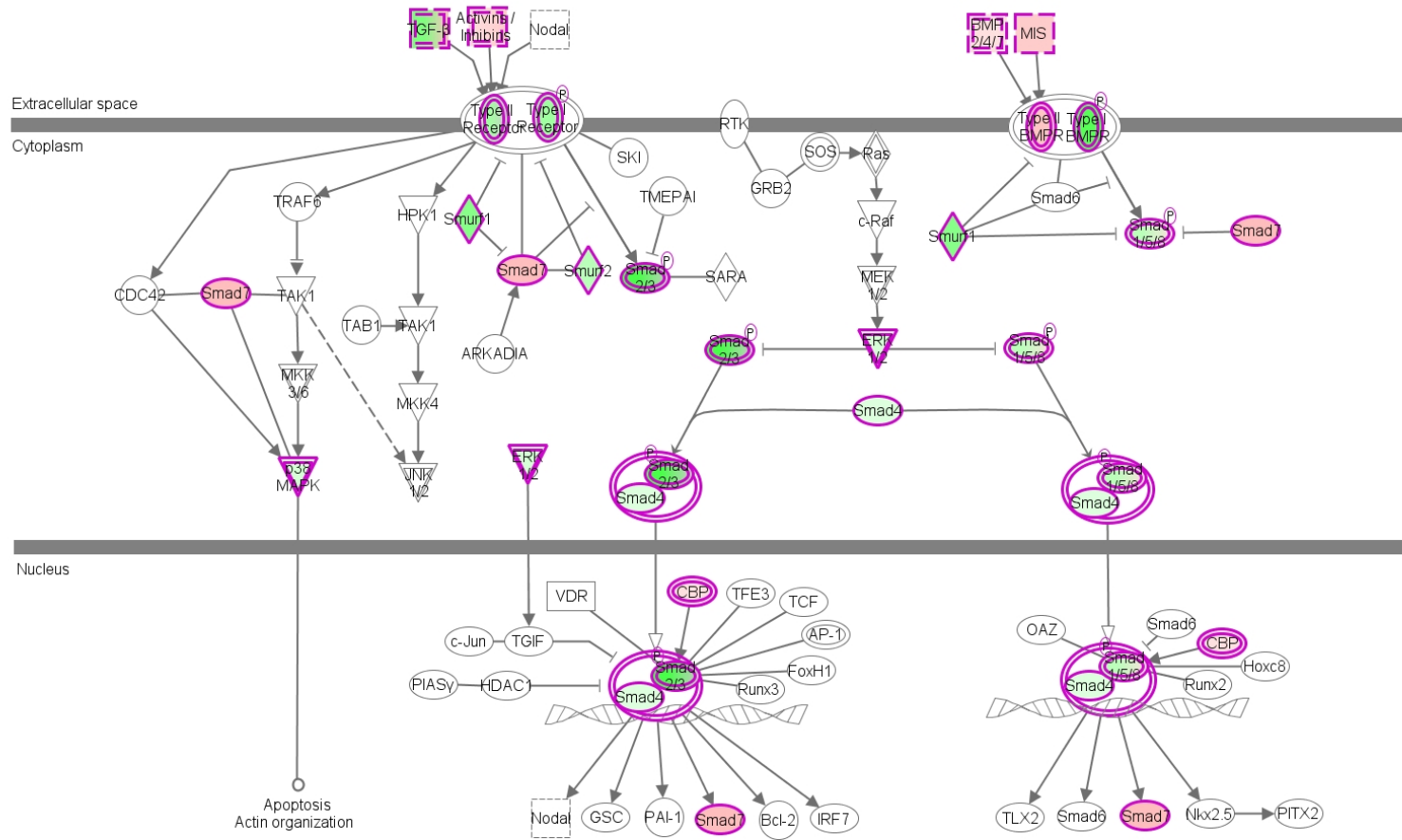


Abbildung 18: Expressionsunterschiede im TGF- β -Signalweg zwischen AA und GK unter LpA-Stimulation.

Die Regulation der TLR-Gene von AA und NAA gegen GK unterscheidet sich am stärksten nach LpA-Stimulation und ist am signifikantesten zwischen NAA und GK (Tabelle 7, p-Wert=0.024/0.015/0.040). Die Effektgröße des Vergleichs zwischen NAA und GK unter LpA-Stimulation der TLR Gengruppe ist mit Hilfe von IPA (Ingenuity) in Abbildung 19 dargestellt (Legende siehe Anhang 6.5).

Toll-like Receptor Signaling

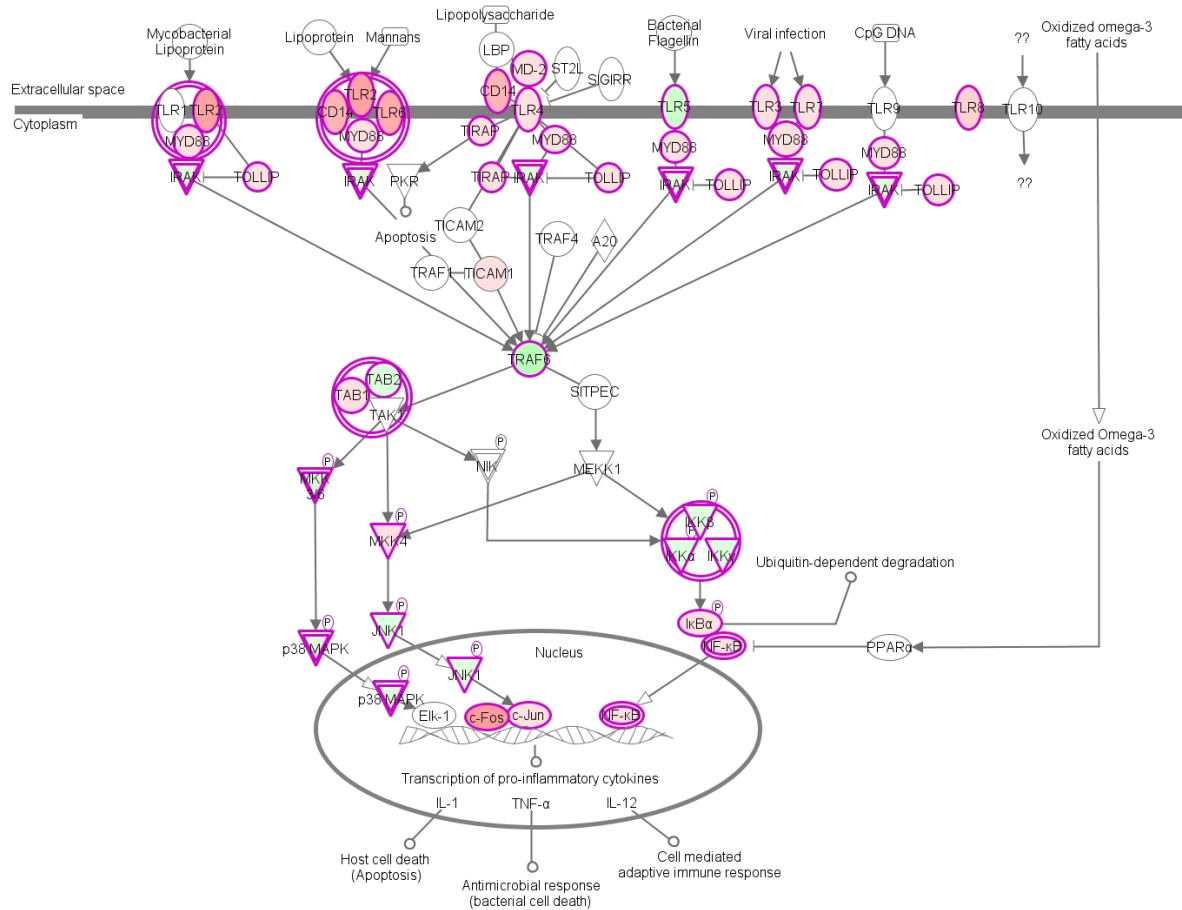


Abbildung 19: Expressionsunterschiede im TLR-Signalweg zwischen NAA gegen GK unter LpA-Stimulation.

Die Expression des Kalzium-Signalwegs ist zwischen den drei Phänotypen ohne Stimulation und nach LpA-Stimulation unterschiedlich (Tabelle 7, alle p-Werte < 0.04; außer zwischen NAA und GK nicht-stimuliert M). Die transkriptionelle Regulation des mTOR-Signalwegs ist zwischen NAA und AA nach CD328-Stimulation und zwischen AA und GK bei allen drei Stimulationsbedingungen unterschiedlich (Tabelle 7; p-Wert: 0.015/0.006/0.034/0.003; Abbildung 20).

mTOR Signaling

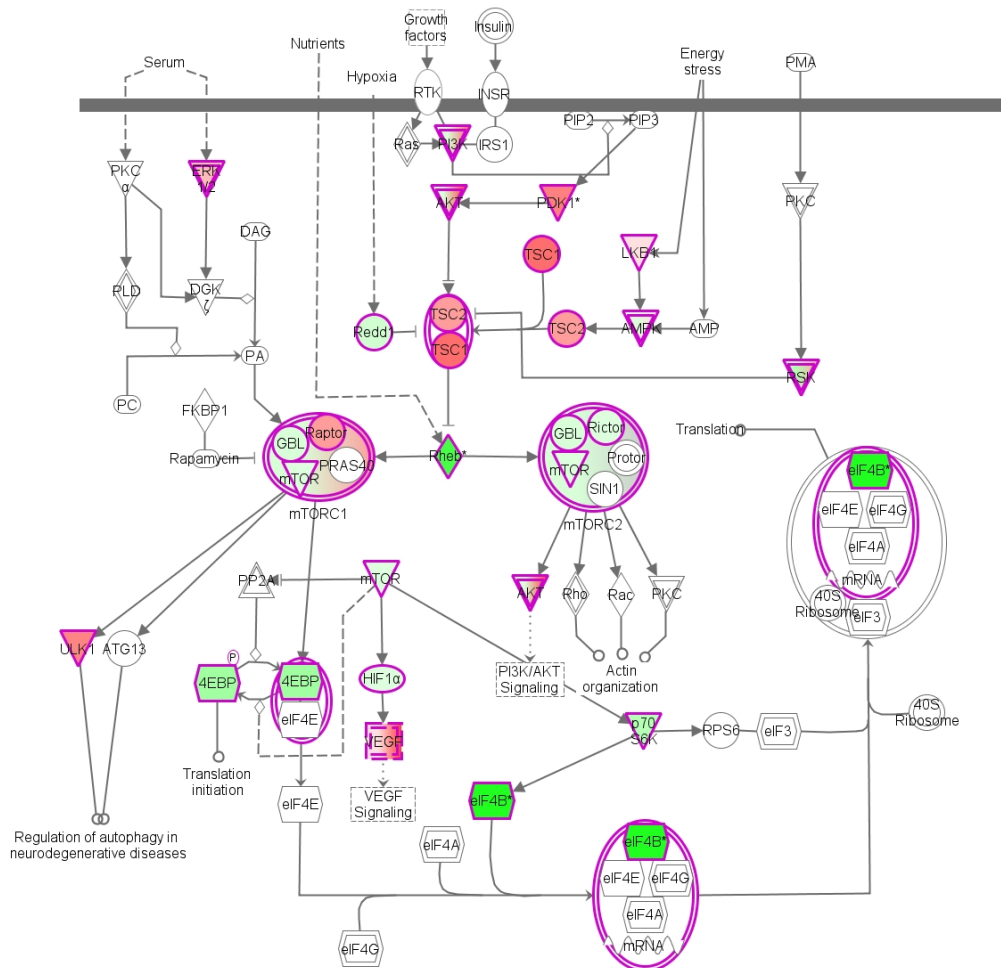


Abbildung 20: Expressionsunterschiede im mTOR-Signalweg zwischen AA gegen GK unter M-Stimulation.

3.2.4.2 Differentielle Genregulation von T-Zellgruppen

Die Gengruppen der T-Zelltypen sind kleiner als die Gengruppen der KEGG-Signalwege (Tabelle 7) und daher kann die differentielle Genregulation der einzelnen Komponenten der Gruppen mittels Balkendiagrammen visualisiert werden (Rädler 2014). Die transkriptionelle Regulation der Zelltypen unterscheidet sich am stärksten in den Gruppen Th2, Treg- und Makrophagen (Tabelle 7). Die Gene der Th2-Gruppe sind am stärksten unterschiedlich reguliert zwischen AA, NAA und GK nach LpA-Stimulation (Tabelle 7, p-Werte: 0.05/0.03). Die Expression ist stärker bei den Asthmatikern (AA und NAA) im Vergleich zur Gruppe GK (gelb: hochreguliert in AA/NAA; rot: hochreguliert in GK). Die Balkendiagramme (Abbildung 21 A und B) zeigen, dass vor allem das Zytokin IL-6 zu dem signifikanten Ergebnis beiträgt. Der Einfluss des einzelnen Gens wird durch die Reduzierung der Quadratsumme berechnet.

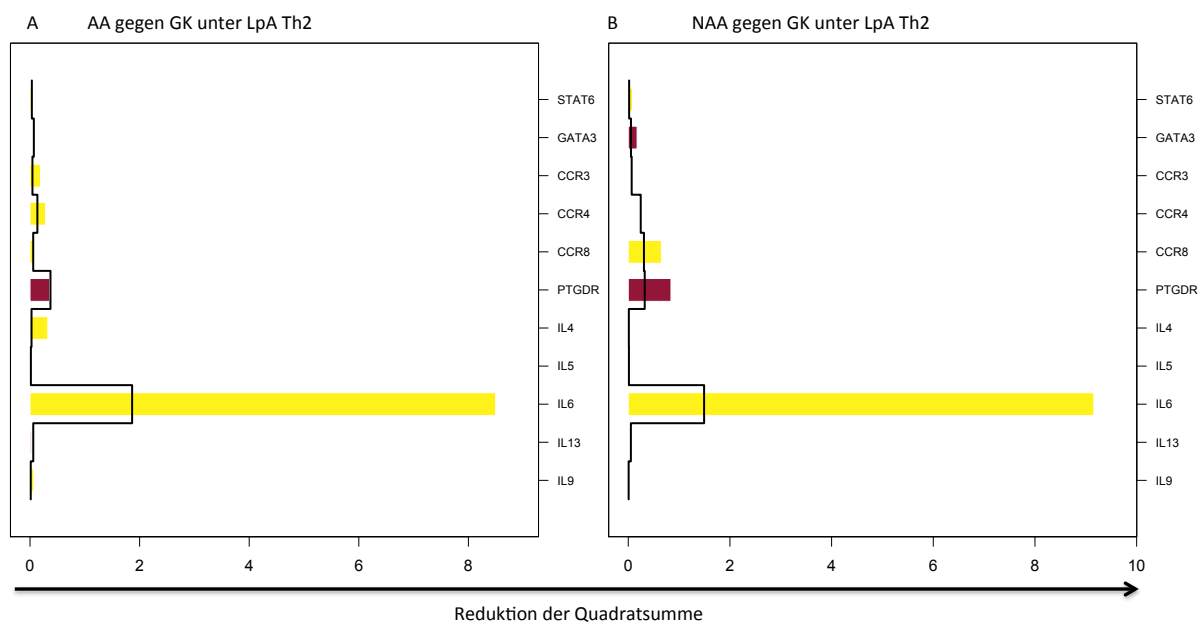


Abbildung 21: Balkendiagramm des Einflusses der einzelnen Gene der Th2-Zellen auf das Ergebnis der Teststatistik.

Die Expression der Treg-assoziierten Gene ist signifikant unterschiedlich in den nicht-stimulierten (M-Stimulation) PBMCs der AA gegen NAA oder GK (Tabelle 7, beide p-Wert < 0.04; Abbildung 22 A und B). Im Vergleich des Phänotyps AA gegen GK ist vor allem das Gen FOXP3 und IL2RA signifikant unterschiedlich reguliert (gelb: hochreguliert in AA; rot: hochreguliert in GK) (Abbildung 21A). Bei dem Vergleich NAA gegen AA sieht man jedoch, dass FOXP3 und

TNFRSF18 signifikant unterschiedlich reguliert sind (gelb: hochreguliert in NAA; rot: hochreguliert in AA). In beiden Analysen ist FOXP3 im Phänotyp AA höher reguliert.

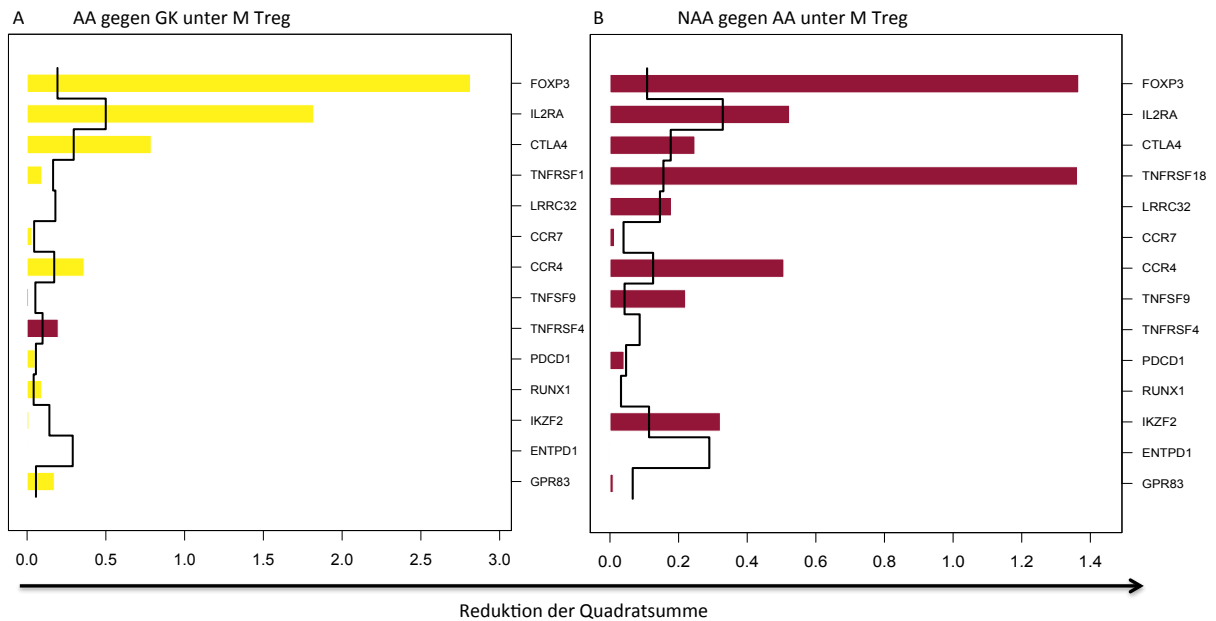


Abbildung 22: Balkendiagramm des Einflusses der einzelnen Gene der Treg-Zellen auf das Ergebnis der Teststatistik.

Die Gene der Makrophagen Gruppe sind signifikanter exprimiert in NAA im Vergleich zu AA und den GK ohne Stimulation und nach LpA-Stimulation (Tabelle 7; alle p-Werte<0.05; Abbildung 23). Kein Gen der Makrophagen ist im Phänotyp GK hochreguliert im Vergleich zu NAA (gelb: hochreguliert in NAA; rot: hochreguliert in GK). Auch im Vergleich der beiden Phänotypen NAA gegen AA unter LpA-Stimulation sind alle Gene in NAA hochreguliert (gelb: hochreguliert in NAA; rot: hochreguliert in AA). Die Hochregulation in NAA des Gens ITGAM trägt vor allem zu dem signifikanten Ergebnis der Gengruppe der Makrophagen bei.

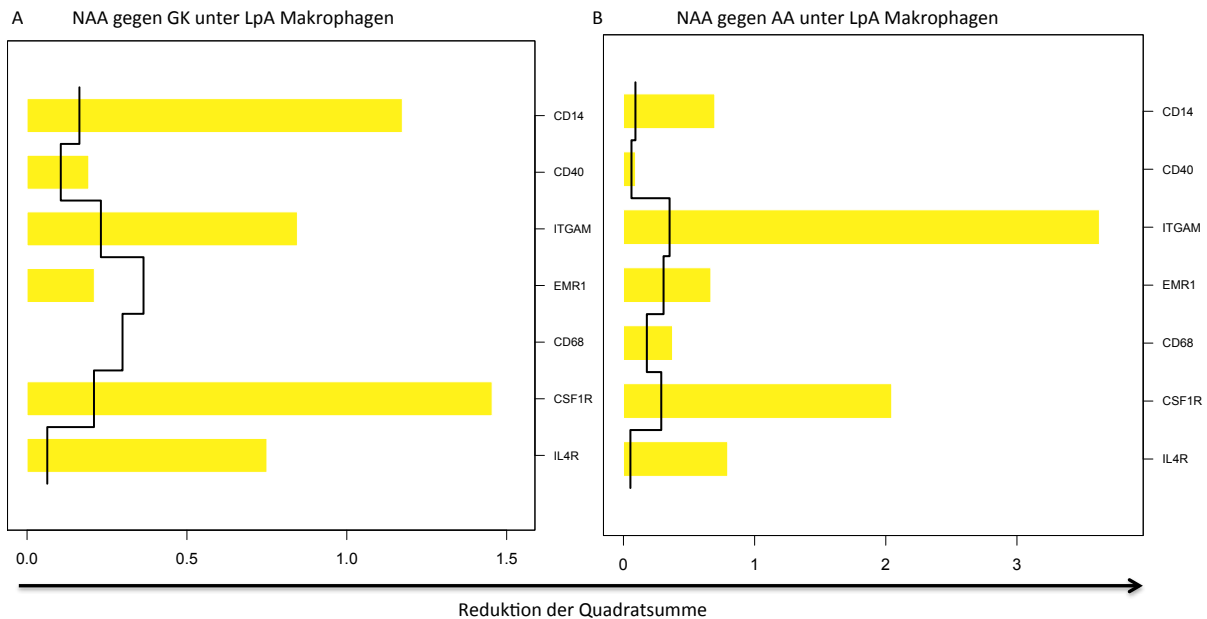


Abbildung 23: Balkendiagramm des Einflusses der einzelnen Gene der Makrophagen auf das Ergebnis der Teststatistik.

3.2.5 Differentielle Genregulation in CD4⁺-Zellen

Die CD4⁺-Zellen stellen eine Untergruppe der PBMCs dar und werden aus den PBMCs sortiert (Rädler 2014). Es wird nur mit M stimuliert. Im Folgenden wird der transkriptionelle Unterschied zwischen den CD4⁺-Zellen und den PBMCs untersucht. Dabei werden nur die PBMC-Microarrays betrachtet, von denen es auch CD4⁺-Microarrays gibt. Es werden also nur zusammenpassende Proben analysiert, die von demselben Probanden stammen. Beide Strategien der Systembiologie werden angewendet, top-down Analyse und bottom-up Analyse.

3.2.5.1 Top-down Analyse der CD4⁺-Zellen

Mittels linearer Regression untersuchte die Autorin dieser Dissertation die Transkriptionsdaten der CD4⁺-Zellen auf Einzelgenebene untersucht. Dabei wird die Kovariate P auf die Genexpression Y abgebildet: $Y \sim P$. Nach Adjustierung des p-Werts gibt es keine signifikant differentiell regulierten Gene (p-Wert < 0.05). Bei den passenden PBMC-Microarrays wird dasselbe Modell auf die Genexpressionsdaten angewendet. Nach Adjustierung des p-Wertes sind 79 Gene differentiell reguliert.

3.2.5.2 Bottom-up Analyse der CD4⁺-Zellen

Die bottom-up Analyse der CD4⁺-Zellen basiert auf GlobalAncova. Dabei analysierte die Autorin dieser Arbeit die differentielle transkriptionelle Regulation von Gengruppen (Tabelle 6) analysiert. Das Ergebnis der CD4⁺-Zellen wurde in Kooperation (Rädler 2014) mit dem Ergebnis der GlobalAncova Analyse der passenden PBMCs verglichen (Tabelle 8). Die transkriptionelle Regulation der Gene in den KEGG Signalwegen ist in CD4⁺-Zellen vergleichbar mit den PBMCs.

Zusätzlich wird ein signifikanter Unterschied der Regulation von MAPK- und BC-Genen in den passenden PBMCs identifiziert. Dagegen sind die Treg-Gene in den passenden PBMCs nicht differentiell exprimiert (Tabelle 8). Die transkriptionelle Regulation der KEGG-Signalwege und T-Zell-spezifischen Gene ist nicht unterschiedlich in den passenden PBMCs und den CD4⁺-Microarrays, außer bei nTh-Genen (Tabelle 8, p-Wert=0.018, Abbildung 24).

KEGG Analyse	Anzahl Gene	PBMC s		CD4 ⁺
		AA vs GK_nicht-stimuliert		
		(14 vs 14)	(5 vs 3)	(5 vs 3)
MAPK	282	0.085	0.0179	0.179
TGF-β	91	0.108	0.286	0.732
Apoptosis	88	0.035	0.018	0.714
NK cyto	142	0.006	0.019	0.732
Cyto-cyto	270	0.282	0.732	0.732
p53	69	0.175	0.732	0.732
TLR	102	0.149	0.196	0.714
Comp	72	0.393	0.732	1.000
TCR	111	0.015	<0.001	0.179
NOD	59	0.057	0.714	0.714
Chemo	199	0.214	0.732	0.732
JAK	159	0.037	<0.001	0.196
Calcium	185	0.017	0.018	0.732
mTOR	58	0.015	0.018	0.196
Zelltypen				
nTh	5	0.079	0.467	0.018
Tfh	12	0.430	0.300	0.179
NK	14	0.527	1.000	0.982
Th1	13	0.153	0.400	0.714
Th2	11	0.244	0.567	0.196
Th9	5	0.643	0.333	0.982
Th17	13	0.156	0.400	0.714
Th22	8	0.198	0.167	0.196
Treg	14	0.034	0.300	0.196
B	9	0.152	<0.001	1.000
CD	11	0.337	0.267	0.732
Mac	7	0.291	0.333	0.714

Tabelle 8: Regulation der KEGG Signalwege und T-Zell Gruppen-spezifischer Gene im Vergleich PBMCs gegen CD4⁺-Zellen

Das Genexpressionsprofil der Gene der Gruppe nTh in Abbildung 24 (schwarz: Hochregulation in AA) zeigt, dass die Hochregulation in der Gruppe AA des Gens Sell wesentlich zu dem signifikanten Ergebnis der Gruppe nTh beiträgt.

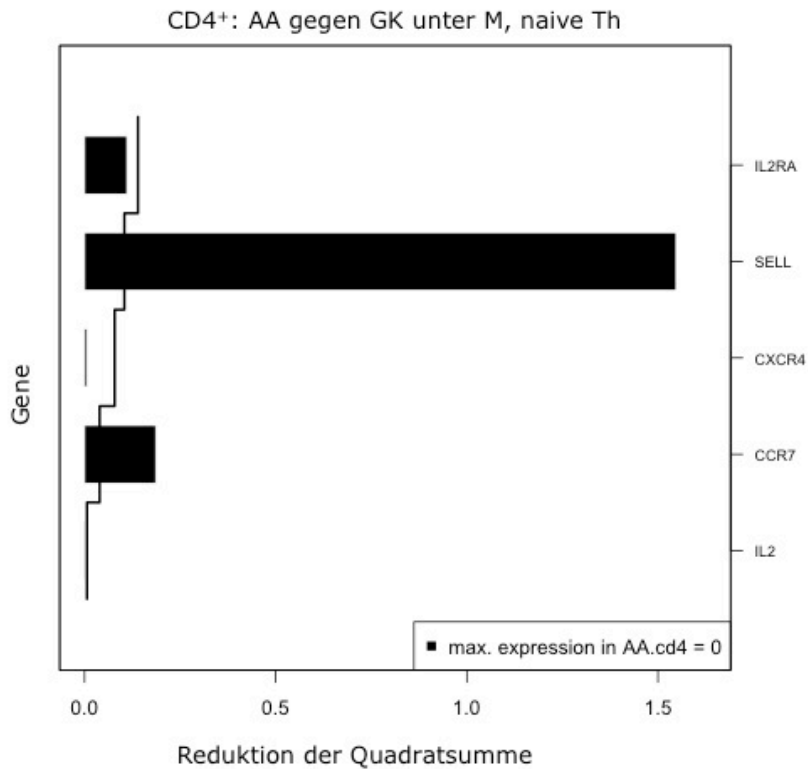


Abbildung 24: Das Balkendiagramm der CD4⁺-Zellen zeigt den Einfluss der einzelnen Gene der naiven Th-Zellen auf das Ergebnis der Teststatistik auf.

Die Analyse der Asthmadaten hat gezeigt, dass der Signalweg mTOR in AA differentiell reguliert ist, während der TLR-Signalweg in NAA unterschiedlich reguliert ist. In beiden Asthmaphänotypen ist die T-Zellgruppe Th2 erhöht im Vergleich zu den gesunden Kontrollen. Die Analyse der CD4⁺-Zellen zeigt nicht mehr differentiell regulierte Signalwege auf.

4 Diskussion

Die Definition von Systembiologie ist nicht eindeutig. Einer der Gründe liegt in der delikaten Balance von experimentellen und mathematischen Ansätzen (Kitano 2002). Die Richtungen der zugrundeliegenden Methoden der Systembiologie unterscheiden sich (Abbildung 1): Top-down Methoden starten bei dem Datensatz eines Systems, z.B. bei einem Transkriptionsdatensatz, um Interaktionen und Komponenten eines Systems zu analysieren. Bei bottom-up Methoden liegt ein *a priori* Wissen über die Datenstruktur vor, welche zu größeren Einheiten zusammengefügt werden soll.

Top-down und bottom-up Methoden analysieren die Systemarchitektur und versuchen funktionelle Eigenschaften zu identifizieren (Bruggeman and Westerhoff 2007). In dieser Arbeit werden beispielhaft beide Strategien der Systembiologie anhand von Transkriptionsdatensätzen angewendet. Dazu wird als visuelles Hilfsmittel der Modellselektion das Eruptionsdiagramm eingeführt.

4.1 Verfahren zur Bestimmung von differentieller Genexpression: Evidenzbasierte Verfahren und Erstellung von Genranglisten

Es gibt zwei verschiedene Verfahren zur Bestimmung differentieller Genexpression: Das evidenzbasierte Verfahren oder die Erstellung von Genranglisten. Das experimentelle Design entscheidet dabei, welche Verfahren sinnvoll sind. Wenn bei dem betreffenden Experiment keine Replikate vorliegen, bietet sich die Erstellung von differentiell exprimierten Genranglisten an. Die Liste wird durch die Stärke der differentiellen Expression sortiert, startet bei den meist hoch-regulierten Genen und endet bei den meist runter-regulierten Genen. Die Länge der Liste wird durch eine gewisse Anzahl von zu erwartenden differentiell exprimierten Genen erstellt. Sie kann aber auch durch Festlegung eines Grenzwertes der Stärke der differentiellen Genexpression bestimmt werden. Dieses Verfahren dient dazu, die richtigen Gene zu finden. Wenn das experimentelle Design Replikate umfasst, kann man durch evidenzbasierte Methoden entscheiden, ob Gene signifikant differentiell exprimiert sind. Einige Verfahren mit Evidenzanspruch haben sich etabliert: der moderate T-Test im Paket „Limma“ (Smyth 2004), der T-Test im Paket „SAM“ (Significance Analysis of Microarrays) (Tusher, Tibshirani et al. 2001), die „shrinkage T-Statistik“ (Opge-Rhein and Strimmer 2007) oder die Intensitätsbasierte moderate T-Statistik (Sartor, Tomlinson et al. 2006).

In der vorliegenden Arbeit wird der moderate T-Test auf verschiedene Datensätze angewendet. Dabei dient das Eruptionsdiagramm als visuelles Hilfsmittel zur Modellselektion um lineare Regressionsmodelle zu selektieren.

4.2 Modellselektionskriterien in Verbindung mit dem Eruptionsdiagramm

Es wird ein evidenzbasiertes Verfahren, der moderate T-Test, angewendet. Vier geschachtelte Kandidatenmodelle werden in einer hierarchischen Struktur zur Analyse des Infektionsdatensatzes angeordnet (Tabelle 4). Modell 1 ist das komplexeste lineare Regressionsmodell. Sukzessive werden von Modell 1 Kovariaten eliminiert bis man bei Modell 4, dem einfachsten Modell, landet. Die hierarchische Ordnung der Modelle erlaubt es, eine rückwärts-gewandte Selektionsstrategie zu nutzen. Die Modelle werden solange verändert, bis man eine Verbesserung der Ergebnisse entsprechend der Selektionskriterien erhält. Das Hauptziel ist es demnach, den Effekt auf die Interaktion $H:T$ durch Streichung von Kovariaten zu bestimmen. Die Ergebnisse der Regressionsmodellen können zum Beispiel mit Hilfe des Determinations-Koeffizienten R^2 bestimmt werden. R^2 gibt an, welcher Teil der vorhandenen Variation von Y von der Variation in X erklärt wird und beschreibt die Anpassungsgüte des kompletten Modells, gibt jedoch keine Information über den Effekt einer einzelnen Kovariate. Des Weiteren gibt R^2 keine Information über die Effektschätzer und p-Werte separat für jedes einzelne Gen an. Weitere Selektionskriterien wie AIC (Akaike Information Criterion) und BIC (Schwarz Bayesian Information Criterion) geben ebenfalls einen Wert für einen Modellvergleich an, aber nicht für jedes einzelne Gen.

Die Tests einzelner Gene können aber auch auf globale Tests von Gengruppen wie GlobalAncova (Hummel, Meister et al. 2008) oder globaltest (Goeman, van de Geer et al. 2004) erweitert werden. Dabei ist das Ergebnis des Modellvergleichs nicht auf Gen-Ebene zu erkennen. Der Effekt wird übergreifend über alle Gene berechnet. Es bedarf also eines visuellen Hilfsmittels, um einen Überblick über alle Gene zu erhalten. Aus diesem Grund wird in dieser Arbeit das Eruptionsdiagramm (Prazeres da Costa, Hoffman et al. 2014) eingeführt, ein einfaches und intuitives Mittel, um die Ergebnisse einer einzelnen Kovariate aus zwei unterschiedlichen Modellen für jedes Gen visuell zu vergleichen (siehe Methoden). Das Eruptionsdiagramm ist eine Erweiterung des Vulkandiagramms (Li 2012), ein häufig verwendetes Punktdiagramm in der Microarray Analyse, das auf der x-Achse die Effektschätzer und auf der y-Achse die Signifikanz (p-Werte) repräsentiert. Dabei entspricht jeder Punkt einem Gen. Beim Eruptionsdiagramm werden die Ergebnisse von zwei verschiedenen Vulkandiagrammen aus zwei unterschiedlichen Modellen derselben Kovariate übereinander gelegt. Hier entspricht ein Gen einem Pfeil, der Startpunkt des Pfeils entspricht dem Ergebnis des ersten Modells und der Endpunkt dem Ergebnis des zweiten Modells. Damit gibt das Eruptionsdiagramm visuell einen Überblick über alle Gene hinweg und deckt die Unterschiede hinsichtlich Effektschätzer und Signifikanz zwischen zwei Modellen derselben Kovariate auf (Abbildung 7). In der Analyse des Infektionsdatensatzes wird die Interaktionskovariate $H:T$ zwischen zwei Modellen verglichen.

Jedoch können auch die Effekte auf andere Kovariaten untersucht werden, wie aus der Analyse des Asthmadatensatzes ersichtlich wird (Abbildung 15).

Das Eruptionsdiagramm ist nicht auf die Analyse von Transkriptionsdatensätzen beschränkt. Es kann auch in der Regressionsanalyse von anderen hochdimensionalen Datensätzen Verwendung finden, wie etwa in Metabolomics- und Proteomics-Daten.

4.3 Pipelines zur Analyse differentieller Genexpression

Im Folgenden werden verschiedene Pipelines zur Analyse von Genexpressionsdaten vorgestellt. Es wird untersucht, ob das Eruptionsdiagramm in die Pipelines integriert werden kann. Die Relevanz des Eruptionsdiagramms liegt insofern nahe, da auf Gen-Ebene ein Modellvergleich interpretiert werden kann.

Generell kann man bei Microarray-Analyse-Verfahren zwischen explorativen (top-down) und konfirmatorischen (bottom-up) Strategien unterscheiden. Der bottom-up Ansatz startet bei den Daten und endet bei der Hypothese. Explorative Methoden generieren Hypothesen durch Modellanpassung. Es werden Strukturen in den Daten gesucht, die Hinweise auf Hypothesen geben. Die generierten Hypothesen müssen in einer Folgestudie überprüft werden, ehe sie als wissenschaftlich abgesichert gelten. Die Ergebnisse der explorativen Methode sind statistisch auffällig.

Im Gegensatz dazu ist der bottom-up Ansatz ist motiviert durch ein Modell, das bei der Hypothese startet und bei den Daten endet. Konfirmatorische Studien sichern *a priori* formulierte Hypothesen ab unter Einhaltung eines Signifikanzniveaus α . Die Fragestellung kann klar definiert und ein entsprechender Versuchsplan aufgestellt werden. Sie kann durch Datenerhebung überprüft oder klar und eindeutig beantwortet werden. Die Ergebnisse müssen statistisch signifikant sein.

Weder explorative noch konfirmatorische Analysen alleine sind ausreichend, es werden beide Arten benötigt (Tukey 1980) und kommen daher in dieser Arbeit zur Anwendung.

4.3.1 Explorative bzw. top-down Verfahren

Explorative Methoden umfassen top-down oder Hypothesen generierende Ansätze der Systembiologie. Beispielhaft wird eine Analysepipeline dargestellt. Diese Pipeline besteht aus mehreren Analyseschritten. Zunächst wird eine Gen-weise differentielle Expression analysiert. Dazu werden zwei verschiedene Methoden angewendet, Limma und SAM. Im nächsten Schritt können die Ergebnisse der differentiellen Genexpression entweder klassifiziert oder auf Anreicherung von Gengruppen untersucht werden. Es gibt zwei verschiedene Methoden der Klassifizierung: supervised und un-supervised. Bei der Analyse der Anreicherung von Gengruppen wird untersucht, ob die Extreme der Genliste mit Kategorien angereichert sind. In

diesem Zusammenhang soll nun festgestellt werden, ob das Eruptionsdiagramm in explorative Methoden eingebaut werden kann.

4.3.1.1 Gen-weise Analyse von differentieller Expression

Das Eruptionsdiagramm findet bei der statistischen Analyse als visuelles Hilfsmittel zur Veranschaulichung von differentieller Genexpression Verwendung. Dabei müssen zwei Ebenen bedacht werden. Bei Gen-weisen Analysen wird jedes Gen einzeln untersucht. Die Annahme liegt zugrunde, dass Genexpressionen unabhängig voneinander sind.

In der Umgebung bioconductor (Gentleman, Carey et al. 2004) stehen zum Beispiel „Limma“ (Smyth 2004) und SAM (Tusher, Tibshirani et al. 2001) zur Verfügung. Die Funktion „anova“, um ANOVA anzuwenden, kann direkt in R (Team 2013) für den Vergleich mehrerer Gruppen als abhängige Variable verwendet werden.

Das Paket „Limma“ (Smyth 2004) wurde bereits in dieser Arbeit vorgestellt und zur Berechnung linearer Regressionsmodelle benutzt. Außerdem wurde das Eruptionsdiagramm für die Modellselektion verwendet. Das Ergebnis wird mit Hilfe einer Anreicherungsanalyse in 4.5 diskutiert.

Im Folgenden soll nun das Anwendungspotential des Eruptionsdiagramms in dem Paket „SAM“ bestimmt werden. SAM berechnet eine Statistik, um die Signifikanz der Genexpressionsunterschiede zu bestimmen. Dabei wird mit Hilfe von Gen-weisen T-Tests die Stärke des Zusammenhangs von Genexpression und einer Antwortvariablen gemessen. Die Analyse basiert auf einer nicht-parametrischen Statistik, da die Daten möglicherweise nicht normalverteilt sind. Die unabhängige Variable X repräsentiert die Expressionsmatrix und die abhängige Variable Y den experimentellen Faktor. Das lineare Modell berechnet nun, welchen Einfluss Y auf die Genexpression X hat. Eine quantitative Antwortvariable ist der lineare Regressionskoeffizient der Gene auf das Ergebnis.

Dabei können nun zwei verschiedene Modelle aufgestellt werden, die sich hinsichtlich dieser Antwortvariablen unterscheiden. Die Ergebnisse beider Modelle können jeweils in einem Vulkandiagramm aufgetragen werden, das an der x-Achse den Expressionsunterschied und an der y-Achse den FDR-Wert aufzeigt. Allerdings ist eine Visualisierung in einem Eruptionsdiagramm nicht möglich, in welchem die Ergebnisse eines Modellvergleichs stets von derselben abhängigen Variablen dargestellt werden, während im vorliegenden Fall die Modelle unterschiedliche unabhängige Variablen enthalten.

4.3.1.2 Klassifikation von Genen

Bei den Analysepipelines folgt nun die Klassifikation der differentiellen Genexpression durch Clustering. Man unterscheidet zwei Methoden, supervised und unsupervised Clustering (Allison, Cui et al. 2006).

Bei supervised Clustering werden Gene bereits existierenden Kategorien zugeordnet, während bei unsupervised Clustering neu gebildete Kategorien zugewiesen werden. Unsupervised Methoden generieren Hypothesen, bei denen keine Annahmen über Strukturen der Daten bekannt sind. Unsupervised Cluster Verfahren von Genen sind z.B. Self-organizing maps SOM (Kohonen 1990), k-means Clustering (MacQueen 1967) als Funktion „kmeans“ und hierarchisches Clustering als Funktion „hclust“ in R. In dieser Arbeit wird hierarchisches Clustering angewendet, um differentiell exprimierte Gene in aggravating und alleviating zu klassifizieren 3.1.5.

4.3.2 Konfirmatorische Gengruppenanalysen bzw. bottom-up Verfahren

In der Forschung beinhalten konfirmatorische Methoden top-down oder Hypothesen-testende Ansätze. Im Speziellen werden hier Analysen zur differentiellen Expression von Gengruppen diskutiert. Im Folgenden werden beispielhaft die Methoden GSEA (Subramanian, Tamayo et al. 2005; Subramanian, Kuehn et al. 2007), GlobalAncova (Hummel, Meister et al. 2008) und SAM-GS (Dinu, Potter et al. 2007) herangezogen. Diese Methoden basieren auf *a priori* Wissen über Datenstrukturen.

Ein globaler Test ist eine statistische Methode, um eine Vereinigungsmenge mehrerer Nullhypothesen, also eine globale Nullhypothese, zu testen.

GlobalAncova (Hummel, Meister et al. 2008) ist eine Form des globalen Tests auf differentielle Expression in funktionellen Gengruppen. Die Nullhypothese lautet, dass kein Gen differentiell exprimiert ist. Das Ziel von GlobalAncova ist es, die Zusammenhänge zwischen der Expression einer Gruppe von Genen X mit einer interessierenden Variablen Y zu quantifizieren. Es wird also nach Gengruppen gesucht, die eine koordinierte Expressionsveränderung unter den Bedingungen von Y zeigen. Die Funktion GlobalAncova kann zwei geschachtelte Modelle mittels ANOVA vergleichen. Das Testergebnis spiegelt sich in der F-Statistik, im permutierten p-Wert und im approximierten p-Wert wider. Da der globale Test für eine Gruppe von Genen durchgeführt wurde, aus dem das Ergebnis für jedes einzelne Gen nicht erkennbar ist, kann an dieser Stelle das Eruptionsdiagramm nicht zum Einsatz kommen.

Bei GSEA handelt es sich um einen kompetitiven Test, da die Nullhypothese lautet, dass Gene in einer Gengruppe höchstens so oft differentiell exprimiert sind wie in der Gruppe aller anderen Gene. Deshalb wird nach einer Anreicherung einer Gengruppe mit differentiell exprimierten Genen gefragt. GSEA Verfahren werden in zwei Schritten ausgeführt. Im ersten Schritt werden die Gene durch eine Gen-weise Analyse in einer Liste angeordnet. Diese Liste gibt an, wie stark eine zu interessierende Gruppe an den Extremen der Liste mit Kategorien angereichert ist. Im

zweiten Schritt wird untersucht, ob die Gen-Ränge angereichert sind. Dazu wird der Kolmogorov-Smirnov-Test angewendet. Ein Score wird berechnet, indem man der geordneten Liste abwärts folgt und eine additive Liste erhöht, wenn sich ein Gen in der Gruppe befindet und erniedrigt wird, wenn sich das Gen nicht in der Gruppe befindet. Die Signifikanz dieses Scores wird ermittelt, indem ein empirischer Phänotyp-basierter Permutationstest durchgeführt wird. Somit kann man entscheiden, welche Werte wirklich extrem sind. Es wird für multiples Testen durch Permutation der Beobachtungen adjustiert.

SAM-GS ist ein in-sich geschlossener Test. Dieses Verfahren basiert auf einer Zusammenfassung von Gen-weisen Scores zu Gruppenstatistiken. Hierfür wird die SAM-Statistik (4.3.1.1) verwendet. Die SAM-Statistiken der Gene innerhalb einer interessierenden Gengruppe werden quadriert und aufsummiert. Dadurch erhält man wieder ein globales Maß für differentielle Expression der Gengruppe. Hierbei wird nach „gleichgerichteter“ Expression gefragt und nicht nach der „absoluten“ Statistik. Ein Nachteil gegenüber dem Verfahren GlobalAncova ist, dass man nur mit einer Kovariate arbeiten kann. Dadurch lassen sich nur Vergleiche zwischen zwei Gruppen durchführen, so dass das Eruptionsdiagramm nicht zum Einsatz kommen kann.

4.4 Interpretation des Eruptionsdiagramms

In dieser Arbeit wird ein passendes lineares Regressionsmodell aufgestellt, indem eine rückwärts-gewandte Modellselektionsstrategie angewendet wird und schrittweise Kovariaten gestrichen werden. Das Eruptionsdiagramm wird verwendet, um den Effekt auf die Interaktionskovariate durch die Streichung von Kovariaten zu veranschaulichen. Dabei hilft die Richtung der Pfeile den Effekt der Streichung der Kovariate zu interpretieren: 1) Effektmodifikation, 2) Störfaktoren oder 3) zusätzliche erklärende Kraft.

Des Weiteren werden in dieser Arbeit geschachtelte lineare Regressionsmodelle auf Microarray Expressionsdaten benutzt, um Interaktionen zwischen mehreren Faktoren und deren Effekte auf die transkriptionelle Antwort zu untersuchen. Ziel ist die Anwendung von Modellen höherer Ordnung auf binäre Faktoren wie Mausstamm (BALB/C und C57BL/6), Phänotyp von Kindern (AA, NAA und GK), Stimulationsbedingungen (IFN- γ und LpA, CD328 und M) und Infektionsstamm (WA(pYV) oder den *Y. enterocolitica* Kontrollstamm WA(pTTS, p60)). Das Eruptionsdiagramm wird als neues Werkzeug entwickelt, um Modelle visuell zu vergleichen und dadurch das passende Modell für einen Datensatz zu identifizieren. Die Richtung der Pfeile wird als Hilfsmittel verwendet, um den Effekt der Streichung von Kovariaten zu interpretieren. So wird gezeigt, dass durch die Hinzunahme der Infektionskovariaten erster Ordnung zusätzliche erklärende Kraft gewonnen wird (Abbildung 9). Außerdem wird dargestellt, dass Effekte zweiter Ordnung in dem Datensatz vorhanden sind, und das korrekte Modell wird basierend auf

statistischer Signifikanz und Effektgröße bestimmt (Abbildung 8). Die Kovariaten dritter Ordnung werden als Störfaktoren identifiziert und aus dem Modell gestrichen (Abbildung 7). Die Interpretation des Eruptionsdiagrammes ist jedoch nicht immer eindeutig, wenn die Pfeile in verschiedene Richtungen zeigen (Abbildung 15). In diesem Fall können Gengruppen, je nach ähnlicher Pfeilrichtung, gebildet werden. In den folgenden Absätzen wird getestet, ob diese Gengruppen mit unterschiedlichen funktionellen Kategorien angereichert sind.

4.5 Anreicherungsanalysen der Ergebnisse der Infektionsdaten

Die Anreicherung von GO-Termen in Gengruppen wird in der vorliegenden Arbeit mit zwei verschiedenen Verfahren, GSEA (Mootha, Lindgren et al. 2003; Subramanian, Tamayo et al. 2005) und DAVID (Huang da, Sherman et al. 2009), getestet. Das GSEA-Verfahren ist bereits in Kapitel 4.3.2 vorgestellt worden. Die zehn signifikantesten Ergebnisse der GSEA-Analyse und DAVID-Analyse sind jeweils in einem Balkendiagramm aufgelistet. Das Balkendiagramm in der $-\log_{10}$ Skala von GSEA zeigt die FDR-Werte und das Balkendiagramm von DAVID die p-Werte.

4.5.1 Signifikante Gene der Interaktionskovariate

Zunächst werden die Ergebnisse der Analyse des Infektionsdatensatzes hinsichtlich der funktionellen Relevanz der Interaktionskovariate diskutiert. Mittels einer rückwärts-gewandten Modelselektionsstrategie wird das passende Regressionsmodell bestimmt. Die Ergebnisse der Regressionsanalyse werden mit Hilfe von Anreicherungsverfahren auf die funktionelle Relevanz hin untersucht.

Die signifikant runter- und hochregulierten Gene der Interaktionskovariate aus Modell 3 (Tabelle 4) werden in einer Anreicherungsanalyse untersucht. Durch die Anreicherungsanalyse wird untersucht, ob Gengruppen statistisch signifikante Unterschiede zwischen zwei Phänotypen zeigen. Die Ergebnisse der Interaktionskovariate von Modell 3 sind in Abbildung 25 in einem Vulkandiagramm dargestellt. Die statistisch signifikanten Gene befinden sich im oberen rechten und linken Bereich der Graphik.

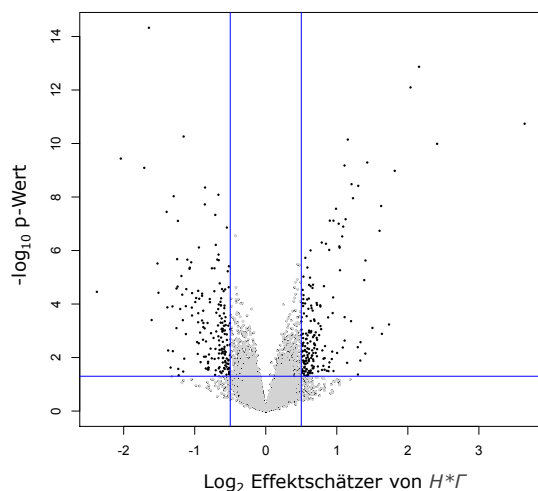


Abbildung 25: Vulkandiagramm der Interaktionskovariate von Modell 3.

Die Gene werden mittels GSEA und DAVID analysiert. Die Ergebnisse der GSEA-Analyse (Abbildung 26) und DAVID-Analyse (Abbildung 27) sind jeweils in einem Balkendiagramm aufgelistet. Auffällig ist der Term „immune system process“, der in beiden Analysen signifikant ist. Dieses Ergebnis passt zu der Datengrundlage, da die IFN- γ -Stimulation eine starke Immunantwort hervorruft.

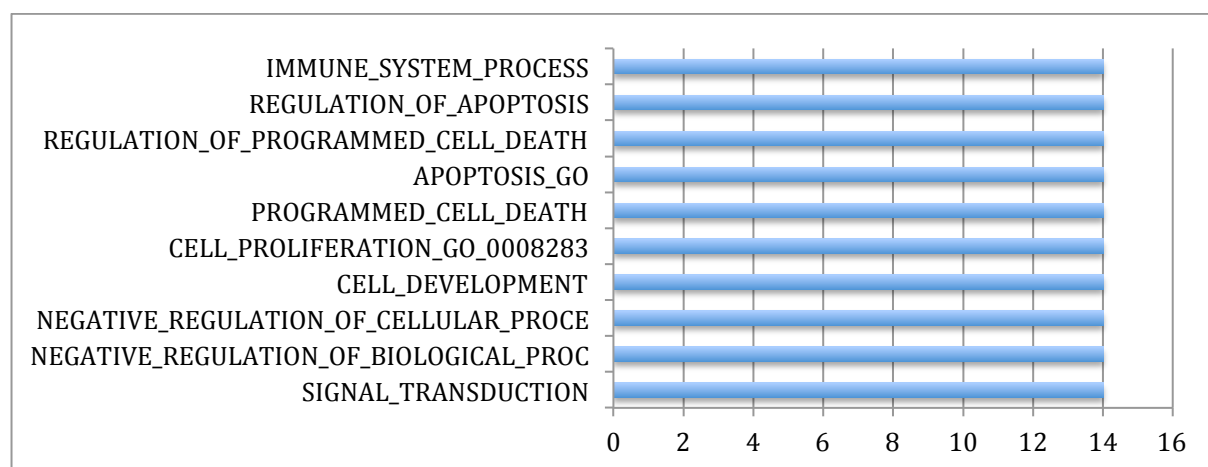


Abbildung 26: GSEA-Analyse von allen signifikanten Genen der Interaktionskovariate von Modell 3.

Die Ergebnisse der GSEA-Analyse (Abbildung 26) zeigen „signal transduction“, „negative regulation of biological process“, und „negative regulation of cellular process“ als stark angereicherte Kategorien. Auch Kategorien, die mit Apoptose und Anhalten von Prozessen assoziiert sind, kommen gehäuft vor: „apoptosis_go“, „programmed cell death“, „regulation of programmed cell death“, „negative regulation of cellular process“ und „negative regulation of biological process“ (Prozesseinstellung oder Abschwächung).

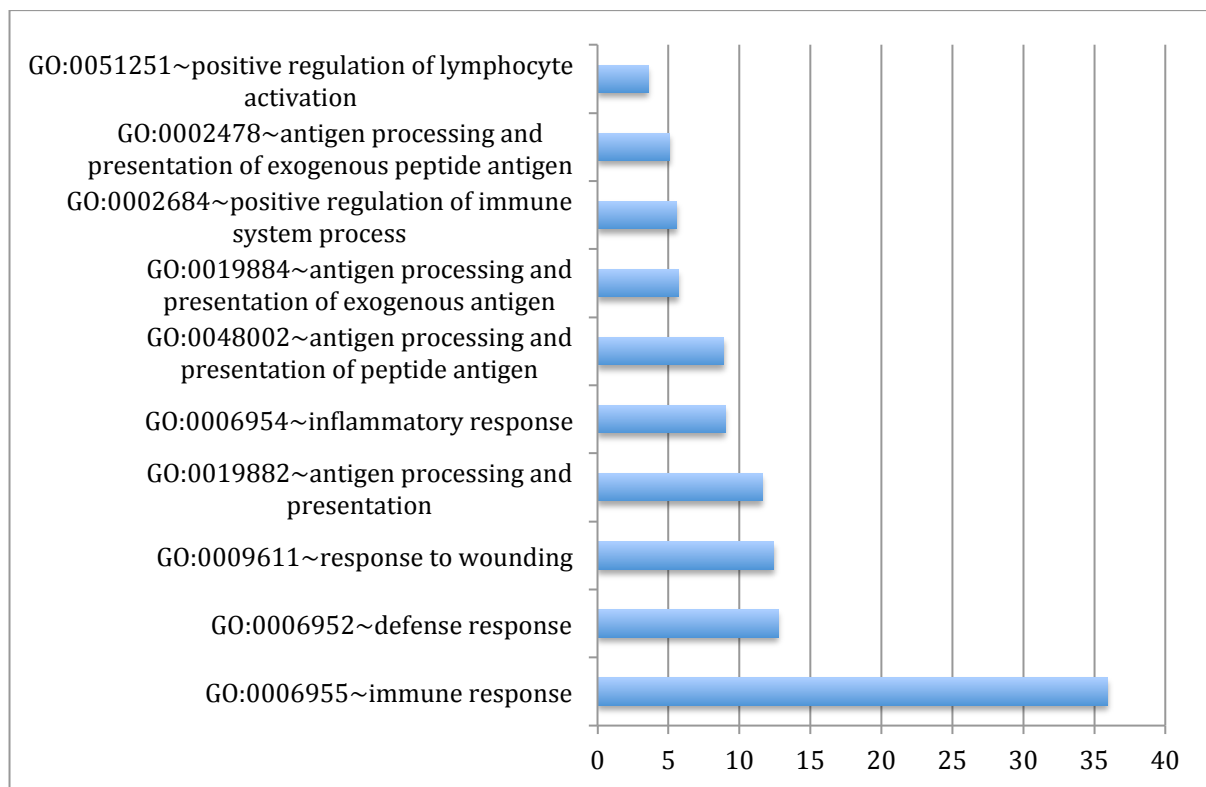


Abbildung 27: DAVID-Anreicherungsverfahren aller signifikanten Gene der Interaktionskovariate von Modell 3.

Die Ergebnisse der DAVID-Anreicherungsanalyse (Abbildung 27) zeigen als stärkste angereicherte Kategorien „immune response“, „defense response“ und „response to wounding“. Alle Kategorien stehen mit der Immunantwort in Verbindung.

Die Ergebnisse der beiden Analysen von GSEA und DAVID unterscheiden sich deutlich. Das liegt einerseits an der ungleichen Datengrundlage und andererseits an den unterschiedlichen zu Grunde liegenden Teststatistiken. Immunsystem-assoziierte Terme kommen in beiden Ergebnissen vor.

4.5.2 Abweichung der zu erwartenden Effekte

Eine genetische Interaktion tritt auf, wenn bei der Mutation eines Gens der Phänotyp von der Mutation eines anderen Gens verdeckt wird (Roth, Lipshitz et al. 2009). Dazu können die Phänotypen einzelner Gendeletionen mit der kombinierten Gendeletion und dem Wildtyp verglichen werden (Beyer, Bandyopadhyay et al. 2007). Wenn der Phänotyp der kombinierten Gendeletion signifikant von dem zu erwartenden Phänotyp (Produkt der unabhängigen einzelnen Deletionen unter Annahme eines multiplikativen Modells) abweicht, spricht man von einer genetischen Interaktion (Beyer, Bandyopadhyay et al. 2007). Ist der Phänotyp der kombinierten Gendeletion größer als der zu erwartende Phänotyp, handelt es sich um eine

alleviating Interaktion. Bei einer aggravating Interaktion ist der Phänotyp der kombinierten Gendeletion dagegen kleiner als der zu erwartende Phänotyp. Der zu erwartende Phänotyp wird unter der Annahme eines multiplikativen Modells berechnet.

In der vorliegenden Arbeit werden Deletionen mit experimentellen Faktoren H und Γ ersetzt und anschließend berechnet, ob die Kombination der einzelnen Faktoren des gemessenen Phänotyps von dem multiplikativen Modell abweicht. Die Kombination von zwei Effekten H und Γ ist unter der Annahme eines multiplikativen Modells der beiden Effekte berechnet werden, $H*\Gamma$. Dabei wird die Kombination wird in einem Modell höherer Ordnung durch die Interaktionskovariate $H:\Gamma$ bestimmt. Abweichungen von der Annahme des multiplikativen Modells können durch die Subtraktion der beiden Terme berechnet werden: $H*\Gamma-H:\Gamma$. Ein negatives Ergebnis wird als alleviating bezeichnet, da $H:\Gamma$ größer als das Produkt von $H*\Gamma$ ist. Ein positives Ergebnis ist aggravating, da die Kombination $H:\Gamma$ kleiner als das erwartete Produkt $H*\Gamma$ ist. Die Effektschätzer von H , Γ , und $H*\Gamma$ und die Abweichung des Produkts und der Kombination $H*\Gamma-H:\Gamma$ sind in 6.2 angegeben. Gene, die bei $H*\Gamma-H:\Gamma$ einen Wert von <-2 und >2 zeigen, werden mittels GO Anreicherungsverfahren analysiert (Huang da, Sherman et al. 2009).

4.5.2.1 Aggravating Effekte

Gene, die aggravating Effekte zeigen, werden durch die Abweichung vom zu erwartenden Effekt berechnet und mittels zwei verschiedener Verfahren auf Anreicherung von Gengruppen analysiert. Beide Ergebnisse sind in einem Balkendiagramm dargestellt.

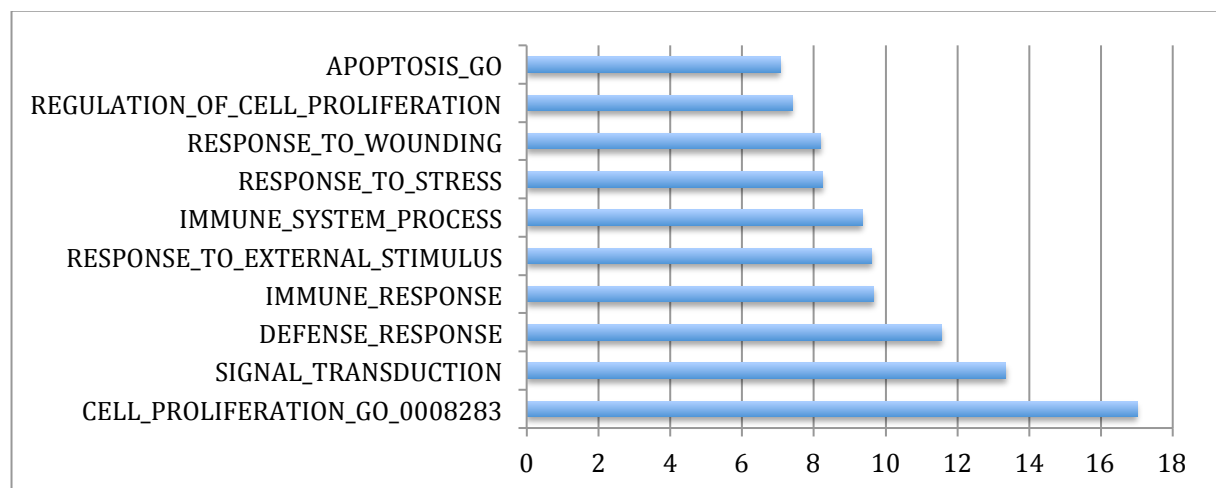


Abbildung 28: Aggravating Effekte der GSEA-Analyse.

Die GSEA-Analyse der aggravating Effekte (Abbildung 28) zeigt die stärkste Anreicherung der Kategorien „cell proliferation“, „signal transduction“ und „defense response“. Die Kategorie „immune response“ ist ebenfalls angereichert.

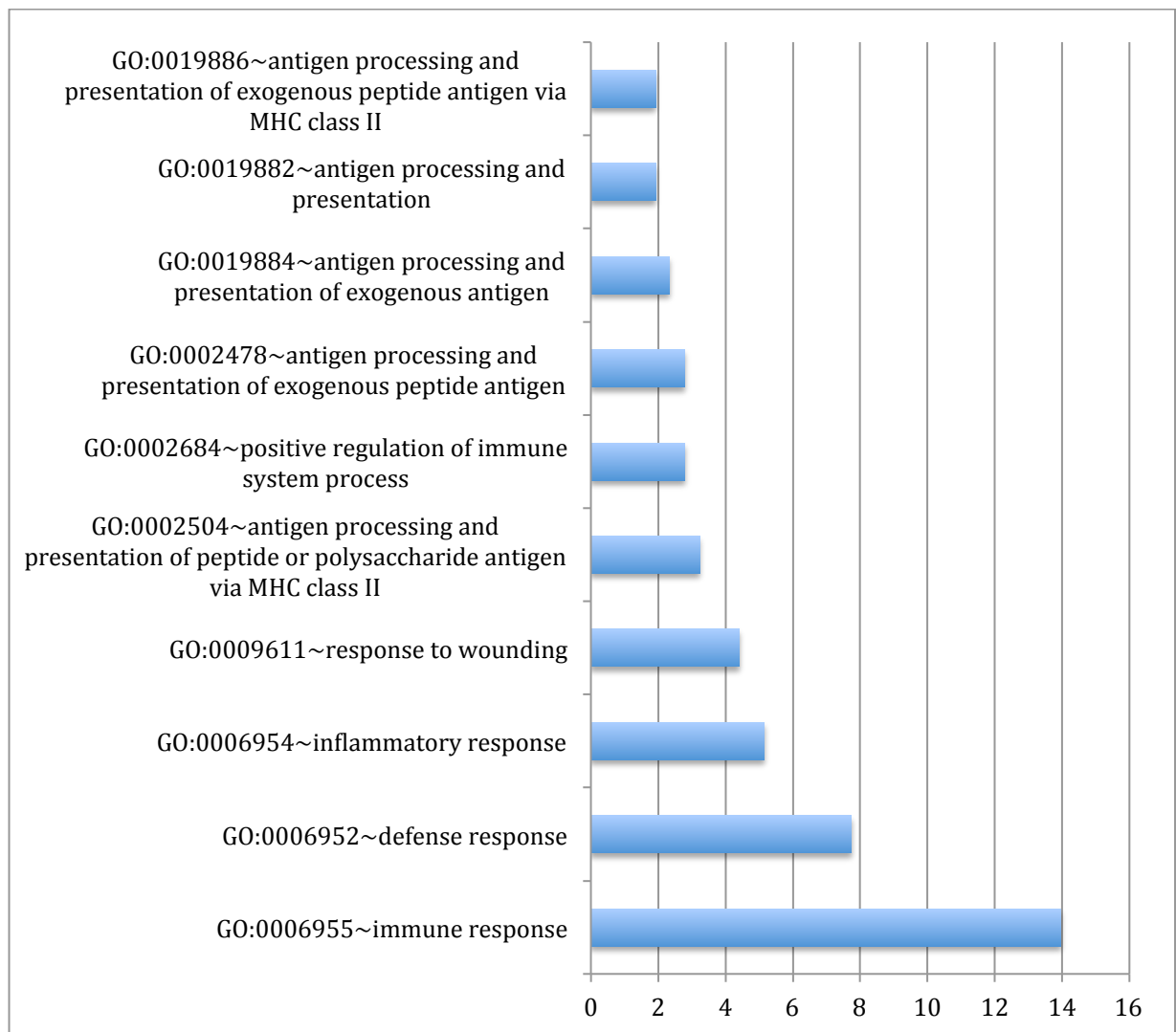


Abbildung 29: Aggravating Effekte der DAVID-Anreicherungsanalyse.

Die Ergebnisse der DAVID-Anreicherungsanalyse (Abbildung 29) zeigen als stärkste angereicherte Kategorien „immune system process“, „defense process“ und „inflammatory response“.

4.5.2.2 Alleviating Effekte

Im folgenden Abschnitt werden Gene auf Anreicherung von Kategorien analysiert, die alleviating Effekte zeigen. Alleviating Effekte sind durch die negative Abweichung des zu erwartenden Effektes gekennzeichnet.

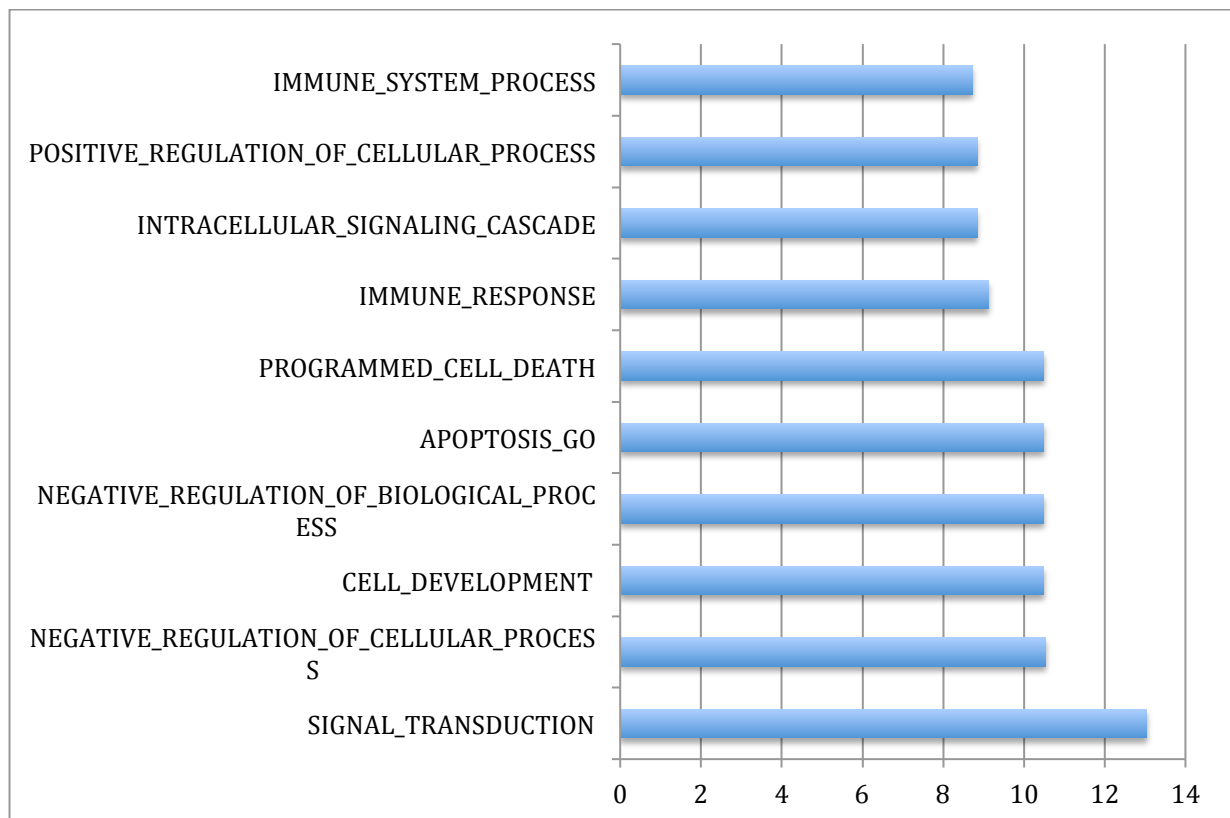


Abbildung 30: Alleviating Effekte der GSEA-Analyse.

Die Ergebnisse der GSEA-Analyse (Abbildung 30) zeigen, dass „signal transduction“, „negative regulation of cellular process“, und „cell development“ am stärksten angereichert sind. Kategorien, die mit Apoptose und Anhalten von Prozessen assoziiert sind, kommen gehäuft vor: „apoptosis“, „programmed cell death“, „negative regulation of cellular process“ und „negative regulation of biological process“ (Prozesseinstellung oder -abschwächung).

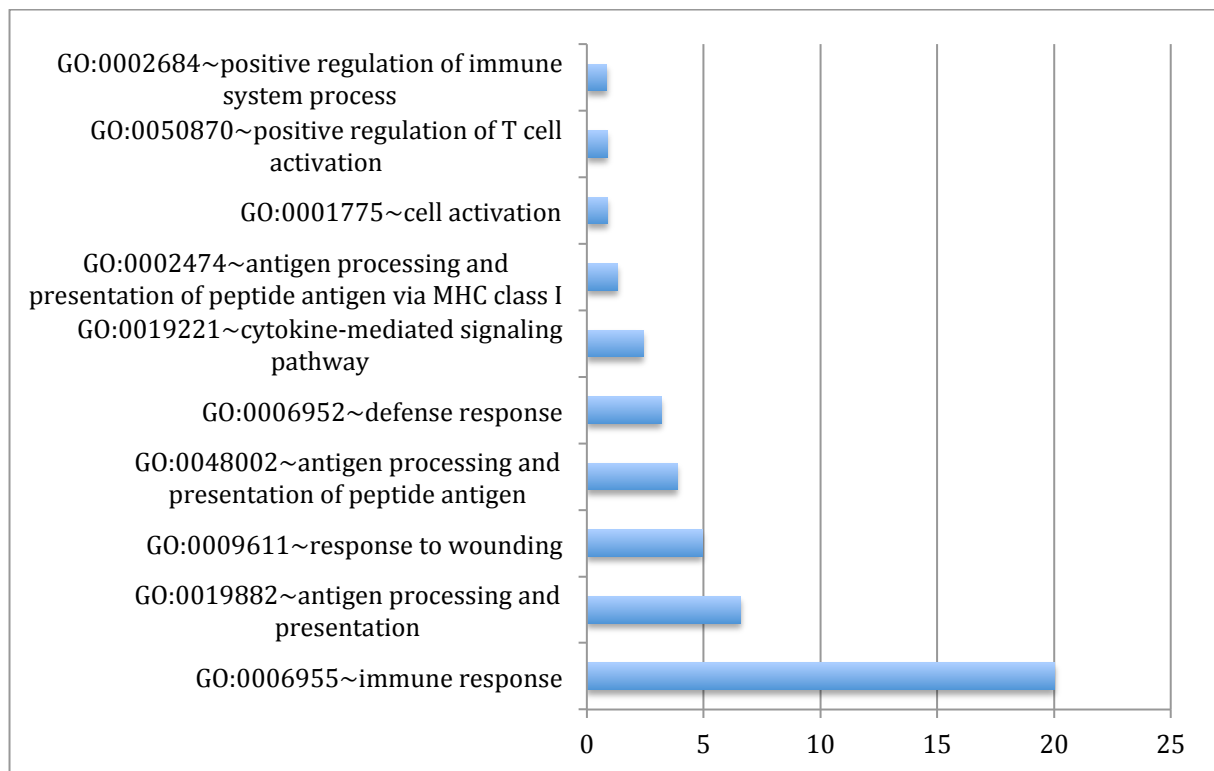


Abbildung 31: Alleviating Effekte der DAVID-Anreicherungsanalyse.

Am stärksten angereicherte Kategorien sind „immune response“, „antigen processing and presentation“ und „response to wounding“ (Abbildung 31). Diese Kategorien sind mit der Immunantwort verknüpft. Die DAVID-Anreicherungsanalyse zeigt keine Kategorien an, die mit Apoptose assoziiert sind.

4.5.3 Vergleich der Ergebnisse: unterschiedliche Ergebnisse der alleviating und aggravating Datensätze

In diesem Abschnitt werden die Ergebnisse aller Gene der Interaktionskovariate mit den Ergebnissen der getrennten Analysen von aggravating und alleviating miteinander verglichen (Tabelle 9). Der Datensatz, der alle signifikant differentiell regulierten Gene der Interaktionskovariate enthält (A), wird analysiert (GSEA: Abbildung 26; DAVID: Abbildung 27). Die Gene werden aufgeteilt in zwei Gruppen und auf Anreicherung von Kategorien untersucht: Gene, die aggravating Effekte (B) (GSEA: Abbildung 28; DAVID: Abbildung 29) und alleviating Effekte (C) (GSEA: Abbildung 28; DAVID: Abbildung 29) zeigen.

Datensatz	GSEA	DAVID
A, B, C	„signal transduction“, „(regulation of) apoptosis“, „immune system process“	„immune response“, „defense response“, „response to wounding“, „antigen processing and presentation“,

		„postive regulation of immune system process“
A, C	„negative regulation of biological process“, „negative regulation of cellular process“, „cell development“, „programmed cell death“, „(regulation of) programmed cell death“	„antigen processing and presentation of peptide antigen“
A, B	„cell proliferation“	„inflammatory response“, „antigen processing and presentation of exogenous antigen“, „antigen processing and presentation of exogenous peptide antigen“
B, C	„apoptosis_go“ , „immune response“	
B	„defense response“, „response to external stimulus“, „response to stress“, „response to wounding“, „regulation of cell proliferation“	„antigen processing and presentation of peptide or polysaccharide antigen via MHC class II“, „antigen processing and presentation of exogenous peptide antigen via MHC class II“
C	„Negative regulation of cellular process“, „Negative regulation of biological process“, „intracellular signalling cascade“, „positive regulation of cellular process“	„cytokine mediated signalling pathway“, „antigen processing and presentation of peptide antigen via MHC class I“, „cell activation, „positive regulation of T cell activation“

Tabelle 9: Vergleich der alleviating und aggravating Anreicherungsanalyse

Der Vergleich der Ergebnisse aus der GSEA-Analyse (Tabelle 9) zeigt, dass die Kategorien „signal transduction“, „(regulation of) apoptosis“ und „immune system process“ in allen drei Datensätzen angereichert sind. Die Kategorien stehen nicht in direktem Zusammenhang. Die DAVID-Analyse weist „immune response“, „defense response“, „response to wounding“, „antigen processing and presentation“ und „postive regulation of immune system process“ in den drei Datensätzen als angereicherte Kategorien auf. Alle Terme sind mit der Immunantwort und Inflammation assoziiert. Die DAVID-Analyse zeigt mehr gemeinsame Kategorien auf im Vergleich zur GSEA-Analyse.

Bei der GSEA-Analyse von A und C überschneiden sich Kategorien: „negative regulation of biological process“, „negative regulation of cellular process, „cell development“, „programmed

cell death“ und „(regulation of) programmed cell death“. Alle Kategorien, außer „cell development“, stehen mit Apoptose in Zusammenhang. Die DAVID-Analyse zeigt hingegen nur eine angereicherte Kategorie auf: „antigen processing and presentation of peptide antigen“.

Es gibt nur eine Kategorie, die sich bei der GSEA-Analyse von A und B überschneidet: „cell proliferation“. Bei der DAVID-Analyse hingegen überschneiden sich drei Kategorien: „inflammatory response“, „antigen processing and presentation of exogenous antigen“ und „antigen processing and presentation of exogenous peptide antigen“. Alle Kategorien sind jedoch mit dem Immunsystem verknüpft.

Zwei Kategorien überschneiden sich bei der GSEA-Analyse von B und C: „apoptosis_go“ und „immune response“. Die Kategorien stehen nicht in direktem Zusammenhang. Die DAVID-Analyse zeigt keine sich überschneidenden Kategorien.

Die Kategorien, die ausschließlich bei der Analyse von B angereichert sind: „defense response“, „response to external stimulus“, „response to stress“, „response to wounding“ und „regulation of cell proliferation“. Alle Kategorien, außer „regulation of cell proliferation“, sind mit Stimulationsantwort verknüpft. Die Kategorien der DAVID-Analyse „antigen processing and presentation of peptide or polysaccharide antigen via MHC class II“ und „antigen processing and presentation of exogenous peptide antigen via MHC class II“ sind beide mit MHC Klasse II Molekülen assoziiert. Die „Eltern-Kategorien“ „antigen processing and presentation of exogenous antigen“ und „antigen processing and presentation of exogenous peptide antigen“ kommen bei der Analyse der drei Datensätzen vor. Die Analyse des Datensatzes B liefert also ein spezifischeres Ergebnis.

Die Ergebnisse der GSEA-Analyse von C sind hauptsächlich mit zellulären Prozessen assoziiert. Die DAVID-Analyse zeigt die Kategorien „cytokine mediated signalling pathway“, „antigen processing and presentation of peptide antigen via MHC class I“, „cell activation“ und „positive regulation of T cell activation“ als angereichert. Die Kategorien sind mit Zellprozessen und der Immunantwort verknüpft.

Interessanterweise sind die Gene der aggravating Effekte mit MHC-Klasse-I assoziiert, während die Gene der alleviating Effekte mit MHC-Klasse-II assoziiert sind.

Es gibt keine Kategorien von A, die entweder in der GSEA-oder der DAVID-Analyse angereichert sind.

Die DAVID-Anreicherungsanalyse der drei Datensätze zeigt ähnliche Ergebnisse wie die GSEA-Analyse. Die DAVID-Analyse zeigt fünf gemeinsame Kategorien und die GSEA-Analyse dagegen drei gemeinsame Kategorien. Die Anreicherung der Kategorien, die mit der Immunantwort assoziiert sind, ist zu erwarten, da hier die Ergebnisse der Interaktionskovariate von H^*I betrachtet werden und die Stimulation mit $IFN-\gamma$ (I) eine Reaktion des Immunsystems hervorruft.

4.6 Anreicherungsanalyse der Gene des Asthmadatensatzes

Für die Analyse des Asthmadatensatzes wird das passende lineare Regressionsmodell mittels eines Eruptionsdiagrammes bestimmt. Das Modell ohne Interaktionskovariate wird mit einem Modell mit Interaktionskovariate verglichen (3.2.2). Bei dem visuellen Vergleich der beiden Modelle durch das Eruptionsdiagramm (Abbildung 15) zeigt sich durch die Pfeilrichtung der Gene, dass sich die Streichung der Interaktionskovariate auf zwei unterschiedliche Arten auswirkt. Einerseits gibt es Pfeile, die bei dem Modellvergleich sowohl eine Veränderung im p-Wert als auch im Effektschätzer zeigen. Die Interaktionskovariate ist in diesem Fall also ein Störfaktor. Zum anderen zeigen Pfeile des Modellvergleichs eine Änderung des p-Wertes und nur eine geringe Veränderung des Effektschätzers. In diesem Fall weist die Interaktionskovariate keine zusätzliche erklärende Kraft auf. Die beiden Gengruppen werden auf Anreicherung von Kategorien hin untersucht.

4.6.1 Alle signifikanten Gene der Kovariate AA

In diesem Abschnitt werden die Gene analysiert, die bei dem Modellvergleich der Kovariate AA signifikante und hohe Effekte ($p\text{-Wert} < 0.05$ und $\text{Effektschätzer} \gg 2/-2$) zeigen. Der Modellvergleich der Kovariate AA ist in einem Eruptionsdiagramm (Abbildung 15) dargestellt. Diese Gene werden mittels einer Anreicherungsanalyse auf Kategorien untersucht.

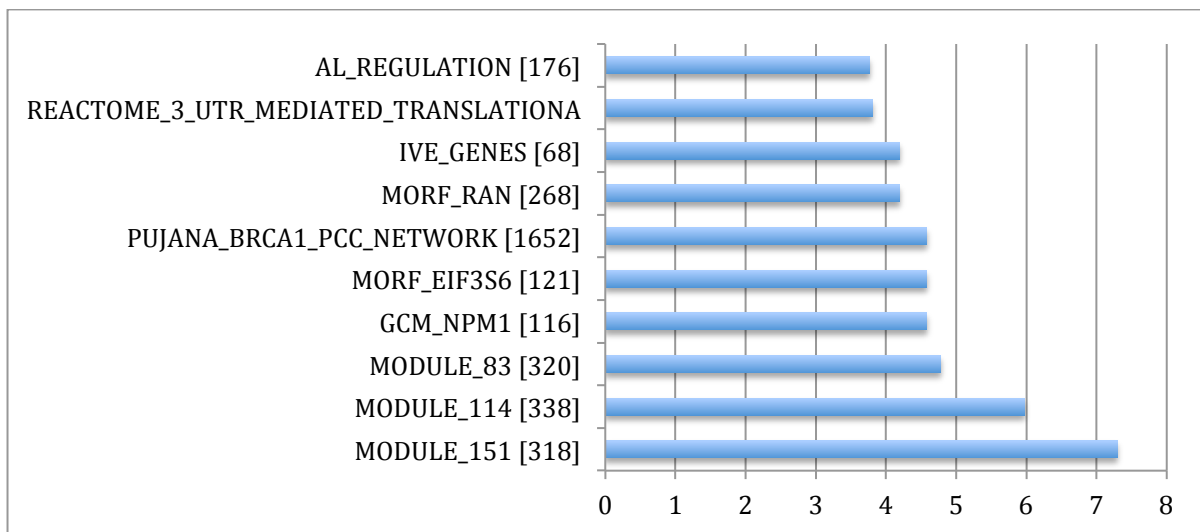


Abbildung 32: GSEA-Analyse aller signifikanten Gene der Kovariate AA.

Die GSEA-Analyse (Abbildung 32) zeigt am stärksten angereicherte Kategorien die mit Krebsmodulen assoziiert sind („module_151“, „module_114“ und „module_83“).

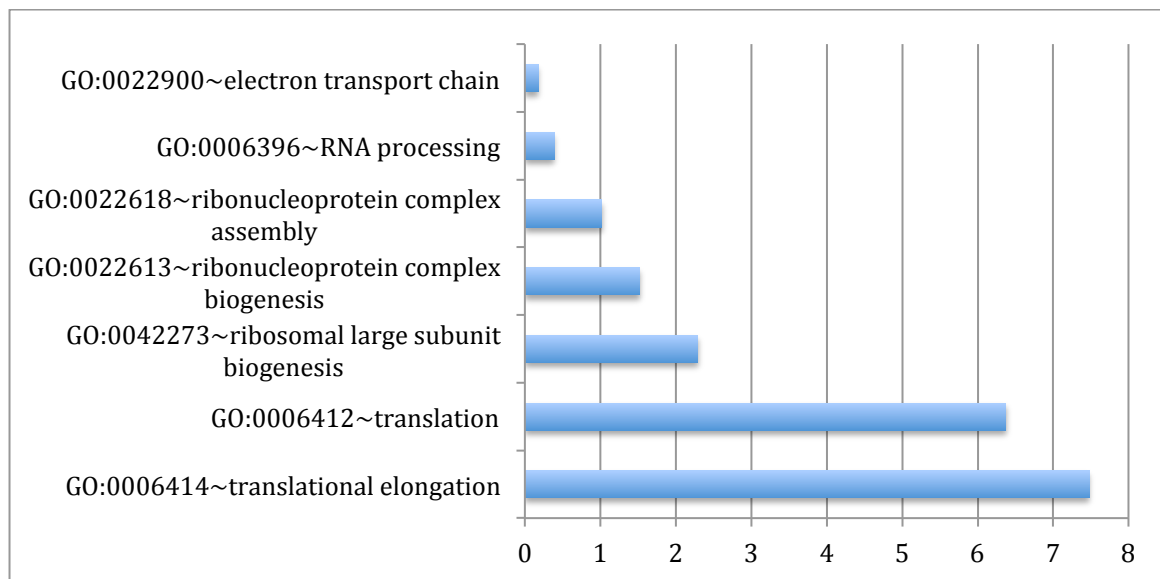


Abbildung 33: DAVID-Anreicherungsanalyse aller signifikanten Gene der Kovariate AA.

Aus der DAVID-Anreicherungsanalyse (Abbildung 33) wird ersichtlich, dass Translationsassoziierte Kategorien am stärksten angereichert sind: „translational elongation“, „translation“, „ribosomal large subunit biogenesis“ und „ribonucleoprotein complex biogenesis“.

4.6.2 Von einem Störfaktor beeinflusste Gene

In diesem Absatz werden die Gene durch Anreicherungsanalysen untersucht, die von einem Störfaktor beeinflusst sind. Die Pfeile der Gene zeigen einen Unterschied im p-Wert und in dem Effektschätzer bei den Ergebnissen beider Modelle.

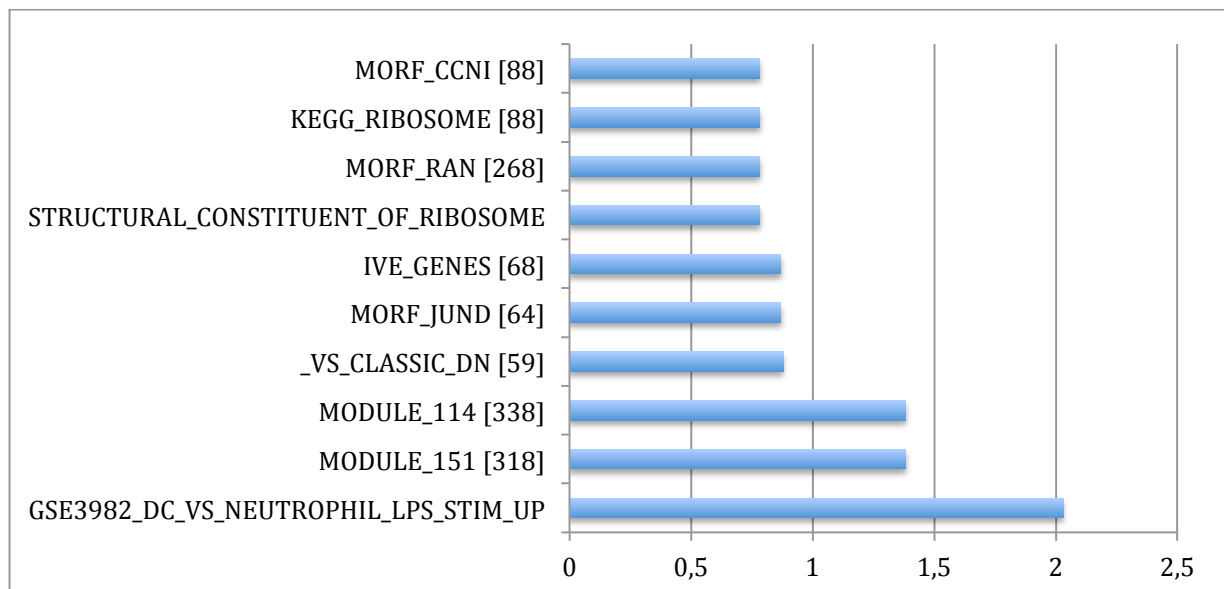


Abbildung 34: GSEA-Analyse der Gene, die von einem Störfaktor beeinflusst werden.

Die GSEA-Analyse (Abbildung 34) zeigt Kategorien signifikant angereichert auf, die bei dem Vergleich von dendritischen Zellen mit LPS zu verschiedenen Zeitpunkten stimuliert werden

(„GSE3982_DC_VS_NEUTROPHIL_LPS_STIM_UP“). Ebenso sind Krebs-Module signifikant angereichert („module_114“ und „module_151“).

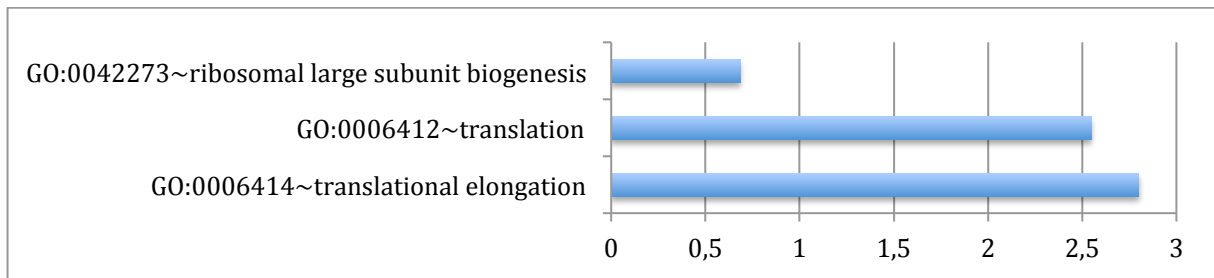


Abbildung 35: DAVID-Anreicherungsanalyse der Gene, die von einem Störfaktor beeinflusst werden.

Die DAVID-Anreicherungsanalyse der Gene, die von einem Störfaktor beeinflusst werden, zeigt zwei signifikant angereicherte Kategorien. Diese sind mit Translation assoziiert: „translational elongation“ und „translation“.

4.6.3 Gene, die bei dem Modellvergleich keine zusätzliche erklärende Kraft der Interaktionskovariate zeigen

In diesem Abschnitt werden die Gene der Kovariate AA analysiert, die bei dem Modellvergleich eine Verbesserung im p-Wert zeigen. Diese Gene ändern den Effektschätzer nicht oder nur minimal. Somit zeigt die Interaktionskovariate, deren Effekt in dem Modellvergleich getestet wird, keine zusätzliche erklärende Kraft. Es werden zwei verschiedene Analyseverfahren auf die Gene angewendet: GSEA- und DAVID-Anreicherungsanalyse.

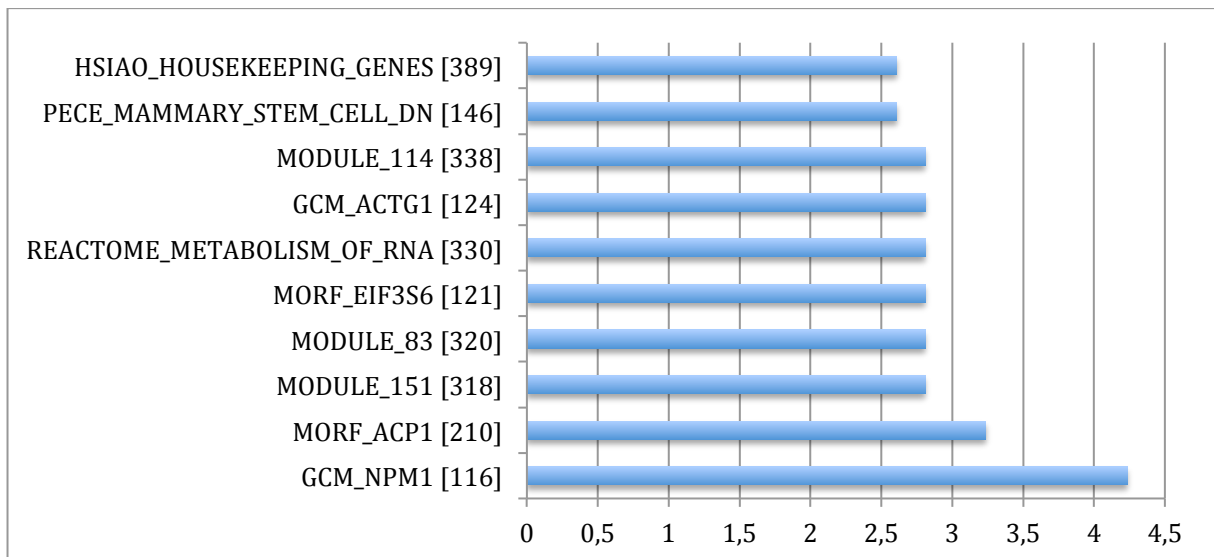


Abbildung 36: GSEA-Analyse der Gene, die eine Veränderung im p-Wert zeigen.

Als stark angereicherte Kategorien zeigen sich „neighborhood of NPM1“ („GCM_NPM1“), „neighborhood of ACP1“ („MORF_ACP1“) und Kategorien, die mit Krebsmodulen assoziiert sind: „module_151“ und „module_83“ (Abbildung 36).

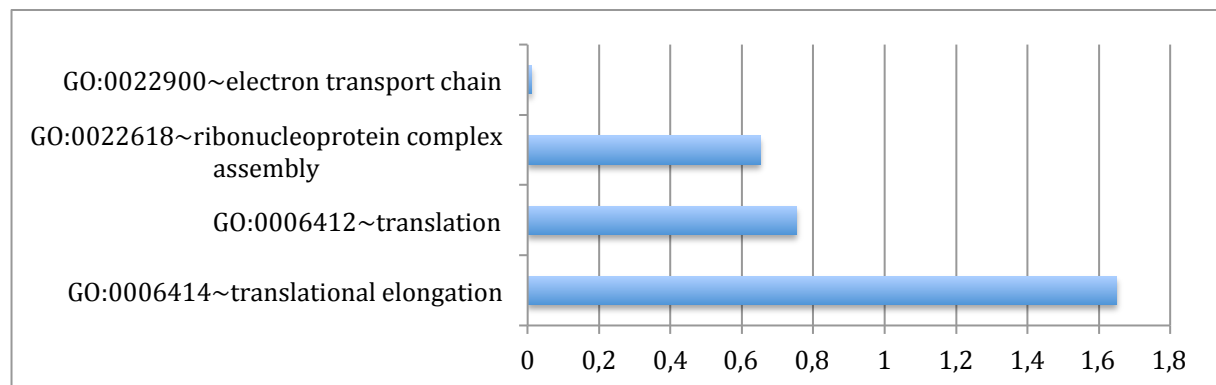


Abbildung 37: DAVID-Anreicherungsanalyse der Gene, die eine Veränderung im p-Wert zeigen.

Die DAVID-Anreicherungsanalyse zeigt nur eine Kategorie als signifikant angereichert: „translational elongation“ (Abbildung 37).

4.6.4 Vergleich der Ergebnisse der Anreicherungsanalysen der Asthmadaten

In diesem Absatz werden die Ergebnisse der Anreicherungsanalysen der Asthmadaten verglichen. Die Tabelle 10 fasst die Ergebnisse der GSEA- (Tabelle 10; Abbildung 34; Abbildung 36) und DAVID- (Tabelle 10; Abbildung 35; Abbildung 37) Analysen zusammen. Beide Anreicherungsverfahren werden auf drei Datensätze angewendet. Der Datensatz, der alle Gene, die bei dem Modellvergleich in 3.2.2 der Kovariate AA signifikant sind, umfasst (A) wird in zwei Teilmengen aufgespalten: Gene, die bei diesem Modellvergleich von einem Störfaktor beeinflusst werden (B) und Gene, die bei dem Modellvergleich keine zusätzliche erklärende Kraft der Interaktionskovariate aufzeigen (C). Kategorien, die sich in den Ergebnissen der Datensätze überschneiden, werden zusammengefasst.

	GSEA	DAVID
A, B, C	„module_151“	„translational elongation“, „translation“
A, B	„module_114“, „MORF_RAN“, „IVE_GENES“	„ribosomal large subunit biogenesis“
A, C	„GCM_NPM1“, „MORF_EIF3S6“,	„ribonucleoprotein complex assembly“,

		„electron transport chain“
A	„module_83“, „PUJANA_BRCA1_PCC_NETWORK“, „REACTOME_3_UTR_MEDIATED_TRANSLATION“, „AL_REGULATION“	„ribonucleoprotein complex biogenesis“, „RNA processing“,
C	„GCM_ACTG1“, „PECE_MAMMARY_STEM_CELL_DN“	
B	„GSE3982_DC_VS_NEUTROPHIL_LPS_STIMULATION“, „VS_CLASSIC_DN“, „MORF_JUND“, „STRUCTURAL_CONSTITUENT_OF_RIBOSOME“, „KEGG_RIBOSOME“, „MORF_CCNI“	

Tabelle 10: Vergleich der Anreicherungsanalysen der Asthmadaten

Die GSEA-Analyse zeigt unterschiedliche Ergebnisse der drei Datensätze. Datensatz B zeigt eine ähnliche Kategorie („KEGG_RIBOSOME“) wie Datensatz A („REACTOME_3_UTR_MEDIATED_TRANSLATION“) der DAVID-Analyse.

Die Tabelle 10 zeigt, dass sich bei der DAVID-Analyse die Ergebnisse der drei Datensätze nicht stark unterscheiden. Alle Kategorien sind mit Translation und dem Ribosom assoziiert. So gibt es keine Kategorien, die nur bei der Analyse der B- und C- Datensätze vorkommen. Es macht also keinen Unterschied, ob man die Gene des gesamten Datensatzes A in die Gruppen B und C aufteilt oder nicht.

4.7 Biologische Relevanz der Ergebnisse der Infektionsdatensatzanalyse

Relevante Interaktionskovariaten wurden als Teil des Modells behalten sowie deren Effekte qualitativ interpretiert und dabei in neutral, alleviating oder aggravating unterteilt. In weiterführenden Analysen wurden die Gene basierend auf ihren Effektgrößen hierarchisch geclustert. Die resultierenden acht Gencluster wurden mittels Signalweg- und TFBS-Analyse untersucht (6.1). Dabei stellte sich heraus, dass der Cluster 6 aus putativen C2TA Zielgenen besteht, welche von MHC-Klasse-II-Genen ko-reguliert werden (Abbildung 11).

Es ist seit langem bekannt, dass die Regulation von MHC-Klasse-II Expression fast ausschließlich von der Bindung des transkriptionellen Transaktivators C2TA abhängig ist, der an einen konstitutiven, aber inaktiven, Promotor-Komplex bestehend aus RFX-AP, -ANK, -5, CREB und NF-Y (Krawczyk and Reith 2006) bindet. Daher würde man erwarten, C2TA in demselben Cluster 6 wie die MHC-Klasse-II-Gene vorzufinden. Jedoch befindet sich C2TA in den Clustern 2 und 4, welche beide durch IFN- γ -Stimulation hochreguliert sind und einen aggravating Interaktionseffekt zeigen. Bei genauerer Betrachtung der Daten stellt man allerdings fest, dass

die Expressionsveränderung zwischen den beiden Mausstämmen nur sehr gering ist. Da beide Probe-Sets von C2TA auf demselben Exon basieren, kann dieses Ergebnis nur durch Rauschen erklärt werden. Es wird daher angenommen, dass die C2TA Expression zwischen beiden Mausstämmen unverändert ist, aber ein starker Effekt durch IFN- γ -Stimulation hervorgerufen wird. Dieser Effekt ist stärker in C57BL/6-Mäusen als in BALB/c-Mäusen. Der Effekt der Hochregulation von C2TA durch IFN- γ -Stimulation ist in BALB/c-Mäusen stärker. Der Effekt der C2TA-Hochregulation durch IFN- γ -Stimulation spiegelt sich auch im Anstieg der Expression der klassischen (und funktionellen) MHC-Klasse-II-Genen wider. Dies entspricht der Erwartung; es ist jedoch interessant zu beobachten, dass der Interaktionseffekt als alleviating berechnet wird. Es wird deshalb postuliert, dass die IFN- γ - und dadurch auch die C2TA-bedingte Hochregulation der MHC-Klasse-II-Gene an diesem Locus ansteigen. Daher kann der Expressionsunterschied zwischen den beiden Stämmen im Grundzustand nicht durch einen Effekt der Präsenz von IFN- γ erklärt werden.

4.8 Biologische Interpretation der Ergebnisse der linearen Regressionsanalyse der Asthmadaten

Durch die Analyse der Asthmadaten wurden in Kooperation (Rädler 2014) neue differentiell regulierte Signalwege und Gene als Biomarker für allergisches und nicht-allergisches Kindheitsasthma in PBMCs und isolierten CD4⁺-Zellen durch Microarrays identifiziert.

Die Identifizierung von spezifischen Biomarkern und zugrundeliegenden immunologischen Mechanismen in allergischem Asthma und nicht-allergischem Asthma bietet die Möglichkeit einer besseren Diagnose und einer spezifischeren Therapie. Bei Erwachsenen geben die Therapien der verschiedenen Asthma-Phänotypen unterschiedliche Antworten (Wenzel 2006). In der vorliegenden Studie werden Biomarker und Gennetzwerke untersucht, um allergisches und nicht-allergisches Asthma bei Kindern durch Transkriptom-Daten immunologisch besser zu charakterisieren. Die Gruppen AA, NAA und GK weisen 1523 differentiell regulierte Gene in PBMCs und eine unterschiedliche Regulation von mTOR-, TLR- und Kalzium-Signalwegs-Genen auf.

Zusammenfassung der vier wichtigsten Ergebnisse (Raedler, Prazeres da Costa, Ballenberger et al., submitted):

- i) AA sind durch unterschiedliche transkriptionelle Regulation des mTOR-Signalwegs charakterisiert sowie durch eine erhöhte Expression von TLR-assoziierten Genen.
- ii) NAA weisen differentielle Genregulation des TLR-Signalwegs und eine erhöhte Expression von Kalzium-Signalweggenen auf.
- iii) Th2-Genexpression ist in beiden Asthmaphänotypen erhöht im Vergleich zu GK.

- iv) Durch die Analyse der PBMCs können mehr differentiell regulierte Signalwege identifiziert werden als durch die Analyse der CD4⁺-Zellen.

In Kooperation (Raedler, Prazeres da Costa, Ballenberger et al., submitted) wurden zusätzlich zu PBMCs CD4⁺-Zellen untersucht, welche vermutlich eine wichtige Rolle bei der Entstehung von Asthma spielen (Robinson 2010). Die transkriptionelle Regulation ist weniger unterschiedlich in den CD4⁺-Zellen als in den PBMCs. Diese Ergebnisse unterstreichen die Komplexität von Asthma, welches nicht alleine auf einer Dysregulation von CD4⁺ basiert, sondern wahrscheinlich eher auf einem Zusammenspiel von mehreren Zelltypen innerhalb der PBMCs. Dies ist wichtig für die Auswahl der Proben für Microarray-Experimente. Jedoch müssen die Ergebnisse der CD4⁺-Zellen in weiteren Analysen mit größerer Gruppengröße repliziert werden, um die statistische Aussage zu verbessern.

In der vorliegenden Arbeit wurden die Daten durch eine zusätzliche Kovariate und eine Genweise Korrektur für einen möglichen Batch-Effekt korrigiert. Im Gegensatz zu vorherigen Microarray-Studien (Gruber, Coldren et al. 2006); (Lu, Pan et al. 2004) kann in den vorliegenden Daten keine altersassoziierte Expression festgestellt werden.

5 Zusammenfassung und Ausblick

Diese Arbeit stellt die zwei Methoden der Systembiologie an Hand von zwei Beispieldatensätzen dar. Diese Methoden der Systembiologie unterscheiden sich hinsichtlich der Anwendung von Vorwissen über die Struktur der Daten. So wird einerseits bei den top-down Methoden kein Vorwissen über die Struktur der Daten angewendet. Experimentelle Daten werden vielmehr anhand mathematischer Methoden untersucht, um biologische Mechanismen aufzudecken. In der vorliegenden Arbeit handelt es sich bei den mathematischen Methoden um lineare Regressionsmodelle, die auf Maus-Transkriptionsdaten angewendet werden. Die Ergebnisse der Regressionsmodelle werden mit einem neuen visuellen Hilfsmittel, dem Eruptionsdiagramm, verglichen.

Bei den bottom-up Methoden wird dagegen *a priori* Wissen über die Struktur von Daten verwendet. In dieser Arbeit wird beispielhaft ein Transkriptionsdatensatz von humanen PBMCs asthmatischer Kinder mittels einer bottom-up Methode analysiert.

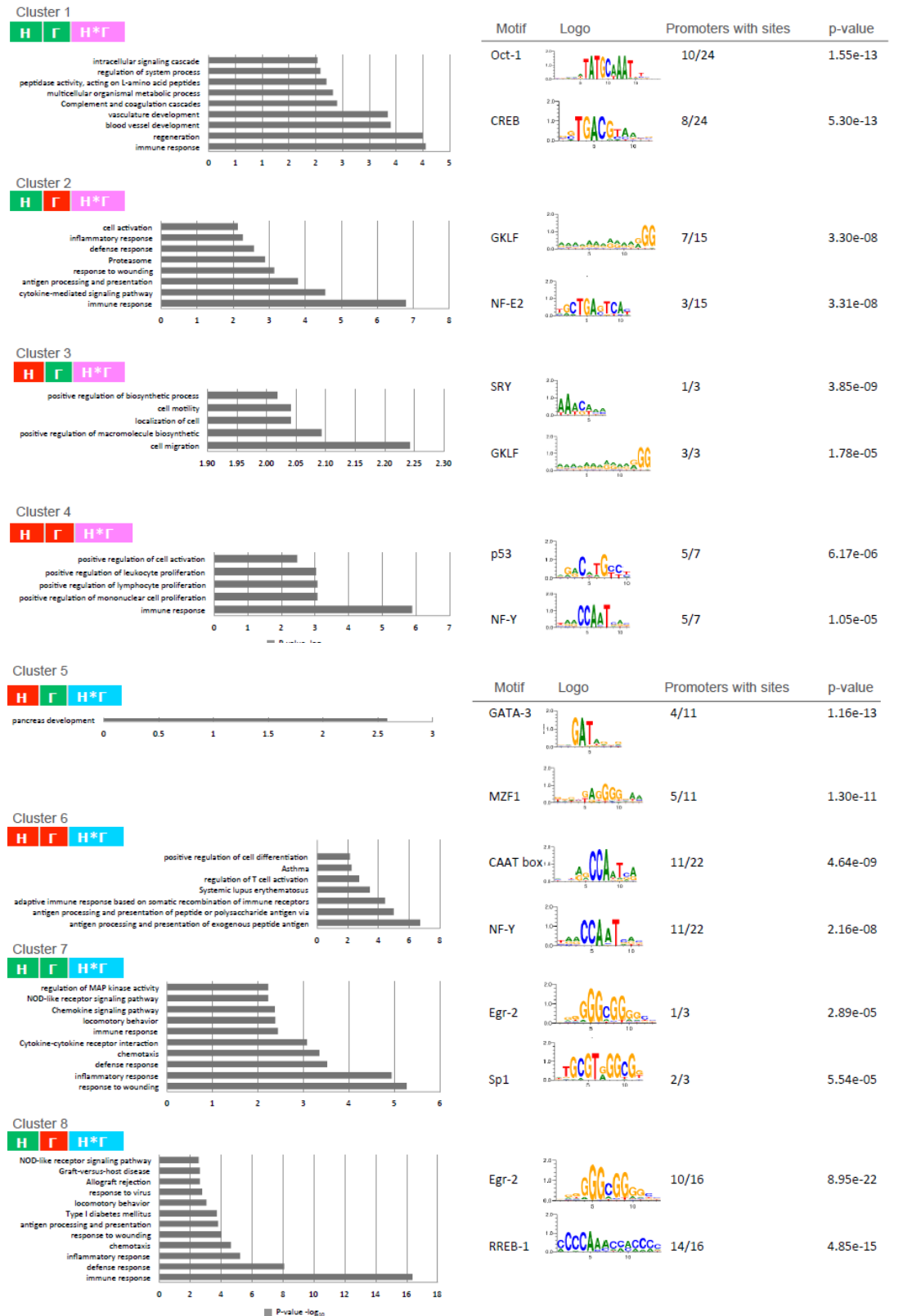
Anschließend werden geschachtelte lineare Regressionsmodelle auf Microarray Transkriptionsdaten angewendet, um Interaktionen zwischen mehreren Faktoren und deren Effekte auf die transkriptionelle Antwort zu untersuchen. Das erste Ziel dieser Arbeit ist die Modellselektion mittels des Eruptionsdiagramms. Modelle höherer Ordnung werden auf binäre Faktoren wie Mausstamm, Anwesenheit (oder Abwesenheit) von Zytokin-Stimulation oder bakterielle Infektion angewendet. Die linearen Modelle werden in einer hierarchischen Struktur angeordnet und durch eine rückwärts-gewandte Strategie verglichen, indem einzelne Kovariaten entfernt werden. Das Eruptionsdiagramm ist eine Ergänzung zu Tests wie GlobalAncova und wurde als neues Werkzeug entwickelt, um visuell das passende Modell für einen Datensatz zu identifizieren. Mithilfe dieser Methodologie werden verschiedene Kombinationen von Modellen höherer Ordnung getestet. So kann gezeigt werden, dass die Infektionskovariaten erster Ordnung zusätzliche erklärende Kraft gewinnen. Die Kovariaten dritter Ordnung werden dabei als irrelevante Kovariaten identifiziert und gestrichen. Effekte zweiter Ordnung werden in dem Datensatz nachgewiesen und das passende Modell wird basierend auf statistischer Signifikanz und Effektgröße bestimmt. Daher sollten Interaktionseffekte höherer Ordnung bei Anwendung von linearen Regressionsmodellen auf multi-Faktor-Datensätze in Betracht gezogen werden.

In der bottom-up Analyse der Asthmadaten werden allergische und nicht-allergische Phänotypen von Kindern untersucht. Die Interaktionen der Komponenten und Strukturen von

Gengruppen werden mittels des globalen Tests, GlobalAncova, analysiert. Die Gengruppen basieren auf bekannten Signalwegen und bestehen aus Gengruppen von T-Zelltypen. Die Gene der T-Zelltypen werden in dieser Arbeit vorgestellt. Der globale Test ist in GlobalAncova implementiert und wird auf die Gengruppen angewendet. Die Nullhypothese lautet, dass kein Gen der Gruppe differentiell exprimiert ist. Die Ergebnisse deuten auf unterschiedliche immunologische Mechanismen beider Phänotypen (nicht-allergischer Asthmatiker und allergischer Asthmatiker) hin. Die Behandlung von Asthma mit Medikamenten, die Kortikosteroide sowie kurze und lange acting beta-2 adrenergic agonists beinhalten, zielt auf Gene ab, die mit dem Kalziumsignalweg assoziiert sind. Der Kalziumsignalweg wird in dieser Arbeit als differentiell regulierter Signalweg beschrieben. Die Beschreibung neuer Genexpressionsprofile in dieser Arbeit könnte zu einer Patienten-spezifischen Behandlung führen.

6 Anhang

6.1 Abbildung der Funktionsanalyse der Cluster von Modell 3



6.2 Tabelle der alleviating und aggravating Interaktionen

Gen Name	H	Γ	$(H*\Gamma) - H:\Gamma$	$H:\Gamma$	Affymetrix IDs
Gbp1	-4,87572152	1,91549979	-12,9833884	3,64394489	1420549_at
H2-Ea-ps	-3,61577199	2,0491654	-5,36872005	-2,0405948	1422892_s_at
Ifi44	-3,55493422	0,85695199	-4,43228787	1,38587993	1423555_a_at
Cxcl11	-1,20329423	2,58014745	-3,30011349	0,19543694	1419697_at
Mpa2l	-0,53830527	4,62464393	-2,91244452	0,4229743	1438676_at
NA	-6,62761057	0,48421729	-2,77036857	-0,43883503	1425614_x_at
Plbd1	-1,10601782	1,72798901	-2,66243742	0,75125079	1448786_at
H28	-4,27440191	0,78815801	-2,56571178	-0,8031923	1421596_s_at
Ch25h	-1,36667521	1,60339133	-2,55697418	0,365659	1449227_at
Bst1	-0,45121003	1,19678642	-2,35563006	1,81562803	1449453_at
Qpct	0,69513578	-0,15785229	-2,26729381	2,15756503	1426622_a_at
H28	-4,17694311	0,62861275	-2,22630373	-0,39937596	1425917_at
Pdgfc	-0,26090494	-1,14770042	-2,11238719	2,4118279	1419123_a_at
NA	-5,14790844	0,46302959	-2,10199121	-0,28164274	1427651_x_at
Plagl1	1,18630888	-0,91798809	-2,06248936	0,97347193	1426208_x_at
Gatm	-1,72385966	0,63863731	-2,0540036	0,95308251	1423569_at
Cnn3	-0,46691924	0,61115221	-2,01985786	1,73449914	1436759_x_at
Upp1	0,04207036	0,82289918	-2,0026498	2,03726947	1448562_at
H2-T24	-1,42410839	1,62998419	-2,00257122	-0,31870294	1422160_at
Smpdl3b	-0,90349481	2,39945504	-1,8451132	-0,32278196	1417300_at
F13a1	0,65226876	-3,52997492	-1,8421563	-0,46033608	1448929_at
Cnn3	-0,32722605	0,62284899	-1,83883617	1,63502376	1455570_x_at
H2-D1	-5,51019025	0,37178676	-1,78923757	-0,25937823	1452544_x_at
Bst1	-0,05449385	0,71547783	-1,66265606	1,62366692	1449454_at
Pla2g16	-0,40380025	2,95602265	-1,58781184	0,39416915	1451611_at
Irf7	-0,81920263	1,99829926	-1,54890493	-0,08810707	1417244_a_at
Sox4	0,95017822	-1,69846131	-1,48856831	-0,12527264	1433575_at
NA	-0,39299866	3,01559665	-1,47454242	0,28941699	1449009_at
H2-K1	-2,070014	1,08144431	-1,45302519	-0,78557968	1451593_at
Itga6	1,0702611	-1,6354771	-1,32585345	-0,42453408	1422444_at
Niacr1	-1,07621058	0,86656194	-1,32127258	0,38866944	1419721_at
Ptgs2	-0,70397223	1,02025186	-1,31803253	0,59980355	1417263_at
H2-DMA	-0,19823428	2,04837762	-1,25319107	0,8471324	1422527_at
Hgsnat	1,89046936	-0,83153352	-1,13459431	-0,43739433	1450868_at
Ptgs2	-0,71280466	0,77098239	-1,13334265	0,5837828	1417262_at
Itga6	0,56225985	-1,94827656	-1,11890489	0,0234672	1422445_at
Igtp	-0,35906535	1,89973698	-1,04562164	0,36349192	1417141_at
Dnase1l3	0,09351537	1,77914379	-1,04101492	1,20739221	1421057_at
NA	1,69687006	0,06367184	-1,01589189	1,12393473	1455869_at
Stat1	-0,32070206	2,41437579	-0,96111811	0,18682281	1450033_a_at

Serpina3g	-0,20523435	5,12136033	-0,92706443	-0,12401463	1424923_at
Hpgd	1,82544075	-1,33957535	-0,92127672	-1,52403871	1419905_s_at
Osgin2	-0,55434174	0,62590693	-0,88326519	0,53629886	1424290_at
Ciita	-0,06569301	1,53507957	-0,88042093	0,77957693	1421210_at
Socs1	-0,15555431	1,36502128	-0,87431843	0,66198349	1450446_a_at
Oasl1	-0,61066634	0,95683173	-0,83005281	0,24574788	1424339_at
Procr	0,67883905	1,03770663	-0,79720298	1,50163876	1420664_s_at
Tmem87a	-2,01962446	0,47620133	-0,79479118	-0,16695666	1424454_at
NA	4,34295527	-0,03930174	-0,77696044	0,60627473	1452426_x_at
Psmb9	-0,32174128	2,01220461	-0,73834736	0,09093807	1450696_at
Pla2g4a	-0,13470674	1,77113448	-0,72454417	0,48596043	1448558_a_at
Slc7a2	-0,58402282	0,64445658	-0,71030838	0,33393103	1426008_a_at
Gja1	-0,09811079	-2,30113141	-0,69497433	0,92074015	1437992_x_at
Il12rb1	-0,04849361	1,81311386	-0,68719527	0,59927083	1418166_at
Slc44a1	0,26587287	-1,50972528	-0,68066304	0,27926804	1433645_at
Rbl1	-0,30281221	1,6576516	-0,6775418	0,17558466	1424156_at
Tmsb10	-0,44951822	1,56165283	-0,67454218	-0,02744921	1436902_x_at
Mmp12	0,39714805	-2,10488175	-0,66926684	-0,16668284	1449153_at
Tmsb10	-0,40137704	1,60876503	-0,66170879	0,01598744	1417219_s_at
Il10	0,33897528	-1,04224073	-0,64804917	0,29475533	1450330_at
Ctgf	-0,1819577	-2,15231537	-0,64678981	1,03842016	1416953_at
Clcf1	-0,11901834	-0,24842524	-0,64581631	0,67538347	1437271_at
Gtpbp2	-0,39241753	1,69922366	-0,63364023	-0,03316492	1416691_at
Ptx3	-0,57898546	-0,90394869	-0,62102411	1,14439725	1418666_at
S100a9	2,1137935	-0,24952094	-0,59082971	0,06339396	1448756_at
A630072M18Rik	0,15593098	-0,80917734	-0,57644015	0,45026434	1427081_at
Mthfd2	-0,2729049	1,67964784	-0,57068308	0,11229896	1419254_at
Gdf15	1,801129	-0,48030689	-0,55210156	-0,3129931	1418949_at
Fabp4	0,25887687	-1,87023578	-0,54935338	0,06519259	1417023_a_at
Il6	0,1432938	-0,89193554	-0,54856224	0,4207534	1450297_at
Zbp1	-0,20680432	2,30311852	-0,54321813	0,06692327	1429947_a_at
Psmb10	-0,26038981	1,67170157	-0,53541522	0,10012116	1448632_at
Gja1	0,09913855	-1,6270082	-0,53288147	0,37158224	1438945_x_at
NA	0,33057656	-0,18034519	-0,52796124	0,46834335	1435137_s_at
S100a8	2,22376192	-0,19230404	-0,52197665	0,09433825	1419394_s_at
Frmd6	0,43081944	-1,11669274	-0,49845336	0,01736042	1451264_at
Ass1	-0,33753894	3,02648171	-0,48797009	-0,53358533	1416239_at
Osgin2	-0,36692149	0,66548732	-0,47992303	0,23574142	1424289_at
S100a4	0,02233544	-1,59492213	-0,47987556	0,44425228	1424542_at
Gja1	0,11460281	-1,58858766	-0,47350013	0,29144352	1415800_at
Mt1	0,23402359	-1,59120249	-0,46935177	0,09697286	1422557_s_at
Tap2	-0,23763218	1,99668271	-0,46202428	-0,01245178	1453913_a_at
Wwp1	0,35845871	-1,92406959	-0,44949427	-0,24020523	1452299_at
Gbp2	0,01976729	2,01686941	-0,44868977	0,48855782	1418240_at

Cxcl11	-0,53585169	1,99956104	-0,44126846	-0,6301997	1419698_at
Psme2	-0,20509541	1,53264399	-0,43945351	0,12511526	1417189_at
Cap1	-2,03361082	0,24162404	-0,43168939	-0,05967986	1417462_at
NA	0,47245509	-0,18540943	-0,42125924	0,33366161	1427932_s_at
Arid5a	-0,07933363	0,11435325	-0,42007336	0,4110013	1451340_at
Wars	-0,04998861	1,84154847	-0,4172083	0,32515185	1415694_at
Tnfsf14	0,00798809	-0,97842699	-0,41671016	0,4088944	1450298_at
Lif	-0,3547003	-0,70157587	-0,41457525	0,66342443	1421207_at
Pim1	-0,98197428	0,2128129	-0,4126586	0,2036818	1423006_at
Mtmr7	-0,28962448	0,11952851	-0,4074508	0,37283242	1427025_at
Ifi202b	-3,7096288	0,75068105	-0,40726567	-2,37748239	1421551_s_at
Cpeb4	0,2704146	-0,10350714	-0,39962638	0,37163654	1449931_at
Slc40a1	0,40189963	-1,70876143	-0,39764095	-0,28910963	1417061_at
BC006779	-0,06053125	0,69111946	-0,39416089	0,35232657	1435454_a_at
Bak1	-0,20936998	1,5897383	-0,39160898	0,0587655	1425716_s_at
Hist1h2bc	1,51501804	-0,5989044	-0,39089079	-0,51646019	1418072_at
Obfc2a	-0,35093452	0,61357939	-0,38985609	0,17452991	1452203_at
Stat1	-0,14009242	2,08030017	-0,38933716	0,09790287	1450034_at
Ank	0,0259561	-1,65163733	-0,37515547	0,3322854	1450627_at
Socs3	-0,23028002	-0,02306584	-0,37272215	0,37803376	1416576_at
Klf6	-0,26556867	0,39129988	-0,37224511	0,26832812	1418280_at
Cited2	0,05697289	-0,50911197	-0,36679266	0,33778707	1452207_at
5430435G22Rik	1,755495	-0,2124661	-0,36036279	-0,0126204	1424987_at
Cdk5r1	0,52185372	-0,06065956	-0,35884473	0,32718931	1433450_at
Arl2bp	2,32281094	-0,11453178	-0,35723532	0,09119965	1429859_a_at
NA	-0,10409268	-0,77610622	-0,35531019	0,43609716	1419858_at
Pim1	-1,13370062	0,01463565	-0,35341228	0,33681984	1435458_at
Trim12a	1,56622835	0,06371817	-0,35173447	0,45153168	1437432_a_at
Slfn1	-0,20409983	2,36609042	-0,35020208	-0,13271657	1418612_at
Tnfsf9	-0,09432313	-0,77590489	-0,34695923	0,42014501	1422924_at
NA	0,28573297	-0,15211212	-0,34550151	0,30203806	1453238_s_at
Pf4	0,22207691	-2,14955685	-0,34348454	-0,1338824	1448995_at
Psmb8	-0,13720782	1,82297591	-0,34260915	0,09248259	1422962_a_at
Pim3	-0,8448627	-0,1295487	-0,34012368	0,44957455	1451069_at
Kcnj10	-1,75145002	0,2035347	-0,33754727	-0,01893358	1419601_at
Igf1	0,53008497	-1,68491965	-0,32459583	-0,56855475	1419519_at
2310016C08Rik	-0,06246771	-0,46995101	-0,32280521	0,35216198	1421031_a_at
Rcan1	0,01830614	-1,17501076	-0,32188841	0,30037849	1416600_a_at
Tagln	-0,15505316	-1,55498523	-0,32037222	0,56147759	1423505_at
Tiparp	-0,2031058	-0,64573645	-0,31994532	0,45109814	1452160_at
Wars	-0,04593873	1,60691747	-0,31406166	0,24024192	1434813_x_at
Nfil3	-0,16600696	0,07002352	-0,30310435	0,29147996	1418932_at
Ifi203	-0,22298601	0,00365251	-0,30062368	0,29980923	1451567_a_at
Rcan1	0,11749045	-1,03735919	-0,28962718	0,16774738	1416601_a_at

Tspo	-0,11966126	1,5176677	-0,28891031	0,10730429	1438948_x_at
F3	0,07321258	-0,04388846	-0,288531	0,28531781	1417408_at
NA	-0,13489184	1,55176318	-0,27495772	0,06563754	1449556_at
Icosl	-0,35944668	0,41454707	-0,27469564	0,12568807	1419212_at
Wars	-0,02247037	1,93982459	-0,2714945	0,22790592	1425106_a_at
Obfc2a	-0,11001934	0,96147448	-0,26445761	0,15867682	1460521_a_at
Slco3a1	0,09362119	1,82565743	-0,26174715	0,43266737	1418030_at
Gm4586	-0,73048012	-0,11628992	-0,26110013	0,34604761	1452418_at
Tnfsf14	-0,03146748	-0,5306063	-0,25827246	0,27496931	1421588_at
Gbp8	0,08765617	4,13821541	-0,25681855	0,61955865	1418776_at
Gadd45b	-0,23927901	-0,18136108	-0,25197379	0,29536969	1449773_s_at
Dusp1	0,38366424	-0,42998012	-0,23930363	0,07433563	1448830_at
Stat1	-0,13830303	1,86771975	-0,23770226	-0,02060905	1420915_at
Il1rn	-0,10901021	0,03643976	-0,23640608	0,23243377	1425663_at
P2ry14	-0,19949375	1,9457736	-0,23405947	-0,15411019	1424733_at
Cited2	0,1164321	-0,34263368	-0,22841318	0,18851962	1421267_a_at
Fpr2	-0,51968554	2,85607661	-0,22707409	-1,25718763	1422953_at
Ccrl2	-0,93906878	-0,03712477	-0,22247116	0,25733387	1427736_a_at
Wwp1	0,31906226	-1,67997866	-0,21939776	-0,31662003	1427098_at
Slc7a2	-0,29441297	0,62064261	-0,21521399	0,03248875	1422648_at
Socs3	-0,24219574	-0,10869997	-0,20699409	0,23332076	1456212_x_at
Gstm1	0,38019889	-1,57247401	-0,20546121	-0,39239166	1416416_x_at
Mir155	-0,33092081	0,31970718	-0,20119108	0,09539333	1428027_at
Ccrn4l	0,27540805	-0,92918735	-0,19922562	-0,05668005	1425837_a_at
Il1rn	-0,33615245	0,34509629	-0,19224537	0,07624041	1451798_at
Tapbpl	-0,1063757	1,94081466	-0,19049756	-0,01595796	1451544_at
Traf1	-0,22540029	1,11960735	-0,18808897	-0,06427085	1423602_at
Gbp2	0,06127962	1,94957168	-0,17800364	0,29747265	1435906_x_at
Id3	-0,39783349	-0,27684261	-0,17596459	0,28610185	1416630_at
Tnf	-0,20464489	0,30505496	-0,16674338	0,10431544	1419607_at
Acta2	-0,19377061	-1,58707862	-0,16019736	0,46772655	1416454_s_at
Ncapg2	3,45896387	-0,10914658	-0,1530454	-0,22448867	1417926_at
ligp1	-0,01296044	3,31435481	-0,15304071	0,11008521	1419042_at
Igf1	0,48225602	-1,61168051	-0,15251753	-0,62472509	1452014_a_at
Tbc1d14	-0,08299963	0,47811971	-0,15045307	0,11076931	1420196_s_at
Slco3a1	0,29823897	2,08850401	-0,14769703	0,77057031	1434537_at
Plaur	-0,01981603	-0,73544242	-0,14496337	0,15953692	1452521_a_at
Tgs1	-0,56408995	1,50507742	-0,1445646	-0,70443446	1450401_at
Ifi47	-0,05869135	1,5661984	-0,14179426	0,04987196	1417292_at
Sfi1	-2,0217219	-0,07123735	-0,13703126	0,28105337	1426787_at
Nt5c3	-0,20282557	0,55415151	-0,13343446	0,02103836	1451050_at
Ppap2b	-0,49517952	-0,7337036	-0,13295356	0,49626856	1429514_at
C3	-0,19420922	2,61775383	-0,1292924	-0,37909952	1423954_at
Rgs1	-0,24289542	0,55570906	-0,125735	-0,00924419	1417601_at

Rabgef1	0,07219952	-0,17430966	-0,12378617	0,1112011	1419068_at
1110002B05Rik	-0,13780766	-1,547224	-0,12233095	0,33555026	1448388_a_at
Gadd45b	-0,13023241	0,04250477	-0,12230433	0,11676884	1450971_at
Phlda1	-0,35023411	-0,75542621	-0,12020298	0,38477901	1418835_at
Tcf4	-0,2358779	-1,57081598	-0,11821248	0,48873325	1434148_at
Cnr2	-0,10421192	0,0632738	-0,11458079	0,10798691	1450476_at
Tiparp	-0,34144105	-0,46024585	-0,11420332	0,27135015	1452161_at
Hist1h1c	0,17750809	-1,6104996	-0,10898283	-0,17689388	1416101_a_at
Slco3a1	0,15992303	1,61291002	-0,10210063	0,36004208	1448918_at
Egr2	-0,38349653	-0,31135328	-0,10184916	0,22125206	1427682_a_at
lfrd1	-0,30824343	-0,07657585	-0,10150041	0,12510442	1416067_at
Zbp1	-0,11986958	2,29552503	-0,09895196	-0,17621166	1419604_at
Chi3l3	2,44727602	-0,17376427	-0,09887283	-0,32637629	1419764_at
Pdxdc1	1,78457423	-0,15357287	-0,09419722	-0,17986497	1452705_at
Cap1	-1,94110646	0,14480056	-0,0920138	-0,18905951	1417461_at
Mmp13	-0,29846775	-0,35250604	-0,08940627	0,19461796	1417256_at
Arl2bp	1,93505669	-0,13436977	-0,08873708	-0,17127605	1417112_at
Cmpk2	-0,44765635	0,51027915	-0,08813595	-0,14029375	1422095_a_at
Maff	-0,13972222	-0,24820104	-0,08774164	0,12242084	1418936_at
Cmpk2	-0,47816589	0,14528093	-0,0875875	0,01811911	1450484_a_at
Ccl2	-0,40901158	0,72708428	-0,0875243	-0,2098616	1420380_at
Il12b	-0,06929672	1,30737883	-0,0844726	-0,00612447	1449497_at
Bhlhe40	-0,14375518	-0,44934532	-0,08151763	0,14611334	1418025_at
Ciita	0,20181098	1,51203668	-0,07922995	0,38437556	1421211_a_at
Samhd1	0,00535748	1,69423918	-0,0771568	0,08623366	1434438_at
Atf3	-0,09847449	0,17859362	-0,07660679	0,05901988	1449363_at
H2-Q7	-0,00785243	1,64091919	-0,07639853	0,06351332	1418536_at
Peli1	-0,17124836	0,16055429	-0,0750392	0,04754454	1417372_a_at
Tiparp	-0,53407896	-0,72021917	-0,07477256	0,45942647	1426721_s_at
AW112010	-0,09555992	2,69204094	-0,07413687	-0,18311435	1434372_at
Mx2	-0,05850649	-0,48035331	-0,06395781	0,0920616	1419676_at
Ifnb1	0,03816433	-0,39619898	-0,06371753	0,04859686	1422305_at
Rabgef1	0,10401694	-0,07884156	-0,05601032	0,04780946	1419069_at
Rabgef1	0,08740529	-0,12873177	-0,05521198	0,04396014	1419067_a_at
Tnfaip3	-0,06730283	-0,02225581	-0,05186299	0,05336087	1450829_at
Rel	0,10625887	-0,14494703	-0,04400785	0,02860594	1420710_at
Peli1	-0,16637934	0,11457393	-0,03509777	0,01603503	1417371_at
Zfp187	-0,51435312	-0,3038198	-0,03188216	0,18815282	1457285_at
NA	1,77654653	-0,09187097	-0,02771672	-0,13549633	1425451_s_at
Aplp2	0,06758068	-1,54220531	-0,02635696	-0,07786632	1421889_a_at
Il15	0,04963649	0,70712215	-0,02596544	0,0610645	1418219_at
Klf7	-0,38895593	-0,20249734	-0,02428588	0,10304842	1419356_at
Ifi203	0,77902093	0,01317869	-0,02416728	0,03443376	1448775_at
Tspo	-0,04788743	1,56631917	-0,02305021	-0,05195679	1456251_x_at

Il1rn	-0,17096768	0,41061001	-0,02042702	-0,04977402	1423017_a_at
Rps9	2,21195162	-0,17839445	-0,01955677	-0,37504313	1434624_x_at
Slc7a11	0,11123817	0,2579455	-0,0180323	0,04672569	1420413_at
Ifit2	-0,06477399	0,8388647	-0,01405789	-0,04027872	1418293_at
Pcdh7	-0,35694497	-0,21106407	-0,01384785	0,0891861	1449249_at
Sesn1	-0,06498709	-0,95376569	-0,01013371	0,07211618	1438931_s_at
Egr1	-0,06845342	0,32929811	-0,00952366	-0,01301793	1417065_at
Cald1	-0,56611308	-1,6587759	-0,00872069	0,94777542	1424769_s_at
NA	3,73508677	0,29517299	-0,00703788	1,10953462	1431213_a_at
Olr1	0,05139031	-0,5499388	-0,00618111	-0,02208042	1419534_at
Hal	-2,89126085	0,0038582	-0,00172816	-0,00942689	1418645_at
Fem1b	-0,0126059	-0,55490082	0,00069134	0,00630369	1418324_at
Mdm2	-0,02072352	0,41728047	0,00170488	-0,0103524	1427718_a_at
Smox	-0,42638879	-0,5994473	0,0053213	0,25027631	1424268_at
Txnip	-0,10585085	0,52435679	0,01652869	-0,07203231	1415997_at
BC013712	0,0115928	1,86960385	0,0170116	0,00466234	1427041_at
Plxnc1	0,0074637	-1,72003095	0,0249361	-0,0377739	1423213_at
NA	-0,38844128	2,58033808	0,03558469	-1,03789451	1452408_at
Klf9	-0,41582832	-1,72696472	0,03594702	0,68217383	1428288_at
Egr2	-0,41458915	-0,48756755	0,04122398	0,16091623	1427683_at
Nfkbiz	-0,40872067	0,00934055	0,04341656	-0,04723423	1448728_a_at
Dnajb4	-0,0870078	-0,31953368	0,04460214	-0,01680021	1451177_at
Ifih1	-0,14721916	0,24567289	0,04563557	-0,08180332	1426276_at
Txnip	-0,01092376	0,58918769	0,051682	-0,05811814	1415996_at
Ptpla	-0,48225337	-1,50549088	0,0552996	0,67072846	1456315_a_at
Hist1h1c	0,08010313	-1,79621877	0,05583979	-0,19972255	1436994_a_at
Plxnc1	-0,07229177	-1,53192869	0,05809818	0,05264766	1450905_at
Myl9	-0,34112495	-1,76609076	0,06046271	0,54199492	1452670_at
Errfi1	-0,41831029	-1,19812925	0,06500714	0,43618265	1419816_s_at
Dnajb4	0,04992493	-0,18995741	0,06525814	-0,07474175	1431734_a_at
Fosb	-0,51765091	-0,45119754	0,07003054	0,16353228	1422134_at
Ngp	1,82734099	-0,0984711	0,07250303	-0,2524433	1418722_at
Sesn1	-0,17678659	-1,19379416	0,07786807	0,13317873	1454699_at
Ppp1r15a	-0,04036346	-0,03821307	0,0786673	-0,07712489	1448325_at
Ccl4	-0,18210089	-0,05472924	0,07989354	-0,0699273	1421578_at
Ifit1	-0,09814934	0,36736919	0,08387715	-0,11993419	1450783_at
Ctla2a	-0,07436403	-1,83889635	0,08590859	0,05083916	1448471_a_at
ligp1	0,01794775	3,74564532	0,08914234	-0,02191645	1419043_a_at
Kctd12	-0,03345324	-1,69661536	0,09384716	-0,03708987	1434881_s_at
Edn1	-0,29435065	0,13796638	0,097142	-0,1377525	1451924_a_at
Pdxdc1	-2,16685178	-0,26146157	0,09762107	0,4689274	1419327_at
Trim25	-0,00984896	-0,22972193	0,10236302	-0,1001005	1425974_a_at
Gstm1	0,11008236	-1,6435955	0,10542924	-0,28636012	1448330_at
Mx1	-0,02395021	0,85283302	0,1062485	-0,12667403	1451905_a_at

Serpine1	-0,43379675	-1,72693434	0,10813038	0,64100812	1419149_at
Ii12b	0,29591209	0,97892192	0,11086038	0,17881445	1419530_at
Rgl1	0,25198801	0,42534827	0,11392019	-0,00673753	1418535_at
Batf2	0,0148469	2,54873073	0,11629847	-0,07845771	1430005_a_at
NA	-0,13880256	1,43861672	0,11847529	-0,31815898	1431591_s_at
Rreb1	-0,03666656	-0,4612699	0,11924388	-0,1023307	1428657_at
Sesn1	-0,06914987	-0,90264305	0,12096597	-0,05854831	1433711_s_at
Trim25	-0,08873188	-0,32701794	0,1243762	-0,09535928	1419879_s_at
Trim25	0,67458525	-0,04172419	0,12636826	-0,15451479	1426415_a_at
Dab2	0,01017873	-2,05613945	0,12681542	-0,14774431	1423805_at
Ifi203	0,92200614	0,16816307	0,12861642	0,02643096	1425008_a_at
Trim34a	1,54641306	-0,02425842	0,12947674	-0,16699028	1424857_a_at
Lcn2	1,68483334	0,00890711	0,14340839	-0,12840139	1427747_a_at
Ii1a	-0,89833905	0,08986352	0,14365733	-0,22438524	1421473_at
Akr1b10	0,12375793	0,09062806	0,14431914	-0,1331032	1425684_at
Fem1b	0,41500113	-0,00067992	0,15861507	-0,15889723	1418323_at
Rsad2	-0,39953738	0,47570356	0,16396772	-0,35402907	1436058_at
Casp8ap2	-0,02817731	0,06141102	0,16784409	-0,16957449	1449217_at
Uba7	-0,25022446	1,71520859	0,17311352	-0,60230066	1426971_at
Klf9	-0,52343459	-1,71464227	0,19394401	0,70355906	1456341_a_at
Arap3	0,10231611	-1,59943164	0,20286276	-0,36651039	1419833_s_at
Rnf2	0,20178559	-0,61268794	0,20379107	-0,32742267	1448042_s_at
1110038F14Rik	-0,00846079	1,62247719	0,20858658	-0,22231402	1416885_at
Ets2	-0,57808621	-0,26594114	0,21145735	-0,05772045	1416268_at
Zfp810	0,1330241	0,18518453	0,21212221	-0,18748821	1451566_at
Hhex	0,13842821	-0,26630016	0,21907425	-0,25593771	1423319_at
Sephs2	-0,29285299	0,13764297	0,22053359	-0,26084275	1418325_at
Samhd1	0,14921742	1,87378595	0,22203646	0,05756505	1418131_at
A830007P12Rik	-0,04390719	0,44119611	0,22245847	-0,24183016	1424829_at
Mfsd6	-0,28750393	-1,54292517	0,22262652	0,22097052	1424463_at
Slc15a2	1,61486757	-0,01823018	0,22533369	-0,25477302	1417600_at
Slc38a1	0,18565095	0,07787267	0,22947451	-0,21501738	1415903_at
Ccl3	-0,05464822	-0,25065536	0,23308595	-0,21938808	1419561_at
Cadm1	-1,53350423	-0,83286137	0,2413169	1,03587954	1431611_a_at
Ccl8	-0,10645074	2,2060997	0,25036058	-0,48520153	1419684_at
B3gnt8	-0,23164693	0,53621952	0,26558774	-0,38980134	1425128_at
Timp2	-0,1008113	-1,60349858	0,2688055	-0,10715472	1433662_s_at
Plk2	0,22188352	-0,19772916	0,26895701	-0,31282985	1427005_at
Errfi1	-0,40285172	-1,41682493	0,27749858	0,29327178	1416129_at
NA	-0,06543178	1,62637994	0,27865461	-0,38507155	1449875_s_at
Fkbp9	-0,47690908	-1,77177619	0,28046515	0,564511	1437687_x_at
Usp18	-0,05153848	1,4507772	0,28765271	-0,36242356	1418191_at
NA	0,1549137	2,13590741	0,29713177	0,03374956	1416016_at
Klf6	0,03664586	0,59743352	0,30274006	-0,2808466	1427742_a_at

Ngrn	-0,88868972	-0,08467996	0,30374075	-0,22848654	1416709_a_at
Cd40	0,24347262	1,57093235	0,30510977	0,07736925	1460415_a_at
Igf1	0,10291316	-2,115879	0,32777682	-0,54552862	1437401_at
Anxa3	-0,2025495	-1,8187206	0,33384232	0,03453863	1460330_at
Lmo4	0,21374063	1,53695092	0,36000602	-0,03149716	1420981_a_at
Cd274	0,3079228	1,15017883	0,36121782	-0,00705154	1419714_at
Dck	0,03677442	1,60301164	0,37379751	-0,31484768	1439012_a_at
Rsad2	-0,25351171	0,79233122	0,37743032	-0,57829556	1421008_at
Rsad2	-0,18473921	0,54146618	0,39967564	-0,49970568	1421009_at
Rnd3	0,48285534	0,57184628	0,40070653	-0,1245875	1416700_at
Rgl1	0,47366263	0,4458161	0,40357487	-0,19240844	1449124_at
Igfbp4	-0,15054123	-1,63478465	0,40538742	-0,15928492	1421991_a_at
Cenpa	-0,43812358	-1,55685604	0,40627883	0,27581651	1450842_a_at
Gem	-0,70951452	-0,56498163	0,41054471	-0,00968204	1426063_a_at
Sema4a	0,16849169	1,85419274	0,41671708	-0,104301	1448110_at
Cadm1	-1,69543405	-0,80730292	0,42703329	0,94169557	1417377_at
Bmp2k	-1,93409776	-0,38038817	0,42722927	0,30847865	1421103_at
Dab2	-0,08386873	-1,66405337	0,43070457	-0,29114252	1420498_a_at
Cd81	-0,1124566	-1,72184548	0,43212879	-0,2384959	1416330_at
Ccl12	-0,10784343	2,09584205	0,45833275	-0,68435556	1419282_at
St3gal5	0,21961702	1,75466714	0,46003895	-0,07468418	1460241_a_at
Fcgr1	0,04626543	1,74033148	0,46743926	-0,38692208	1417876_at
Sulf2	0,06901742	-1,91213917	0,47452683	-0,60649774	1430388_a_at
Sephs2	-0,31769623	0,3225633	0,4801631	-0,58264024	1449040_a_at
Rcbtb2	-0,45526622	-1,56607821	0,49142686	0,22155564	1416390_at
Timp2	-0,20548943	-2,00725072	0,51469611	-0,10222731	1454677_at
S1pr1	-0,40138783	-1,51639546	0,52701123	0,08165145	1423571_at
H2-Ea-ps	-0,66203472	1,68065262	0,532436	-1,64508639	1422891_at
Igfbp4	-0,21787824	-1,78718939	0,53850721	-0,14911753	1423757_x_at
Rpgrip1	3,201191	0,05987887	0,54641898	-0,35473527	1421144_at
Rnd3	0,58755482	0,61733681	0,55068551	-0,18796629	1416701_at
Ms4a4c	-0,21810078	2,98174751	0,5541821	-1,20450357	1420671_x_at
Cflar	0,20288218	0,98502884	0,55651079	-0,35666599	1449317_at
Cxcl2	-0,34519834	-0,79969227	0,57391524	-0,2978628	1449984_at
Gm9706	0,14334668	1,01097116	0,59915968	-0,45424032	1453939_x_at
Ifit3	0,29068019	0,47956924	0,60205621	-0,46265493	1449025_at
Il2rg	0,17120868	1,81228075	0,61209246	-0,30181427	1416296_at
Emp1	-0,27847623	-2,07758553	0,6168789	-0,03832072	1416529_at
NA	-0,28002284	1,71892749	0,6223059	-1,10364486	1435792_at
Igfbp4	-0,1347113	-1,64811693	0,63307528	-0,4110553	1437405_a_at
Idh1	-0,23374648	-1,61312313	0,6332367	-0,25617484	1419821_s_at
Zdhhc14	-0,19498463	-1,56405395	0,63683072	-0,33186424	1437614_x_at
Zdhhc14	-0,29058818	-1,58771957	0,68608363	-0,22471109	1438619_x_at
Cflar	0,17671689	0,87549362	0,70042045	-0,54570594	1425686_at

Igfbp4	-0,26772956	-1,95689404	0,72622511	-0,20230674	1421992_a_at
Il1b	-1,31204127	-0,30496843	0,73677963	-0,33664847	1449399_a_at
H2-Aa	0,24533154	2,14625349	0,74891727	-0,22237359	1452431_s_at
Cd28	-0,54542427	-1,922823	0,75120479	0,29754955	1417597_at
Ccl7	-0,20793531	0,0725984	0,78519841	-0,80029418	1421228_at
Aif1	0,17832687	1,96285318	0,79575605	-0,44572659	1418204_s_at
Cd5l	3,06512134	-0,28080296	0,84977985	-1,71047499	1449193_at
Cd28	-0,54059587	-2,12285152	0,85621189	0,29139288	1437025_at
Ms4a4c	-0,14500039	3,50558346	0,8656921	-1,37400308	1450291_s_at
NA	0,69953556	2,16693988	0,86732667	0,64852484	1418638_at
NA	2,60591982	0,29070187	0,86966558	-0,1121198	1435872_at
Lox	-1,04113979	-2,13327982	0,92433616	1,29670633	1448228_at
Plac8	0,36331977	1,98412258	0,9377482	-0,21687725	1451335_at
Serpinb2	-1,19004642	-1,24230623	0,9387158	0,53968628	1419082_at
Sephs2	0,23558805	0,53003629	0,94430552	-0,8194353	1435602_at
Serpinb9	0,55206601	1,995979	0,94651452	0,15539765	1422601_at
Igfbp4	-0,2211647	-2,29351116	0,9648472	-0,4576035	1423756_s_at
Cst7	0,04269178	2,91782785	0,9937364	-0,86916914	1419202_at
Lox	-1,06597959	-2,06579571	1,01368031	1,18841575	1416121_at
Mmp8	-0,94328486	-1,79719251	1,01402665	0,68123783	1449366_at
Ctse	-2,08148256	-0,77595654	1,02676412	0,58837589	1418989_at
Fcrls	-0,01607431	-1,83603934	1,06681785	-1,03730479	1448891_at
Clec4n	-0,69954514	-1,66682592	1,08359948	0,08242048	1419627_s_at
Gas6	-0,426594	-2,77120588	1,10090781	0,08127198	1417399_at
Klrk1	0,34684945	2,05550131	1,12936859	-0,41641909	1450495_a_at
Ly6i	0,15137333	1,93656558	1,14775625	-0,85461188	1425890_at
Bst2	0,1057137	2,06480742	1,16098143	-0,94270299	1424921_at
Clec4n	-0,717175	-2,045449	1,22285056	0,24409433	1425951_a_at
NA	1,6488687	0,4181065	1,22562165	-0,53621893	1438398_at
H2-Ab1	0,4727679	2,28973421	1,228859	-0,14634617	1425477_x_at
Acsl1	0,35632808	2,40881359	1,23442103	-0,37609311	1450643_s_at
Acsl1	0,42728964	2,04593497	1,28524201	-0,4110352	1423883_at
Mrc1	-0,28729847	-2,75708069	1,31814159	-0,52603651	1450430_at
Lair1	0,51371308	1,95237357	1,37003719	-0,36707735	1430447_a_at
Cp	0,1288018	1,55710034	1,37602987	-1,17547255	1417495_x_at
Isg20	-0,10812176	1,71364297	1,42124025	-1,60652233	1419569_a_at
Acsl1	0,41803607	2,61667842	1,48617233	-0,39230637	1422526_at
Il18bp	0,25039955	2,93973359	1,49389838	-0,75779043	1450424_a_at
Cfb	0,0647717	3,92705945	1,50210024	-1,24773793	1417314_at
Fam20c	-0,79594862	-1,65079914	1,54834916	-0,23439786	1417688_at
Fcgr4	0,42095803	3,47426599	1,56720744	-0,10468729	1425225_at
NA	0,67428071	1,60590044	1,57593659	-0,4931089	1453299_a_at
NA	0,34591758	2,0952671	1,62123407	-0,89644434	1421571_a_at
H2-Ab1	0,8034623	1,93539814	1,66126193	-0,10624249	1450648_s_at

Cxcl1	-1,08464651	-1,55874743	1,79431176	-0,10362179	1419209_at
Fgl2	0,3886823	3,92625508	1,82462315	-0,2985573	1421854_at
Cp	0,21747225	1,83368591	1,90935264	-1,51057683	1455393_at
Cxcl9	0,37612836	4,39709635	1,91372819	-0,25985556	1418652_at
Arg2	-2,27213492	-1,17719788	1,93679078	0,73796163	1418847_at
Cd74	0,90495353	1,6719976	2,02491898	-0,51183886	1425519_a_at
Pyhin1	2,51091299	0,83848557	2,03361679	0,07174752	1435330_at
Plau	-3,46146127	-0,75754288	2,03946809	0,58273724	1422138_at
H2-Ab1	0,90706862	2,21517902	2,05628283	-0,04696346	1451721_a_at
F2r	-2,02175953	-1,36788573	2,05884575	0,70669025	1437308_s_at
H2-Eb1	0,9280893	2,03778862	2,1049672	-0,21371739	1417025_at
Epsti1	0,91444875	1,73460092	2,30622922	-0,72002558	1454169_a_at
Fgl2	0,53603139	3,56801528	2,32595828	-0,41339007	1421855_at
Slamf9	0,80214122	2,41915162	2,35579628	-0,41529504	1419315_at
Gdap10	0,72287554	1,92040181	2,37068979	-0,9824783	1420342_at
Lpl	1,58031434	0,93402036	2,42725462	-0,95120885	1431056_a_at
H2-Aa	1,09226462	1,79271403	2,44909549	-0,49097739	1435290_x_at
Colec12	-2,40648428	-1,68426674	2,4515748	1,60158661	1419693_at
Gab1	-1,23371834	-1,52933186	2,53934614	-0,65258138	1417694_at
Trim30d	0,78910948	1,68267376	2,58440635	-1,25659253	1438716_at
NA	-2,73284679	-1,35668336	2,62359301	1,08401475	1438841_s_at
Cd52	1,09837576	1,67485442	2,69730153	-0,85768203	1460218_at
Plau	-4,02605298	-0,85497603	2,87979557	0,56238323	1422139_at
Pyhin1	1,28385881	1,88806631	3,23274424	-0,80873368	1435331_at
Ctsh	1,19040417	2,01221281	3,5514126	-1,15606608	1418365_at
Slamf8	1,4639598	2,83234745	3,82499698	0,32144584	1425294_at
NA	1,11814964	3,10282901	3,97217141	-0,50274428	1449580_s_at
Slamf7	1,44199561	2,30872604	4,45342256	-1,12424975	1453472_a_at
Ly6a	1,13436323	3,67866187	5,34029846	-1,16735969	1417185_at

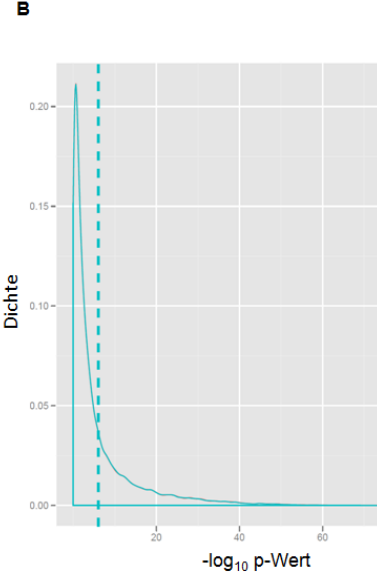
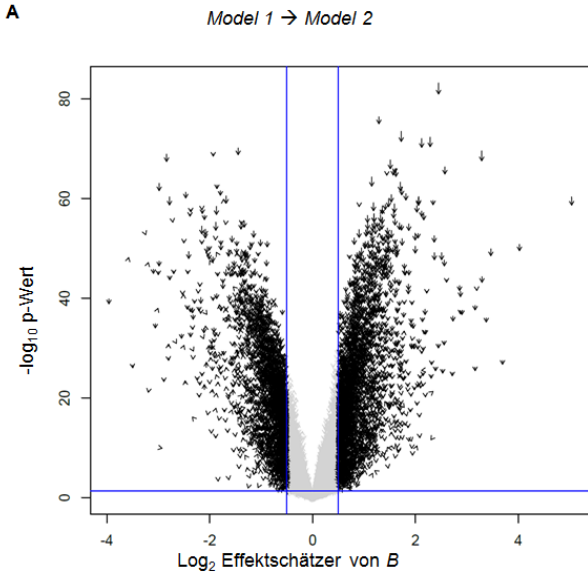
6.3 Anreicherungsanalyse der Interaktionskovariate

Category	Term	Count	PValue	Genes	List Total	Benjani
GOTERM_BP_FAT	GO:0006955~immune response	32	5,19E+03	1417378_AT, 1417936_AT, 1422891_AT, 1417976_AT, 1419135_AT, 1425407_S_AT, 1450424_A_AT, 1422892_S_AT, 1437726_X_AT, 1431611_A_AT, 1420361_AT, 1420549_AT, 1418638_AT, 1421210_AT, 1423153_X_AT, 1452425_AT, 1416978_AT, 1417377_AT, 1425406_AT, 1422527_AT, 1420664_S_AT, 1450582_AT, 1448620_AT, 1422013_AT, 1424832_AT, 1419693_AT, 1431008_AT, 1417376_A_AT, 1417063_AT, 1451593_AT, 1418666_AT, 1417314_AT, 1451644_A_AT, 1425548_A_AT, 1425519_A_AT, 1427236_A_AT, 1448898_AT, 1449580_S_AT, 1419728_AT, 1450876_AT, 1434366_X_AT, 1418480_AT, 1421596_S_AT, 1419297_AT	208	8,68E+06
GOTERM_BP_FAT	GO:0048002~antigen processing and presentation of peptide antigen	11	7,14E+03	1427797_S_AT, 1422891_AT, 1431008_AT, 1422892_S_AT, 1420361_AT, 1451593_AT, 1418989_AT, 1418638_AT, 1451644_A_AT, 1422476_AT, 1425519_A_AT, 1449580_S_AT, 1422527_AT, 1419297_AT, 1448620_AT	208	5,97E+07
GOTERM_BP_FAT	GO:0019882~antigen processing and presentation	14	6,21E+05	1427797_S_AT, 1422891_AT, 1431008_AT, 1422892_S_AT, 1420361_AT, 1451593_AT, 1418989_AT, 1451644_A_AT, 1418638_AT, 1422476_AT, 1425519_A_AT, 1416978_AT, 1449580_S_AT, 1422527_AT, 1450582_AT, 1420664_S_AT, 1419297_AT, 1448620_AT	208	3,46E+08
GOTERM_BP_FAT	GO:0002478~antigen processing and presentation of exogenous peptide antigen	9	1,02E+07	1427797_S_AT, 1422476_AT, 1422891_AT, 1425519_A_AT, 1422892_S_AT, 1449580_S_AT, 1451593_AT, 1418989_AT, 1418638_AT, 1422527_AT, 1448620_AT, 1419297_AT	208	4,27E+08
GOTERM_BP_FAT	GO:0019884~antigen processing and presentation of exogenous antigen	9	6,06E+04	1427797_S_AT, 1422476_AT, 1422891_AT, 1425519_A_AT, 1422892_S_AT, 1449580_S_AT, 1451593_AT, 1418989_AT, 1418638_AT, 1422527_AT, 1448620_AT, 1419297_AT	208	2,03E+09
GOTERM_BP_FAT	GO:0002495~antigen processing and presentation of peptide antigen via MHC class II	7	8,20E+07	1427797_S_AT, 1422476_AT, 1422891_AT, 1425519_A_AT, 1422892_S_AT, 1449580_S_AT, 1418989_AT, 1418638_AT, 1422527_AT, 1419297_AT	208	2,28E+11
GOTERM_BP_FAT	GO:0019886~antigen processing and presentation of exogenous peptide antigen via MHC class II	7	8,20E+07	1427797_S_AT, 1422476_AT, 1422891_AT, 1425519_A_AT, 1422892_S_AT, 1449580_S_AT, 1418989_AT, 1418638_AT, 1422527_AT, 1419297_AT	208	2,28E+11
GOTERM_BP_FAT	GO:0002504~antigen processing and presentation of	7	2,67E+06	1427797_S_AT, 1422476_AT, 1422891_AT, 1425519_A_AT, 1422892_S_AT, 1449580_S_AT, 1418989_AT, 1418638_AT, 1422527_AT,	208	6,39E+10



































	peptide or polysaccharide antigen via MHC class II			1419297_AT		
GOTERM_BP_FAT	GO:0050778~positive regulation of immune response	12	7,26E+09	1417378_AT, 1422891_AT, 1417976_AT, 1417376_A_AT, 1431008_AT, 1422892_S_AT, 1431611_A_AT, 1437726_X_AT, 1420361_AT, 1417063_AT, 1451593_AT, 1417314_AT, 1451644_A_AT, 1423153_X_AT, 1417377_AT, 1450876_AT, 1422527_AT, 1434366_X_AT, 1419431_AT, 1448620_AT	208	0.0015168636532536528
GOTERM_BP_FAT	GO:0002684~positive regulation of immune system process	14	1,63E+11	1417378_AT, 1417976_AT, 1422891_AT, 1417376_A_AT, 1431008_AT, 1422892_S_AT, 1431611_A_AT, 1437726_X_AT, 1417063_AT, 1420361_AT, 1451593_AT, 1451644_A_AT, 1417314_AT, 1425519_A_AT, 1416111_AT, 1423153_X_AT, 1417377_AT, 1450876_AT, 1434366_X_AT, 1422527_AT, 1419431_AT, 1448620_AT	208	0.003025568132322465
GOTERM_BP_FAT	GO:0002252~immune effector process	11	2,25E+09	1422891_AT, 1431008_AT, 1422892_S_AT, 1437726_X_AT, 1420361_AT, 1417063_AT, 1418666_AT, 1417314_AT, 1451644_A_AT, 1427236_A_AT, 1425519_A_AT, 1423153_X_AT, 1450876_AT, 1422527_AT, 1434366_X_AT, 1448620_AT	208	0.003750502537754552
GOTERM_BP_FAT	GO:0006909~phagocytosis	7	9,42E+10	1419693_AT, 1452784_AT, 1420361_AT, 1450883_A_AT, 1418666_AT, 1421198_AT, 1450884_AT, 1423166_AT, 1422869_AT, 1448620_AT	208	0.014212653164402678
GOTERM_BP_FAT	GO:0048584~positive regulation of response to stimulus	12	1,32E+12	1417378_AT, 1422891_AT, 1417976_AT, 1417376_A_AT, 1431008_AT, 1422892_S_AT, 1431611_A_AT, 1437726_X_AT, 1420361_AT, 1417063_AT, 1451593_AT, 1417314_AT, 1451644_A_AT, 1423153_X_AT, 1417377_AT, 1450876_AT, 1422527_AT, 1434366_X_AT, 1419431_AT, 1448620_AT	208	0.01819790506492991
GOTERM_BP_FAT	GO:0051050~positive regulation of transport	10	1,87E+12	1450852_S_AT, 1425548_A_AT, 1451020_AT, 1434557_AT, 1419135_AT, 1422468_AT, 1449041_A_AT, 1420361_AT, 1418666_AT, 1448620_AT, 1422467_AT	208	0.023778833181729575
GOTERM_BP_FAT	GO:0002460~adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily	8	2,89E+11	1422891_AT, 1425519_A_AT, 1422892_S_AT, 1450424_A_AT, 1437726_X_AT, 1417063_AT, 1420361_AT, 1434366_X_AT, 1417314_AT, 1422527_AT, 1448620_AT	208	0.0339204859121951

	domains					
GOTERM _BP_FAT	GO:0002250~adaptive immune response	8	2,89E+ 11	1422891_AT, 1425519_A_AT, 1422892_S_AT, 1450424_A_AT, 1437726_X_AT, 1417063_AT, 1420361_AT, 1434366_X_AT, 1417314_AT, 1422527_AT, 1448620_AT	208	0.033 92048 59121 951
GOTERM _BP_FAT	GO:0022402~cell cycle process	17	3,46E+ 12	1451020_AT, 1424278_A_AT, 1415849_S_AT, 1435177_A_AT, 1420172_AT, 1417910_AT, 1455869_AT, 1448635_AT, 1422513_AT, 1427236_A_AT, 1421546_A_AT, 1428848_A_AT, 1417420_AT, 1418719_AT, 1419431_AT, 1433543_AT, 1427062_AT, 1453307_A_AT	208	0.037 85291 83583 4492
GOTERM _BP_FAT	GO:0001775~cell activation	13	3,94E+ 11	1456174_X_AT, 1417976_AT, 1419135_AT, 1450976_AT, 1420172_AT, 1448765_AT, 1420361_AT, 1423413_AT, 1420760_S_AT, 1435386_AT, 1425548_A_AT, 1427236_A_AT, 1425519_A_AT, 1437271_AT, 1437270_A_AT, 1422527_AT, 1448620_AT, 1450262_AT	208	0.040 30295 99011 2198
GOTERM _BP_FAT	GO:0002443~leukocyte mediated immunity	8	4,12E+ 11	1422891_AT, 1425519_A_AT, 1427236_A_AT, 1422892_S_AT, 1437726_X_AT, 1417063_AT, 1420361_AT, 1434366_X_AT, 1417314_AT, 1422527_AT, 1448620_AT	208	0.039 74353 64294 6782

6.4 Abbildung des Eruptionsdiagramms von B



6.5 Legende der Ingenuity Signalwege

Network Shapes	Path Designer Shapes
 Cytokine	 Cytokine / Growth Factor
 Growth Factor	 Drug
 Chemical / Drug / Toxicant	 Chemical / Toxicant
 Enzyme	 Enzyme
 G-protein Coupled Receptor	 G-protein Coupled Receptor
 Ion Channel	 Ion Channel
 Kinase	 Kinase
 Ligand-dependent Nuclear Receptor	 Ligand-dependent Nuclear Receptor
 Peptidase	 Peptidase
 Phosphatase	 Phosphatase
 Transcription Regulator	 Transcription Regulator
 Translation Regulator	 Translation Regulator
 Transmembrane Receptor	 Transmembrane Receptor
 Transporter	 Transporter
 microRNA	 microRNA
 Complex / Group	 Complex / Group
 Other	 Other

7 Quellenverzeichnis

- Akbari, O., P. Stock, et al. (2003). "Essential role of NKT cells producing IL-4 and IL-13 in the development of allergen-induced airway hyperreactivity." *Nat Med* 9(5): 582-588.
- Allison, D. B., X. Cui, et al. (2006). "Microarray data analysis: from disarray to consolidation and consensus." *Nat Rev Genet* 7(1): 55-65.
- Ambarus, C. A., S. Krausz, et al. (2012). "Systematic validation of specific phenotypic markers for in vitro polarized human macrophages." *J Immunol Methods* 375(1-2): 196-206.
- Annunziato, F., L. Cosmi, et al. (2012). "Defining the human T helper 17 cell phenotype." *Trends Immunol* 33(10): 505-512.
- Annunziato, F. and S. Romagnani (2009). "Do studies in humans better depict Th17 cells?" *Blood* 114(11): 2213-2219.
- Autenrieth, I. B., M. Beer, et al. (1994). "Immune responses to *Yersinia enterocolitica* in susceptible BALB/c and resistant C57BL/6 mice: an essential role for gamma interferon." *Infect Immun* 62(6): 2590-2599.
- Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B-Methodological* 57(1): 289-300.
- Beyer, A., S. Bandyopadhyay, et al. (2007). "Integrating physical and genetic maps: from genomes to interaction networks." *Nat Rev Genet* 8(9): 699-710.
- Breiman, L. (2001). "Statistical Modeling: The Two Cultures." *Statistical Science* 16(3): 199-231.
- Breslow, N. E. and N. E. Day (1980). "Statistical methods in cancer research. Volume I - The analysis of case-control studies." *IARC Sci Publ*(32): 5-338.
- Bruggeman, F. J. and H. V. Westerhoff (2007). "The nature of systems biology." *Trends in Microbiology* 15(1): 45-50.
- Capaldi, A. P., T. Kaplan, et al. (2008). "Structure and function of a transcriptional network activated by the MAPK Hog1." *Nat Genet* 40(11): 1300-1306.
- Chen, X. and J. J. Oppenheim (2011). "Resolving the identity myth: key markers of functional CD4+FoxP3+ regulatory T cells." *Int Immunopharmacol* 11(10): 1489-1496.
- Collins, S. R., K. M. Miller, et al. (2007). "Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map." *Nature* 446(7137): 806-810.
- Cosmi, L., F. Annunziato, et al. (2001). "Chemoattractant receptors expressed on type 2 T cells and their role in disease." *Int Arch Allergy Immunol* 125(4): 273-279.
- Crotty, S. (2011). "Follicular helper CD4 T cells (TFH)." *Annu Rev Immunol* 29: 621-663.
- Defrance, M. and H. Touzet (2006). "Predicting transcription factor binding sites using local over-representation and comparative genomics." *Bmc Bioinformatics* 7: 396.
- Dinu, I., J. D. Potter, et al. (2007). "Improving gene set analysis of microarray data by SAM-GS." *Bmc Bioinformatics* 8: 242.
- Edgar, R., M. Domrachev, et al. (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." *Nucleic Acids Res* 30(1): 207-210.

- Fanger, N. A., K. Wardwell, et al. (1996). "Type I (CD64) and type II (CD32) Fc gamma receptor-mediated phagocytosis by human blood dendritic cells." J Immunol 157(2): 541-548.
- Fazilleau, N., L. Mark, et al. (2009). "Follicular helper T cells: lineage and location." Immunity 30(3): 324-335.
- Gentleman, R. C., V. J. Carey, et al. (2004). "Bioconductor: open software development for computational biology and bioinformatics." Genome Biol 5(10): R80.
- Gentleman, R. C., V. J. Carey, et al. (2004). "Bioconductor: open software development for computational biology and bioinformatics." Genome Biology 5(10).
- Gibson, S. E., H. Y. Dong, et al. (2006). "Expression of the B cell-associated transcription factors PAX5, OCT-2, and BOB.1 in acute myeloid leukemia: associations with B-cell antigen expression and myelomonocytic maturation." Am J Clin Pathol 126(6): 916-924.
- Goeman, J. J., S. A. van de Geer, et al. (2004). "A global test for groups of genes: testing association with a clinical outcome." Bioinformatics 20(1): 93-99.
- Gruber, M. P., C. D. Coldren, et al. (2006). "Human lung project: evaluating variance of gene expression in the human lung." Am J Respir Cell Mol Biol 35(1): 65-71.
- Hancock, G. E., R. W. Schaedler, et al. (1986). "Yersinia enterocolitica infection in resistant and susceptible strains of mice." Infect Immun 53(1): 26-31.
- Hill, J. A., M. Feuerer, et al. (2007). "Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature." Immunity 27(5): 786-800.
- Hofer, T., H. Nathansen, et al. (2002). "GATA-3 transcriptional imprinting in Th2 lymphocytes: a mathematical model." Proc Natl Acad Sci U S A 99(14): 9364-9368.
- Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc 4(1): 44-57.
- Hummel, M., R. Meister, et al. (2008). "GlobalANCOVA: exploration and assessment of gene group effects." Bioinformatics 24(1): 78-85.
- Irizarry, R. A., B. Hobbs, et al. (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." Biostatistics 4(2): 249-264.
- Jaimovich, A., R. Rinott, et al. (2010). "Modularity and directionality in genetic interaction maps." Bioinformatics 26(12): i228-236.
- Jin, W., R. M. Riley, et al. (2001). "The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster." Nat Genet 29(4): 389-395.
- Kanehisa, M. (2002). "The KEGG database." Novartis Found Symp 247: 91-101; discussion 101-103, 119-128, 244-152.
- Kaplan, M. H. (2013). "Th9 cells: differentiation and disease." Immunol Rev 252(1): 104-115.
- Kitano, H. (2002). "Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology." Curr Genet 41(1): 1-10.
- Kohonen, T. (1990). "The Self-Organizing Map." Proceedings of the Ieee 78(9): 1464-1480.
- Korn, T., E. Bettelli, et al. (2009). "IL-17 and Th17 Cells." Annu Rev Immunol 27: 485-517.
- Krawczyk, M. and W. Reith (2006). "Regulation of MHC class II expression, a unique regulatory system identified by the study of a primary immunodeficiency disease." Tissue Antigens 67(3): 183-197.
- Lanza, A. M., N. C. Crook, et al. (2012). "Innovation at the intersection of synthetic and systems biology." Curr Opin Biotechnol 23(5): 712-717.

- Li, W. (2012). "Volcano plots in analyzing differential expressions with mRNA microarrays." *J Bioinform Comput Biol* 10(6): 1231003.
- Lipscomb, M. F. and B. J. Masten (2002). "Dendritic cells: immune regulators in health and disease." *Physiol Rev* 82(1): 97-130.
- Lu, T., Y. Pan, et al. (2004). "Gene regulation and DNA damage in the ageing human brain." *Nature* 429(6994): 883-891.
- Ma, C. S., S. G. Tangye, et al. (2010). "Human Th9 cells: inflammatory cytokines modulate IL-9 production through the induction of IL-21." *Immunol Cell Biol* 88(6): 621-623.
- MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*: 281-297.
- Marcinowski, L., M. Lidschreiber, et al. (2012). "Real-time transcriptional profiling of cellular and viral gene expression during lytic cytomegalovirus infection." *PLoS Pathog* 8(9): e1002908.
- Matys, V., O. V. Kel-Margoulis, et al. (2006). "TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes." *Nucleic Acids Res* 34(Database issue): D108-110.
- McDonald, J. F. (2011). "Integrated cancer systems biology: current progress and future promise." *Future Oncol* 7(5): 599-601.
- Mootha, V. K., C. M. Lindgren, et al. (2003). "PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes." *Nat Genet* 34(3): 267-273.
- Murray, P. J. and T. A. Wynn (2011). "Protective and pathogenic functions of macrophage subsets." *Nat Rev Immunol* 11(11): 723-737.
- Opgen-Rhein, R. and K. Strimmer (2007). "Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach." *Stat Appl Genet Mol Biol* 6: Article9.
- Pan, X., P. Ye, et al. (2006). "A DNA integrity network in the yeast *Saccharomyces cerevisiae*." *Cell* 124(5): 1069-1081.
- Prazeres da Costa, O., A. Hoffman, et al. (2014). "Selection of higher order regression models in the analysis of multi-factorial transcription data." *PLoS One* 9(3): e91840.
- Rädler, D. (2014). Mechanisms of immune regulation during development of atopic diseases in childhood: Analysis of T cell subpopulations considering genetic and epigenetic influences. Dissertation, LMU München: Medizinische Fakultät.
- Robinson, D. S. (2010). "The role of the T cell in asthma." *J Allergy Clin Immunol* 126(6): 1081-1091; quiz 1092-1083.
- Roth, F. P., H. D. Lipshitz, et al. (2009). "Q&A: epistasis." *J Biol* 8(4): 35.
- Sallusto, F., D. Lenig, et al. (1999). "Two subsets of memory T lymphocytes with distinct homing potentials and effector functions." *Nature* 401(6754): 708-712.
- Sartor, M. A., C. R. Tomlinson, et al. (2006). "Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments." *BMC Bioinformatics* 7: 538.
- Schmetterer, K. G., A. Neunkirchner, et al. (2012). "Naturally occurring regulatory T cells: markers, mechanisms, and manipulation." *FASEB J* 26(6): 2253-2276.
- Schneider, H. C. and T. Klabunde (2013). "Understanding drugs and diseases by systems biology?" *Bioorg Med Chem Lett* 23(5): 1168-1176.
- Smyth, G. K. (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." *Stat Appl Genet Mol Biol* 3: Article3.

- Spitzer, M., E. Griffiths, et al. (2011). "Cross-species discovery of syncretic drug combinations that potentiate the antifungal fluconazole." Mol Syst Biol 7: 499.
- Subramanian, A., H. Kuehn, et al. (2007). "GSEA-P: A desktop application for Gene Set Enrichment Analysis." Bioinformatics 23(23): 3251-3253.
- Subramanian, A., H. Kuehn, et al. (2007). "GSEA-P: a desktop application for Gene Set Enrichment Analysis." Bioinformatics 23(23): 3251-3253.
- Subramanian, A., P. Tamayo, et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A 102(43): 15545-15550.
- Szabo, S. J., S. T. Kim, et al. (2000). "A novel transcription factor, T-bet, directs Th1 lineage commitment." Cell 100(6): 655-669.
- Szabo, S. J., B. M. Sullivan, et al. (2003). "Molecular mechanisms regulating Th1 immune responses." Annu Rev Immunol 21: 713-758.
- Tan, C. and I. Gery (2012). "The unique features of Th9 cells and their products." Crit Rev Immunol 32(1): 1-10.
- Team, R. C. (2013). "R: A Language and Environment for Statistical Computing." from <http://www.R-project.org>.
- Thomas-Chollier, M., M. Defrance, et al. (2011). "RSAT 2011: regulatory sequence analysis tools." Nucleic Acids Res 39(Web Server issue): W86-91.
- Tosolini, M., A. Kirilovsky, et al. (2011). "Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, th2, treg, th17) in patients with colorectal cancer." Cancer Res 71(4): 1263-1271.
- Trulzsch, K., A. Roggenkamp, et al. (2003). "Analysis of chaperone-dependent Yop secretion/translocation and effector function using a mini-virulence plasmid of *Yersinia enterocolitica*." Int J Med Microbiol 293(2-3): 167-177.
- Tukey, J. W. (1980). "We Need Both Exploratory and Confirmatory." American Statistician 34(1): 23-25.
- Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A 98(9): 5116-5121.
- van Erp, K., K. Dach, et al. (2006). "Role of strain differences on host resistance and the transcriptional response of macrophages to infection with *Yersinia enterocolitica*." Physiol Genomics 25(1): 75-84.
- Vandenabeele, S., H. Hochrein, et al. (2001). "Human thymus contains 2 distinct dendritic cell populations." Blood 97(6): 1733-1741.
- VanderWeele, T. J. (2009). "On the distinction between interaction and effect modification." Epidemiology 20(6): 863-871.
- Vivier, E., D. H. Raulet, et al. (2011). "Innate or adaptive immunity? The example of natural killer cells." Science 331(6013): 44-49.
- Warringer, J., D. Anevski, et al. (2008). "Chemogenetic fingerprinting by analysis of cellular growth dynamics." BMC Chem Biol 8: 3.
- Wenzel, S. E. (2006). "Asthma: defining of the persistent adult phenotypes." Lancet 368(9537): 804-813.
- Wong, W. F., K. Kohu, et al. (2011). "Interplay of transcription factors in T-cell differentiation and function: the role of Runx." Immunology 132(2): 157-164.
- Zhang, J., S. Vaga, et al. (2011). "Mapping the interaction of Snf1 with TORC1 in *Saccharomyces cerevisiae*." Mol Syst Biol 7: 545.

- Zhang, N., H. F. Pan, et al. (2011). "Th22 in inflammatory and autoimmune disease: prospects for therapeutic intervention." Mol Cell Biochem 353(1-2): 41-46.
-

8 Abkürzungsverzeichnis

AA	Allergische Asthmatiker;
Aff. ID,	Affymetrix Gene ID;
AIC	Akaike Information Criterion;
BC,	B cell;
BH	Benjamini-Hochberg;
BIC	Baysian Information Criterion;
DAVID	Database for Annotation, Visualization and Integrated Discovery;
DC,	Dendritic cell;
FDR	False discovery rate;
GEO	Gene expression omnibus;
GK	Gesunde Kontrollen;
GO	Gene ontology; Gen-Ontologie;
GSEA	Gene set enrichment analysis;
IFN- γ	Interferon-gamma;
IgE	Immunglobulin E;
IL	Interleukin;
IPA	Ingenuity pathway analysis;
KEGG	Kyoto encyclopaedia
nTh,	Naïve T helper cell;
NAA	Nicht-allergische Asthmatiker;
NK	Natural killer cell;
LpA	Lipid A;
M	Medium;
Mac,	Macrophage;
MvA	Mean versus average;
PCA	Principal component analysis;
PBMC	Peripheral blood monocytes;
RAST	Radioallergosorbent test;
ROI	Region of interest;
RMA	Robust multi-array average;
RNA	Ribonucleic acid;
SAM	Significance Analysis of Microarrays;
SFB-TR22	Sonderforschungsbereich Transregio 22;

SOM	Self-organizing map;
TC	T cell;
TF	Transkriptionsfaktor
TFBS	Transkriptionsfaktor-Bindestelle
Tfh,	Follicular T helper cell;
TGF- β	Transforming growth factor beta;
Th cells	T helper cell;
Th1,	T helper cell type 1;
Th2,	T helper cell type 2;
Th9,	T helper cell type 9;
Th17,	T helper cell type 17;
Th22,	T helper cell type 22;
TLR	Toll-like receptor;
Treg,	Regulatory T cell;

9 Abbildungsverzeichnis

Abbildung 1: Entgegengesetzte Richtungen der Strategien in der Systembiologie.....	2
Abbildung 2: Schematische Darstellung von aggravating und alleviating Effekten.....	4
Abbildung 3: Schematische Darstellung der Interpretationsmöglichkeiten (rot, blau, grün) des Eruptionsdiagramms.....	8
Abbildung 4: Eruptionsdiagramme angewendet auf einen Simulationsdatensatz. A: Komponente B ist relevant; B: Komponente B ist irrelevant.....	9
Abbildung 5: Eruptionsdiagramme angewendet auf Simulationsdatensätze zur Identifikation von Störfaktoren und Effektmodifikation.	10
Abbildung 6: Vulkandiagramm der Interaktionskovariate von Modell 1, angewendet auf den Infektionsdatensatz.....	17
Abbildung 7: A: Eruptionsdiagramm von Modell 1 und Modell 2. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 1 und Modell 2.....	19
Abbildung 8: A: Eruptionsdiagramm von Modell 2 und Modell 3. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 2 und Modell 3.....	20
Abbildung 9: A: Eruptionsdiagramm von Modell 3 und Modell 4. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 3 und Modell 4.....	21
Abbildung 10: Clusteranalyse der Effekte und P-Werte von Modell 3.	22
Abbildung 11: GO- und TFBS Analyse von Cluster 6 der Clusteranalyse.....	23
Abbildung 12: Genexpression von H2-Ea-ps des genetischen Hintergrund <i>H</i> , BALB/ <i>c</i> und C57BL/6.....	24
Abbildung 13: PCA Analyse der Asthmadaten. Jede Farbe kennzeichnet ein Scan Datum der Microarrays.....	26
Abbildung 14: Boxplots der Asthmadaten.	27
Abbildung 15: A: Eruptionsdiagramm von AA. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 1 und Modell 2.....	28
Abbildung 16: A: Eruptionsdiagramm von NAA. B: Dichtediagramm der $-\log_{10}$ p-Werte von Modell 1 und Modell 2.....	29
Abbildung 17: Differentiell exprimierte Gene der Asthma-Microarray-Analyse.	30
Abbildung 18: Expressionsunterschiede im TGF- β -Signalweg zwischen AA und GK unter LpA-Stimulation.....	37
Abbildung 19: Expressionsunterschiede im TLR-Signalweg zwischen NAA gegen GK unter LpA-Stimulation.....	39

Abbildung 20: Expressionsunterschiede im mTOR-Signalweg zwischen AA gegen GK unter M-Stimulation.....	41
Abbildung 21: Balkendiagramm des Einflusses der einzelnen Gene der Th2-Zellen auf das Ergebnis der Teststatistik.....	41
Abbildung 22: Balkendiagramm des Einflusses der einzelnen Gene der Treg-Zellen auf das Ergebnis der Teststatistik.....	42
Abbildung 23: Balkendiagramm des Einflusses der einzelnen Gene der Makrophagen auf das Ergebnis der Teststatistik.....	43
Abbildung 24: Das Balkendiagramm der CD4 ⁺ -Zellen zeigt den Einfluss der einzelnen Gene der naiven Th-Zellen auf das Ergebnis der Teststatistik auf.....	45
Abbildung 25: Vulkandiagramm der Interaktionskovariate von Modell 3.....	53
Abbildung 26: GSEA-Analyse von allen signifikanten Genen der Interaktionskovariate von Modell 3.....	53
Abbildung 27: DAVID-Anreicherungsverfahren aller signifikanten Gene der Interaktionskovariate von Modell 3.....	54
Abbildung 28: Aggravating Effekte der GSEA-Analyse.....	55
Abbildung 29: Aggravating Effekte der DAVID-Anreicherungsanalyse.....	56
Abbildung 30: Alleviating Effekte der GSEA-Analyse.....	57
Abbildung 31: Alleviating Effekte der DAVID-Anreicherungsanalyse.....	58
Abbildung 32: GSEA-Analyse aller signifikanten Gene der Kovariate AA.....	61
Abbildung 33: DAVID-Anreicherungsanalyse aller signifikanten Gene der Kovariate AA.....	62
Abbildung 34: GSEA-Analyse der Gene, die von einem Störfaktor beeinflusst werden.....	62
Abbildung 35: DAVID-Anreicherungsanalyse der Gene, die von einem Störfaktor beeinflusst werden.....	63
Abbildung 36: GSEA-Analyse der Gene, die eine Veränderung im p-Wert zeigen.....	63
Abbildung 37: DAVID-Anreicherungsanalyse der Gene, die eine Veränderung im p-Wert zeigen.....	64

10 Tabellenverzeichnis

Tabelle 1: Experimenteller Aufbau und Anzahl der Replikate des Infektionsdatensatzes.....	5
Tabelle 2: Anzahl der PBMCs Microarrays.....	14
Tabelle 3: Anzahl der CD4 ⁺ -Microarrays	14
Tabelle 4: Die linearen Regressionsmodelle des Infektionsdatensatzes.	18
Tabelle 5: Lineare Regressionsmodelle der Asthmadaten.....	28
Tabelle 6: Charakteristische Gene von verschiedenen T-Zellen.....	31
Tabelle 7: Regulation der KEGG Signalwege und T-Zell Gruppen-spezifische Gene.....	34
Tabelle 8: Regulation der KEGG Signalwege und T-Zell Gruppen-spezifischer Gene im Vergleich PBMCs gegen CD4 ⁺ -Zellen	45
Tabelle 9: Vergleich der alleviating und aggravating Anreicherungsanalyse.....	59
Tabelle 10: Vergleich der Anreicherungsanalysen der Asthmadaten.....	65

11 Danksagung

Da ich diese Dissertation am Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie der Medizinischen Fakultät durchgeführt habe, möchte ich meinem Doktorvater Prof. Dr. rer. nat. Ulrich Mansmann für die Themenwahl dieser Dissertation herzlich danken. Die konstruktiven Diskussionen haben maßgeblich zu der Veröffentlichung der beiden Fachartikel und der vorliegenden Dissertation beigetragen. In Zusammenarbeit mit Dr. rer. nat. Achim Tresch konnte ein Fachartikel publiziert werden, der Thema dieser Dissertation ist.

Auch Prof. Dr. med. Busch und PD Dr. med. Reinhard Hoffmann bin ich sehr zu Dank verpflichtet, da sie mir die Möglichkeit gaben, am Institut für Medizinische Mikrobiologie, Immunologie und Hygiene meine Tätigkeit als wissenschaftliche Mitarbeiterin durchzuführen. In Vorträgen habe ich wertvolles Feedback erhalten. Auch in den wöchentlichen Meetings mit Prof. Dr. rer. nat. Buch habe ich meine Forschungsarbeit besprechen können.

Ich möchte den Mitgliedern der Arbeitsgruppe Buch (Jane Beil, Lynsey Fairbairn, Karin Mink, Pushpalatha Palle, Ravindra Sapkota, Martin Skerhut, Silvia Weidner) für Ihre Unterstützung danken.

Großer Dank geht an die Deutsche Forschungsgemeinschaft, die mir die Möglichkeit gab im Rahmen des Sonderforschungsbereich TR22 die Serviceplattform „Microarrays“ zu führen. Dies gab mir die Möglichkeit zu vielfältigen Kooperationen. Aus einer dieser Kooperationen geht das zweite Projekt dieser Arbeit hervor. Daher möchte ich den Mitgliedern dieser fruchtbaren Zusammenarbeit (PD Dr. med. Bianca Schaub, Dr. rer. nat. Diana Rädler und Dr. rer. hum. biol. Nikolaus Ballenberger) danken.

Schließlich möchte ich meinem Mann für objektive Kritik, meiner Tochter für nötige Ablenkung und meinen Eltern für Ihre Zeit als Babysitter danken.

12 Lebenslauf

Olivia Wanda Maria Prazeres da Costa, geboren am 30.03.1978 in München

30.03.1978	Geburt in München
1984-1988	Besuch der Grundschule Taufkirchen
1988-1998	Besuch des Gymnasiums Unterhaching
25.Juni 1998	Allgemeine Hochschulreife
2000-2004	Bachelorstudium Bioinformatik an der Ludwig-Maximilians-Universität München und der Technischen Universität München
23.07.2004	Bachelor of Science
2004-2006	Masterstudium Genetik (Schwerpunkt Evolutionsgenetik) an der Universität Aut3noma de Barcelona
27.10.2006	Master of Research in Genetik
2007 – 2008	Projektmanagerin bei Extedo GmbH
2008-2009	Unternehmensberaterin bei Accenture GmbH
Seit 01.09.2009	Wissenschaftliche Mitarbeiterin am Institut für Mikrobiologie, Immunologie und Hygiene der Technischen Universität München
Seit 27.07.2011	Gastwissenschaftlerin am Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie der Ludwig-Maximilians-Universität München

Eidesstattliche Versicherung

Name, Vorname

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift Doktorandin/Doktorand