

**Detektion und Analyse synergistischer  
Einflüsse multipler Stimuli auf zelluläre  
Signaltransduktionsnetzwerke in  
Säugetieren**

**Dissertation**

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades  
"Doctor rerum naturalium"  
der Georg-August-Universität Göttingen

im Promotionsprogramm Environmental Informatics  
der Georg-August University School of Science (GAUSS)

vorgelegt von  
Sebastian Zeidler

aus Karlsburg  
Göttingen, 2013

Betreuungsausschuss:

Prof. Dr. Edgar Wingender, Bioinformatik, UMG

Prof. Dr. Stephan Waack, Theoretische Informatik und Algorithmische Methoden

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Edgar Wingender

Korreferent: Prof. Dr. Stephan Waack

weitere Mitglieder der Prüfungskommission:

Prof. Dr. Burkhard Morgenstern

Prof. Dr. Tim Beißbarth

Prof. Dr. Carsten Damm

Prof. Dr. Dieter Hogrefe

Tag der mündlichen Prüfung: 18.09.2013

# Danksagung

Ich möchte mich bei jenen bedanken, die mich durch ihr fachliches und persönliches Engagement während dieser Arbeit begleitet und unterstützt haben. Zuerst sei dies Prof. Dr. Edgar Wingender, der mir die Bearbeitung dieses spannenden Themas ermöglichte, der mich mit Diskussionen und Anregungen aus seinem umfangreichen wissenschaftlichen Erfahrungsschatz unterstützte und maßgeblich zum Abschluss beitrug. Ich möchte außerdem Prof. Dr. Stephan Waack für den wissenschaftlichen Diskurs danken sowie für seine Tätigkeit als Korreferent und Mitglied im Betreuungsausschuss. Ein großer Dank gilt Dr. Anatolij Potapov, der mir durch viele intensive Diskussionen half, tiefer in die biologische Problematik einzutauchen und Brücken zwischen den verschiedenen Disziplinen zu bauen. Ein besonderer Dank geht an Dr. Björn Goemann, der mir mit seiner Expertise und Diskussionsbereitschaft als Freund und Kollege immer zur Verfügung stand und dabei half die verschiedenen Fragestellungen unter neuen Blickwinkeln zu betrachten.

Den Mitgliedern und Alumni des Instituts für Bioinformatik sei dafür gedankt, dass sie stets ein offenes Ohr für Fragen hatten und mir so halfen die Tücken des wissenschaftlichen Alltags zu meistern. Ich bedanke mich bei Doris Waldmann und Carmen Modrok, die mich bei administrativen Fragen unterstützten und ohne die das Einleben in Göttingen schwieriger gewesen wäre. Außerdem bedanke ich mich bei Torsten Schöps für die technische Unterstützung sowie bei Raphaël Zöllinger für die Durchführung der Experimente und die Bereitstellung der Daten.

Ein tief empfundener Dank geht an Anne, Ulrike, Christian, Kathrin, Matthias und die drei Michas sowie allen hier Ungenannten, die mir als alte oder neue Freunde in den Hochs und Tiefs den Rücken und den Blick für das Wesentliche schärften. Gedankt sei auch den Göttinger 7, die den wesentlichen Kontrast zur wissenschaftlichen Arbeit setzten und mit denen ich tolle Zeiten erleben durfte.

Der größte Dank jedoch gilt meinen Eltern, ohne deren permanente Unterstützung dieser Bildungsweg für mich nicht möglich gewesen wäre und die mich immer im Streben bestärkten nach den *Sternen* zu greifen.





# Inhaltsverzeichnis

Inhaltsverzeichnis . . . . .	i
Abkürzungsverzeichnis . . . . .	v
Tabellenverzeichnis . . . . .	vii
Abbildungsverzeichnis . . . . .	ix
1 Einleitung . . . . .	1
2 Grundlagen . . . . .	5
2.1 Kommunikation in mehrzelligen Organismen . . . . .	6
2.1.1 Informationsaustausch zwischen Zellen . . . . .	6
2.1.2 Signaltransduktion in Zellen . . . . .	9
2.1.3 Netzwerke ermöglichen einen systemischen Blick auf Signal- transduktion . . . . .	12
2.1.4 Interferenz und Synergie zwischen verschiedenen Signalpfaden	14
2.2 Der Einfluss verschiedener Signale auf plasmacytoide dendritische Zellen im menschlichen Immunsystem . . . . .	19
2.2.1 Angeborenes und adaptives Immunsystem des Menschen . . .	19
2.2.2 Plasmacytoide dendritische Zellen verbinden angeborenes und adaptives Immunsystem . . . . .	20
2.2.3 Interleukin-3, ein aktivierendes inflammatorisches Signal . . .	24
2.2.4 Das Influenza-A-Virus im Immunsystem . . . . .	26
3 Material und Methoden . . . . .	29
3.1 Datenbanken . . . . .	29
3.1.1 TRANSFAC® . . . . .	29
3.1.2 TRANSPATH® . . . . .	30
3.1.3 Kyoto Encyclopedia of Genes and Genomes (KEGG) . . . . .	32
3.2 Programme . . . . .	33
3.2.1 GraphViz . . . . .	33
3.2.2 R . . . . .	33

3.3	Bestimmung von Genexpression und Identifikation differentiell exprimierter Gene . . . . .	35
3.3.1	Microarrays - globale und simultane Bestimmung von Genexpression . . . . .	35
3.3.2	Identifikation differentiell exprimierter Gene mittels <i>limma</i> . . . . .	36
3.4	Datensatz - Experimentell bestimmte Genexpression in pDC . . . . .	40
3.5	Datenabstraktion bei der Netzwerkmodellierung . . . . .	43
4	Ergebnisse und Diskussion . . . . .	47
4.1	Detektion von Synergien in pDC mittels Genexpressionsanalyse . . . . .	47
4.1.1	Durch IL-3- und Flu- induzierte Synergie in pDC . . . . .	49
4.1.2	Funktionen synergistischer Gene . . . . .	55
4.1.3	Zusammenfassung . . . . .	61
4.2	Erstellung eines IL-3- und Flu-spezifischen Netzwerks . . . . .	63
4.2.1	Netzwerkmodellierung mittels bipartiter Graphen . . . . .	64
4.2.2	Inkorporation spezieübergreifender Informationen . . . . .	68
4.2.3	Einbindung unterschiedlich granularen Wissens . . . . .	70
4.2.4	Zellspezifität für pDC durch Entfernung nicht exprimierter Gene und Proteine . . . . .	73
4.2.5	IL-3- und Flu-spezifische Signaltransduktion in pDC . . . . .	76
4.2.6	Zusammenfassung . . . . .	80
4.3	Detektion und Propagierung synergistischer Effekte im bipartiten Signaltransduktionsnetzwerk . . . . .	82
4.3.1	Detektion von Synergiekandidaten . . . . .	88
4.3.2	Entfernung uninformativer synergistischer Kandidaten . . . . .	91
4.3.3	Detektion von Abhängigkeiten zwischen synergistischen Molekülen . . . . .	96
4.3.4	Transkriptionsfaktoren propagieren Synergien auf die genomische Ebene . . . . .	99
4.3.5	Funktionen synergistisch beeinflusster Gene . . . . .	100
4.3.6	Anwendung des Algorithmus auf das IL-3- und Flu-spezifische Signaltransduktionsnetzwerk . . . . .	101
4.3.7	Zusammenfassung . . . . .	117

4.4 Vergleich zwischen experimenteller und algorithmischer Methode . . . . .	119
4.4.1 Methodischer Vergleich der Ansätze zur Detektion von Syn- ergien . . . . .	119
4.4.2 Vergleich der Immunrelevanz der experimentell und algorithmisch bestimmten Gene . . . . .	122
4.4.3 Zusammenfassung . . . . .	127
5 Fazit und Ausblick . . . . .	129
Literatur . . . . .	133
Appendix . . . . .	157



# Abkürzungsverzeichnis

<b>Abb.</b>	Abbildung
<b>bp</b>	Basenpaare
<b>bspw.</b>	beispielsweise
<b>bzw.</b>	beziehungsweise
<b>ca.</b>	zirka
<b>DC</b>	dendritische Zellen
<b>d.h.</b>	das heißt
<b>DNA</b>	(engl.) deoxyribonucleic acid
<b>DNA-bP</b>	DNA-bindende-Proteine
<b>etc.</b>	<i>et cetera</i>
<b>ff.</b>	fort folgende
<b>Flu</b>	Influenza-A-Virus
<b>GCRMA</b>	(engl.) GC-robust multi-array average
<b>GEF</b>	(engl.) guanine-nucleotide exchange factor
<b>GEO</b>	(engl.) Gene Expression Omnibus
<b>GO</b>	(engl.) The Gene Ontology
<b>GSEA</b>	(engl.) Gene Set Enrichment Analysis
<b>h</b>	Stunden
<b>HRE</b>	Hauptregulationsebene

<b>IL</b>	Interleukin(e)
<b>JAK</b>	Janus Kinase
<b>KEGG</b>	Kyoto Encyclopedia of Genes and Genomes
<b>limma</b>	(engl.) linear models for microarray analysis
<b>MHC</b>	(engl.) major histocompatibility complex
<b>mRNA</b>	(engl.) messenger-RNA
<b>NK-Zellen</b>	natürliche Killerzellen
<b>pDC</b>	plasmacytoide dendritische Zellen
<b>PDI</b>	(engl.) pairwise disconnectivity index
<b>RMA</b>	(engl.) robust multi-array average
<b>RNA</b>	(engl.) ribonucleic acid
<b>STAT</b>	(engl.) signal transducer and activator of transcription
<b>sd</b>	(engl.) standard deviation
<b>ssRNA</b>	(engl.) single stranded RNA
<b>Tab.</b>	Tabelle
<b>TF</b>	Transkriptionsfaktor(en) <i>pl.</i> TFs
<b>TFBS</b>	Transkriptionsfaktorbindestelle(n)
<b>TLR</b>	<i>toll-like</i> Rezeptor
<b>usw.</b>	und so weiter
<b>z.B.</b>	zum Beispiel

# Tabellenverzeichnis

4.1	Repräsentation der 22 experimentell bestimmten und am stärksten synergistisch regulierten Gene und ihrer Funktionen . . . . .	57
4.2	Darstellung wichtiger Parameter des rekonstruierten IL-3-, Flu-signalspezifischen Netzwerks in pDC . . . . .	78
4.3	Darstellung der Beziehungen zwischen den TF-Regulationsebenen und ihren regulierten Genen sowie Betrachtungen zur Bedeutung von TF-Komplexen . . . . .	103
4.4	Immunrelevanz der 18 synergistisch regulierten TFs der Hauptregulationsebene und ihre Funktionen . . . . .	107
4.5	Repräsentation der algorithmisch bestimmten synergistisch regulierten Gene und ihrer Funktionen . . . . .	112
4.6	Vergleich zwischen experimenteller und algorithmischer Methode zur Bestimmung von Synergien . . . . .	120
4.7	Ergebnisvergleich zwischen algorithmischer und experimenteller Methode zum Auffinden von Synergien . . . . .	123
5.1	Darstellung von in Tab. 4.1 nicht aufgeführten experimentell identifizierten synergistischen Genen und die Bewertung ihrer Rolle in der Immunantwort . . . . .	161





# Abbildungsverzeichnis

2.1	Kommunikationsformen zum Signalaustausch zwischen verschiedenen Zellen . . . . .	7
2.2	Schematischer Ablauf von enzymrezeptorabhängiger Signaltransduktion . . . . .	11
2.3	Schematische Darstellung einer antagonistischen Interaktion zwischen zwei Signalmolekülen . . . . .	14
2.4	Schematische Darstellung einer agonistischen Interaktion zwischen zwei Signalmolekülen . . . . .	15
2.5	Schematische Darstellung eines durch zwei Signalmoleküle ausgelösten synergistischen Signalpfades . . . . .	17
2.6	Illustration einer zu pDC morphologisch ähnlichen Plasmazelle . . . . .	21
2.7	pDCs in der hämatopoietischen Stammzelllinie . . . . .	23
3.1	Orthologe und hierarchische Abbildungen in TRANSPATH® . . . . .	31
3.2	Experimenteller Ablauf der Extraktion von pDC aus Blut und Gewinnung der Daten . . . . .	41
3.3	Orthologe und hierarchische Abstraktionsprinzipien . . . . .	45
3.4	Zusammenfassung orthologer und hierarchischer Netzwerkabstraktionsprinzipien am Beispiel der Vav-Proteinfamilie . . . . .	46
4.1	Darstellung der differentiellen Genexpression anhand der auf einem Microarray aufgebrachten Sonden in den verschiedenen experimentellen Bedingungen . . . . .	50
4.2	Identifikation von differentiell exprimierten, synergistisch regulierten Genen durch ihre Sonden mit Hilfe eines Venn-Diagramms . . . . .	51
4.3	Genexpressionsunterschiede der synergistischen Gene im Vergleich zwischen Synergiebedingung und Kontrolle . . . . .	54

---

4.4	Schematische unipartite Darstellung von zwei Beispielreaktionen zur Synthese von CO und Probleme bei ihrer Abbildung . . . . .	67
4.5	Schematische Darstellung der Beispielreaktionen durch einen bipartiten Netzwerk-Modellierungsansatz . . . . .	67
4.6	Integration von orthologen und hierarchischen Informationen zur Komplettierung des zu rekonstruierenden Netzwerks . . . . .	69
4.7	Abstraktion von Wissen auf eine gemeinsame speziesunspezifische orthologe Ebene während der Netzwerkrekonstruktion . . . . .	71
4.8	Hierarchische Beziehungen zwischen Molekülen und ihre Abbildung bei der Modellierung orthologer bipartiter Prozesse . . . . .	72
4.9	Zelltypspezifität durch Filterung nicht-exprimierter Knoten und die Auswirkungen auf die hierarchischen Abhängigkeiten . . . . .	74
4.10	Anpassung der Interaktionen im Netzwerkgraphen nach Filterung der nicht-exprimierten Knoten . . . . .	75
4.11	Schematische Ableitung des signalabhängigen Netzwerkgraphen .	76
4.12	Vergleich von PDI und Synergie in bipartiten Graphen . . . . .	84
4.13	Schematische Darstellung von Signaltransduktion im Bezug auf Synergien . . . . .	85
4.14	Detektion von Synergiekandidaten durch den Algorithmus in einem Beispielgraph . . . . .	90
4.15	Algorithmische Entfernung von Synergiekandidaten aus dem Beispielgraph . . . . .	95
4.16	Darstellung der Abhängigkeiten zwischen detektierten synergistischen Molekülen im Beispielgraph . . . . .	98
4.17	Transformation von Abhängigkeiten zwischen synergistischen Molekülen in gruppierende Regulationsebenen . . . . .	102
4.18	Beziehungen zwischen Regulationsebenen im pDC-spezifischen IL-3-, Flu-Signaltransduktionsnetzwerk . . . . .	102
4.19	Schema der proteomischen Regulation zwischen den TFs und ihre Wirkungen auf die genomische Ebene . . . . .	106
4.20	Schematische Darstellung von gewebespezifischer Genregulation durch einen TF und die Folgen bei einer orthologen Modellierung	108

4.21	Änderung der Genexpression der algorithmisch bestimmten synergistischen Gene im experimentellen Vergleich von synergistischer Bedingung und Kontrolle . . . . .	110
4.22	Schema der aktivierten Immunantwort während 24h, ausgelöst durch die Signale IL-3 und Flu sowie die Differenzierung von pDC zu DC . . . . .	116
4.23	Unterschiede zwischen Dominanzbeziehungen und Synergie . . . .	121
4.24	Vergleich zwischen experimentell und algorithmisch verifizierten immunrelevanten Genen bezüglich verschiedenen Funktionsgruppen	123
5.1	Algorithmische Bestimmung der Abhängigkeiten zwischen verschiedenen synergistischen Molekülen im Beispielgraph . . . . .	160



# 1 Einleitung

Das Immunsystem schützt Organismen vor Krankheitserregern (Pathogenen) der Umwelt. Als Folge von Verletzungen dringen Pathogene in den Organismus ein, wo sie von spezialisierten Immunzellen perzipiert werden. Die Pathogene aktivieren dadurch intrazelluläre Signalfade in der Immunzelle. Entlang von Signalweiterleitungskaskaden werden die aufgenommenen Signale in der Zelle prozessiert und regulieren verschiedene Zielgene. Die Gesamtheit an biologischen Prozessen von der Signalperzeption über die Weiterleitung bis hin zur Genregulation nennt man Signaltransduktion. Durch die Regulation der Zielgene und den sich daraus ableitenden Änderungen des Proteoms oder der ausgeschütteten Signale ist die Zelle in der Lage auf ein Pathogen zu reagieren und es zu bekämpfen.

Von verschiedenen Pathogenen aktivierte Signalfade interagieren während der Signalweiterleitung in ihren einzelnen Komponenten, z.B. Proteinkomplexen, miteinander, wodurch die Zelle die „Bedrohungslage“ ganzheitlich bewertet und eine adäquate Immunantwort auslöst. Dadurch können Pathogene beispielsweise zusammen stärkere Effekte auslösen als jeder Krankheitserreger allein, oder die gemeinsame Wirkung verschiedener Erreger erzeugt neue Effekte, die nicht durch die einzelnen Pathogene induziert werden. Derartige Effekte nennt man Synergien, sie wurden bereits im Immunsystem von Säugern beobachtet (siehe Rissoan *et al.*, 1999; Shelton *et al.*, 2004; Gouwy *et al.*, 2004; Doreau *et al.*, 2009).

Die Vielzahl an Pathogenen, die miteinander interagieren können, sowie die unüberschaubare Menge an Möglichkeiten, wo Synergien während der Signalweiterleitung in den Zellen auftreten, erschwert ihre Identifikation. Bislang jedoch sind Synergien aufgrund ihrer Komplexität experimentell selten untersucht worden und werden systematisch vernachlässigt. Die Detektion von Synergien beschränkt sich momentan auf experimentell gemessene Mengenunterschiede zwischen verschiedenen Messgrößen. Dadurch finden systemische Betrachtungen nur

auf der Ebene der Genexpression statt. Dies detektiert die synergistisch regulierten Gene, aber die kausalen Zusammenhänge zwischen Signalpfadinteraktion und der beobachteten Änderung in der Genexpression werden nicht aufgeklärt. Um Synergien ganzheitlich zu verstehen, muss verstanden werden, wie verschiedene Signale eine gemeinsame synergistische Wirkung erzeugen. Dazu ist die Aufklärung der Orte, an denen Synergie auftritt, ebenso notwendig wie die Verfolgung der Prozessierung während der Signaltransduktion und die Identifikation synergistisch regulierter Gene und ihrer Effekte.

Deshalb existiert ein großer Bedarf an der Entwicklung ganzheitlicher Analysemethoden, mit denen die komplexe Reaktion eines Organismus auf Umwelteinflüsse analysiert werden kann. Mithilfe von systembiologischen Untersuchungen können solche Sachverhalte untersucht werden. Dadurch werden neue Einblicke in die Funktionsweise des Organismus gewonnen, etwa darüber, wie Synergien auf das Immunsystem einwirken und es regulieren. Bisher jedoch sind für derartige Fragestellungen keine adäquaten Methoden verfügbar.

Im Rahmen des GlobCell-Projekts - was auch diese Arbeit umklammert - wurden Synergien systembiologisch untersucht. Dafür wurde hier ein neues Netzwerk-basiertes Verfahren entwickelt, welches die Orte synergistischer Signalpfadinteraktion von verschiedenen Stimuli anhand von strukturellen Eigenschaften im Signaltransduktionsnetzwerk identifiziert. Die einzelnen Prozessierungsschritte während der Signalweiterleitung wurden bis zu den regulierten Genen verfolgt. Dadurch werden alle beteiligten Schritte von der Perzeption der Signale bis zu den, durch die regulierten Gene, induzierten Effekten betrachtet.

Synergien wurden im GlobCell-Projekt exemplarisch in plasmacytoiden dendritischen Zellen, einem Zelltyp im Immunsystem des Menschen, untersucht. Diese Zellen verbinden angeborenes und erworbenes Immunsystem miteinander, wobei das angeborene Immunsystem das erworbene Immunsystem aktiviert und zum Teil auch reguliert. Plasmacytoide dendritische Zellen stellen eine Schlüsselstelle im Immunsystem dar, die auf eine Vielzahl verschiedener Pathogensignale reagiert. Die hier eingesetzten Stimuli, Interleukin-3 und das Influenza A-Virus, interagieren mit den plasmacytoiden dendritischen Zellen und repräsentieren zwei „Krankheitszustände“ des Organismus. Während das Influenza-A-Virus direkt die angeborene Immunantwort auslöst und eine neue Infektion darstellt,

---

repräsentiert Interleukin-3 modellhaft eine bereits bestehende Erkrankung, die auf das erworbene Immunsystem einwirkt. Durch den Netzwerkansatz wurde versucht zu verstehen, wie zwei Pathogene die Signalweiterleitung modifizieren. Es wurde untersucht, wie die beiden Signale miteinander interagieren und wie die Zelle diese Signale integriert, interpretiert bzw. prozessiert und inwieweit das Immunsystem dadurch reguliert wird.

Die Synergien und ihre regulierten Gene werden in dieser Arbeit zum einen durch Auswertung eines durch den französischen Projektpartner zur Verfügung gestellten experimentellen Datensatzes (Abschnitt 4.1) und zum anderen durch die hier neu entwickelte systembiologische Methode (Abschnitt 4.3) bestimmt. Durch ein, im Rahmen der Arbeit, rekonstruiertes Signaltransduktionsnetzwerk für plasmacytoide dendritische Zellen (Abschnitt 4.2) werden mithilfe systembiologischer Methoden die Orte der interagierenden Signalpfade von Interleukin-3 und des Influenza-A-Virus identifiziert. Außerdem werden die durch die weitere Prozessierung regulierten Gene und ihre Effekte bestimmt. Im letzten Teil der Arbeit (Abschnitt 4.4) werden die beiden Methoden miteinander verglichen. Dabei werden prinzipielle Gesichtspunkte von experimenteller und algorithmischer Methode beurteilt. Die Ergebnisse werden abschließend bezüglich ihrer Relevanz in der Immunantwort eingeordnet.





## 2 Grundlagen

Der erste Abschnitt des Grundlagenkapitels beschäftigt sich mit Kommunikation in mehrzelligen Organismen im Allgemeinen. Dabei wird auf grundlegende Prinzipien eingegangen, wie Signale in einem Organismus ausgetauscht werden und welche verschiedenen Formen auftreten. In Unterabschnitt 2.1.2 wird die intrazelluläre Signaltransduktion betrachtet, da externe Signale nur hier verarbeitet werden können. Für die ganzheitliche systembiologische Betrachtung von Signaltransduktion ist es hilfreich, Signaltransduktion als Netzwerk zu begreifen, worauf in Unterabschnitt 2.1.3 eingegangen wird. Von besonderer Relevanz in dieser Arbeit ist jedoch die Signalinterferenz, welche beschreibt, wie verschiedene Signale innerhalb einer Zelle miteinander interagieren, was in 2.1.4 dargestellt wird.

Im Immunsystem von Säugern spielt Signaltransduktion und Interferenz zwischen verschiedenen Immunsignalen eine besondere Rolle. Diese Arbeit beschäftigt sich mit der Identifikation von Synergien und ihrer Effekte während der Signaltransduktion, weshalb in Abschnitt 2.2 die experimentelle Basis der Arbeit vorgestellt wird. Neben einer kurzen Einführung in das Immunsystem des Menschen (2.2.1) werden die wesentlichen Charakteristika der eingesetzten plasmacytoiden dendritischen Zellen (siehe Unterabschnitt 2.2.2) und der Signale Interleukin-3 (Unterabschnitt 2.2.3) und des Influenza-A-Virus betrachtet (Unterabschnitt 2.2.4).

## 2.1 Kommunikation in mehrzelligen Organismen

### 2.1.1 Informationsaustausch zwischen Zellen

Die Zelle ist der universelle Grundbaustein des Lebens, sie trennt über eine Phasengrenze (Zellmembran, Zellwand) den intrazellulären vom extrazellulären Raum (Alberts *et al.*, 2008). Dadurch herrscht innerhalb der Zelle ein anderes Mikromilieu als außerhalb. Das Erbgut, welches Informationen über Aufbau und Struktur der Zelle beinhaltet, liegt innerhalb der Zelle vor und wird durch ein doppelsträngiges helikales Desoxyribonukleinsäure-Molekül (DNA) kodiert (Watson und Crick, 1953; Alberts *et al.*, 2008).

Vor etwa 1,5 Milliarden Jahren entwickelten sich die ersten Eukaryoten, in denen das Erbgut erstmals durch einen Zellkern geschützt wird (Martin *et al.*, 2002; Martin und Martin-Granel, 2006; Alberts *et al.*, 2008). Aus den frühen Eukaryoten entstanden vor etwa 800 Millionen Jahren die ersten mehrzelligen Organismen, aus denen sich durch Evolution und Selektion die heute vorherrschende Fauna und Flora entwickelte. Durch Spezialisierung von Zellen sowie durch Arbeitsteilung entstanden verschiedene Gewebe und Organe, die jeweils spezifische Funktionen im Organismus erfüllen. Beispielsweise sind die Lungen der Säugetiere auf die Aufnahme von Sauerstoff spezialisiert. Das Blut transportiert den an Hämoglobin gebundenen Sauerstoff zu anderen Zellen, wo er für Zellatmung und zum Verrichten von Arbeit benötigt wird (Campbell *et al.*, 2009; Hall, 2011). Neben dem Austausch von Nährstoffen bedingt die Spezialisierung der Zellen auch den Austausch von Informationen. Dadurch übermitteln Zellen ihre Bedürfnisse an den Gesamtorganismus und die Kommunikation zwischen Zellen im Organismus und der Umgebung gewährleistet das Überleben.

Bei der Kommunikation der Zellen untereinander unterscheidet man verschiedene Kommunikationsformen. Bei juxtakriner Kommunikation werden über direkten Zellkontakt Informationen ausgetauscht (siehe Abb. 2.1a). Die Informationsübermittlung ist lokal wie zeitlich begrenzt und kommt von Prokaryota bis hin zum Menschen vor, bei letzterem beispielsweise in der Entwicklung neuronaler Zellen (Hitoshi *et al.*, 2002; Aguirre, Rubio und Gallo, 2010).

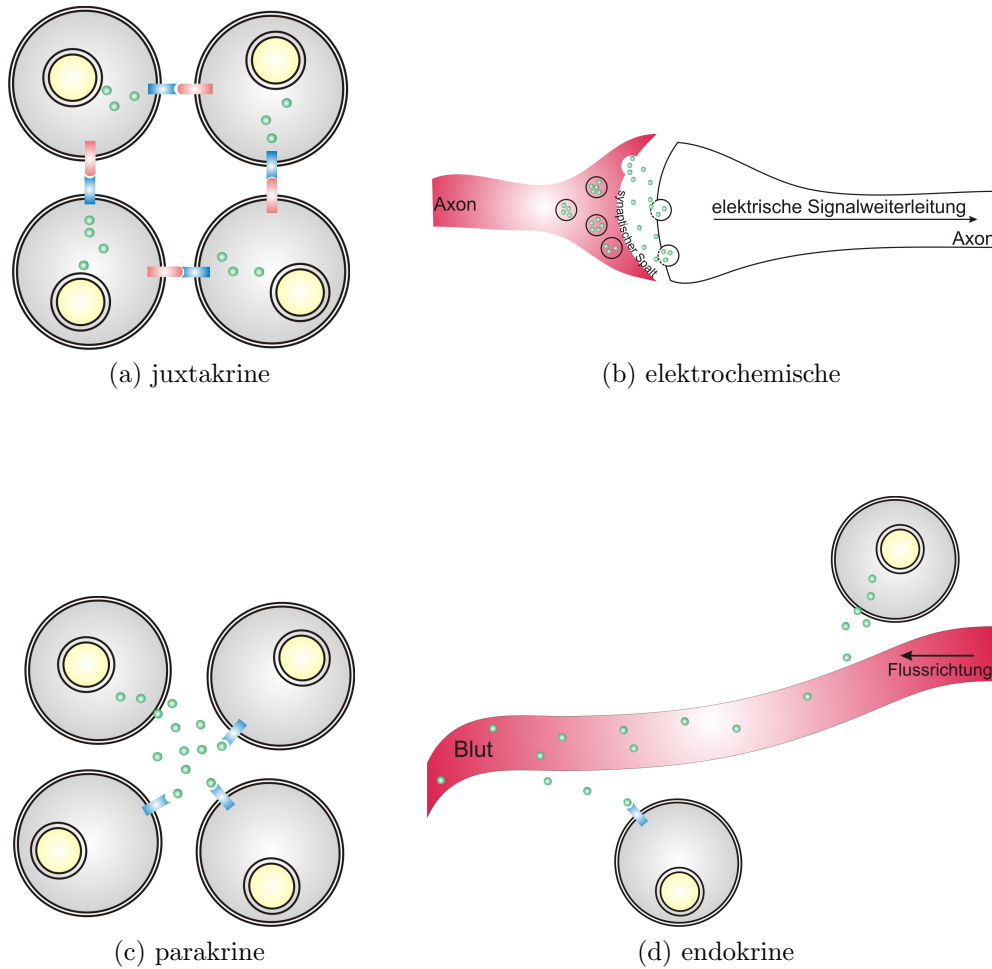


Abbildung 2.1: Signalaustausch zwischen Zellen. (a) und (c) zeigen zwei lokal wirkende Kommunikationsformen zwischen Zellen. Bei juxtakriner Kommunikation (a) löst direkter Zell-Zell-Kontakt (rot-blauer Kontakt) die Signalkaskade aus, während bei parakriner Kommunikation (c) ein Signalmolekül (grün) in den extrazellulären Raum freigesetzt wird, welches von umgebenden Zellen aufgenommen wird. Abbildung (b) und (d) zeigen systemische Kommunikationsformen, welche über Entfernungen wirken. Bei der elektrochemischen Kommunikation (b) wird ein Signal elektrisch durch Nervenzellaxone weitergeleitet. Die Synapse verbindet zwei Nervenzellen. Durch Austausch von Neurotransmittern zwischen den Zellen wird ein elektrischer Impuls ausgelöst. Bei der endokrinen Kommunikation (d) wird ein Signal (z.B. Hormon) in Blut oder Lymphe sekretiert, wodurch es sich im Körper verteilt und von Zielzellen über Rezeptoren perzipiert wird.

Die schnellste Form, Signale über weite Strecken zu übertragen, ist die elektrochemische Signalweiterleitung (siehe Abb. 2.1b), wie sie beispielsweise in den Nervenbahnen von Säugern stattfindet. Über diesen phylogenetisch neuen Signalweg<sup>1</sup> werden besonders wichtige Signale transportiert, wie etwa Schmerzsignale. Für die Übertragung sind hochgradig spezialisierte Zellen (Neuronen) verantwortlich, die miteinander durch Synapsen verknüpft sind. An den Synapsen findet die Weiterleitung an benachbarte Zellen durch Freisetzung von Neurotransmittern in den synaptischen Spalt statt. Die Aufnahme durch die angrenzende Zelle löst, anhand von Aktionspotentialen, einen elektrischen Impuls aus, welcher die Signale zum Gehirn oder zum Rückenmark leitet, wo sie verarbeitet werden (Campbell *et al.*, 2009). Reflexe wie der Patellasehnenreflex oder der Lidschlussreflex sind Beispiele für elektrochemische Signalweiterleitung (Wehner und Gehring, 1995).

Eine Anpassung über längere Zeiträume oder an sich langsam ändernde Umweltgrößen, wie die organismische Anpassung an die Jahreszeiten, wird durch para- und endokrine Kommunikation ermöglicht. Beides führt zu einer langsameren aber persistenten Anpassung, wobei chemische Signale als Informationsträger fungieren (Campbell *et al.*, 2009). Signale werden von den Zellen ins extrazelluläre Medium sekretiert und verteilen sich im Organismus (Alberts *et al.*, 2008). Parakrine Signale beeinflussen die Zellen in der Umgebung (z.B. im selben Organ oder Zellverband, siehe Abb. 2.1c), wogegen endokrine Signale auf entfernte Ziele und andere Organe einwirken (Abb. 2.1d) (Alberts *et al.*, 2008; Campbell *et al.*, 2009). Endokrine Signale sind beispielsweise Hormone, welche durch Blut oder Lymphe verteilt werden. Beide Kommunikationsformen sind in der Immunantwort relevant, wo spezialisierte Zellen Cytokine, Hormone oder andere Signale ausschütten, welche entweder ein Pathogen angreifen oder die Infektionsstelle markieren und weitere Zellen zum Infektionsort rekrutieren (Arai *et al.*, 1990; Viau und Zouali, 2005; Alberts *et al.*, 2008; Iwasaki und Medzhitov, 2010). Ein Beispiel ist die Aktivierung des erworbenen (adaptiven) Immunsystems (Janeway *et al.*, 2001).

---

<sup>1</sup>Dieser Signalweg existiert seit zirka 600 Millionen Jahren und tritt erstmals in einigen Stämmen der Eumetazoa auf (z.B. bei den Bilateria) (Wehner und Gehring, 1995; Nielsen, 2012).

### 2.1.2 Signaltransduktion in Zellen

Kommunikation zwischen Zellen bedingt auch die Aufnahme von Signalen aus der Umgebung sowie ihre Verarbeitung. Dabei existieren zwei wesentliche Probleme. Zum einen können verschiedene Zelltypen unterschiedlich auf ein Signal reagieren, beispielsweise dehnen sich Herzmuskelzellen bei Perzeption von Acetylcholin, einem Neurotransmitter, aus, während Skelettmuskelzellen kontrahieren<sup>2</sup> (Alberts *et al.*, 2008). Zum anderen müssen die extrazellulären Signale für die Initiierung der intrazellulären Signalverarbeitung die Zellbegrenzung (Plasmamembran, Zellwand) überwinden.

Kleine Moleküle oder Ionen wie  $K^+$  oder  $Ca^{2+}$  werden passiv durch Diffusion oder aktiv durch Ionenkanäle ins Innere der Zelle transportiert (Nelson und Cox, 2001; Alberts *et al.*, 2008). Größere Moleküle, die aufgrund ihrer chemischen Eigenschaften die Zellbarriere überwinden können, werden durch intrazelluläre Rezeptoren perzipiert, wie z.B. das Influenza-A-Virus, welches durch den *toll-like* Rezeptor 7 (TLR7) detektiert wird (siehe Unterabschnitt 2.2.4, Finberg, Wang und Kurt-Jones, 2007). Im Allgemeinen ist dieser Transportweg für größere Moleküle jedoch nicht möglich. Diese extrazellulären Signale werden durch spezifische Rezeptoren über ein Schlüssel-Schloss-Prinzip an der Zelloberfläche perzipiert. Nur selten werden die Signale durch Endozytose<sup>3</sup> in die Zelle transferiert, meistens kommt es zu reversiblen Ligand<sup>4</sup>-Rezeptor-Interaktionen<sup>5</sup> (Alberts *et al.*, 2008; Gomperts, Kramer und Tatham, 2009).

Die Interaktionen mit dem Rezeptor sind entweder agonistisch (positiv) oder antagonistisch (negativ). Bei einer agonistischen Ligand-Rezeptor-Interaktionen ändert sich die räumliche Struktur des Rezeptors auf der intrazellulären Seite (siehe Abb. 2.2). Dadurch können direkt chemische Reaktionen ausgelöst werden oder es werden Domänen<sup>6</sup> verfügbar, an denen Adaptoren, Modifikatoren oder

---

<sup>2</sup>Im Beispiel liegen unterschiedliche Rezeptoren vor, was jedoch nicht zwangsweise der Fall ist (Alberts *et al.*, 2008).

<sup>3</sup>Endozytose beschreibt einen Transport, bei dem extrazelluläre Signale durch eine Membran umschlossen und anschließend durch die Zellmembran transferiert werden.

<sup>4</sup>Mit Liganden werden Proteine und Moleküle bezeichnet, die an einen Rezeptor binden.

<sup>5</sup>Die Liganden binden zumeist nicht kovalent (dauerhaft) an den Rezeptor und bilden einen Ligand-Rezeptor-Komplex aus.

<sup>6</sup>Bereiche der Aminosäurekette der Rezeptorproteine, die auf der intrazellulären Seite mit anderen Proteinen oder Molekülen interagieren.

Kinasen<sup>7</sup> binden. Diese rekrutieren entweder weitere Proteine<sup>8</sup> oder übernehmen selbst die Rolle als Signalträger. Ein extrazelluläres Signal wird dadurch auf einen intrazellulär vorliegenden Signalträger übertragen, wodurch die Signaltransduktion initiiert wird (siehe Abb. 2.2).

Das Gegenteil zu agonistischen Ligand-Rezeptor-Interaktionen sind Antagonistische. Diese blockieren die Signalweiterleitung oder schränken sie zumindest ein (Smith *et al.*, 2003; Alberts *et al.*, 2008). Antagonistische Effekte sind in Medizin und Pharmazie von besonderem Interesse, da sie beispielsweise Rezeptoren oder Enzyme deaktivieren und somit krankheitsspezifische Signalkaskaden unterbrechen. Ein Beispiel hierfür sind MAO-Hemmer<sup>9</sup>,  $\beta$ -Blocker<sup>10</sup> oder gängige Schmerzmittel wie Ibuprofen<sup>11</sup>, welche Rezeptorbindestellen oder Enzyme blockieren und dadurch die Signalweiterleitung bzw. die Rezeptoraktivierung verhindern (Lüllmann, Mohr und Hein, 2010; Herdegen *et al.*, 2010).

### Signalweiterleitung durch konsekutiv katalytische Reaktionen

Ein Signalträger katalysiert weitere chemische Reaktionen, wodurch das Signal sukzessive übertragen wird. Eine Reaktion ermöglicht so die Folgende usw.. Die aktivierten chemischen Reaktionen und Prozesse können jedoch auch mehrfach aktiviert werden, wodurch die Signalstärke moduliert wird. Da die Produkte einer katalysierten Reaktion, in der Regel aktivierte Proteine, als Katalysator für weitere Reaktionen dienen, entstehen Signalfade entlang derer die zu übermittelnde Information weitergegeben wird (siehe Abb. 2.2).

Die Stärke eines Signalfades, also die Möglichkeit eine Information entlang des Signalweges weiterzugeben, wird durch verschiedene Parameter definiert: ers-

---

<sup>7</sup>Viele Enzymrezeptoren verfügen über eine intrinsische Kinasedomäne, dies bedeutet, dass die Aktivierung der Rezeptoren ihre Struktur derart verändert, sodass sie direkt spezifische Phosphorylierungsreaktionen katalysieren. Bei Rezeptoren, bei denen diese Funktionalität fehlt, werden weitere intrazellulär vorliegende Kinasen gebunden, welche die katalytischen Aufgaben übernehmen.

<sup>8</sup>Ein Protein ist eine Kette von Aminosäuren, welche durch Translation und Prozessierung aus der in der DNA-kodierten Information produziert wird.

<sup>9</sup>Moleküle, die enzymatisch die Monoaminoxidasen (MAO) hemmen, wodurch verschiedene Hormone und Neurotransmitter verlangsamt abgebaut werden.

<sup>10</sup> $\beta$ -Blocker interagieren mit den  $\beta$ -Adrenorezeptoren und vermindern so die Bindung von Adrenalin oder Noradrenalin und somit die Rezeptoraktivität.

<sup>11</sup>Ibuprofen wirkt enzymatisch und unspezifisch inhibierend auf Cyclooxygenasen (COX-1 und COX-2) ein, wodurch eine Entzündungsreaktion vermindert wird.

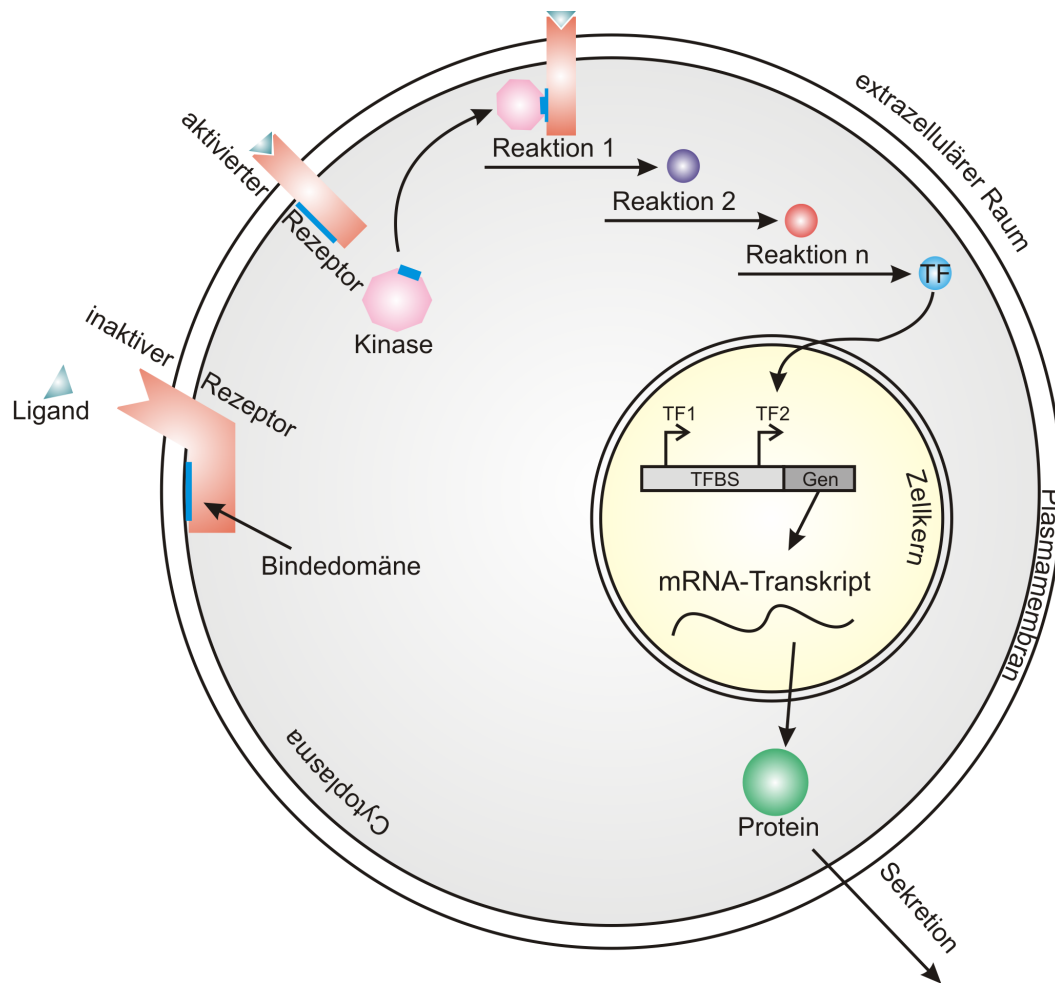


Abbildung 2.2: Schematischer Ablauf von enzymrezeptorabhängiger Signaltransduktion. Ein Ligand bindet extrazellulär an einen Rezeptor, der dadurch aktiviert wird, seine Konformation ändert und eine Bindedomäne freilegt. An der nun freien intrazellulären Domäne bindet eine Kinase, welche eine Phosphorylierungsreaktion katalysiert. Das Produkt dieser Reaktion katalysiert die nächste Reaktion usw. bis ein Transkriptionsfaktor (TF) aktiviert wird. Dieser wird in den Zellkern transferiert und bindet an geeignete Transkriptionsfaktorbindestellen (TFBS), wo er genregulatorische Effekte auslöst. Die betreffenden Gene werden in mRNA transkribiert und schließlich in Proteine translatiert, die entweder intrazelluläre Aufgaben erfüllen oder als neues Signal in den extrazellulären Raum sekretiert werden.

tens von der Stabilität des Ligand-Rezeptor-Komplexes, welcher in seiner aktiven Form bestimmt, wie oft ein Signalpfad ausgelöst wird. Je stabiler dieser Komplex ist, desto häufiger wird die initiale Phosphorylierungsreaktion ausgelöst. Zweitens von der Lebensdauer der jeweils aktivierten Katalysatoren, dadurch wird determiniert, wie gut eine Information von einem Schritt der Signalweiterleitung an den nächsten weitergegeben wird. Die Anzahl der um ein begrenztes Edukt konkurrierenden Reaktionen ist ein dritter Faktor, welcher die einzelnen Schritte während der Signalweiterleitung limitiert und reguliert.

### **Transkriptionsfaktoren regulieren genomische Programme**

Die Ziele von aktivierten Signalpfaden sind oft Transkriptionsfaktoren (TFs), welche im Cytoplasma oder im Zellkern vorliegen (siehe Abb. 2.2). Der Transfer von Proteinen, also auch TFs, vom Cytosol (intrazellulärer Raum) in den Zellkern erfolgt durch Importine, welche entsprechend markierte TFs durch die Zellkernporen schleusen (Kalderon *et al.*, 1984; Izaurralde und Adam, 1998). Die TFs binden an spezifische Sequenzen der DNA, die Transkriptionsfaktorbindestellen (TFBS), welche in den regulatorischen Bereichen der Gene liegen. Durch Bindung verschiedener TFs an ihre TFBS wird die Genexpression grundsätzlich ermöglicht, aber auch feinjustiert. Dadurch kann die Zelle adäquat auf alle einwirkenden Faktoren reagieren (Wingender, 1993; Alberts *et al.*, 2008; Campbell *et al.*, 2009). Im Allgemeinen endet die Signaltransduktion an dieser Stelle, jedoch wird das Transkript, die Kopie eines Gens, in Protein umgewandelt, wodurch die Reaktion der Zelle ausgelöst wird. Dies kann eine interne Anpassung oder die Ausschüttung weiterer Signale in den extrazellulären Raum sein. Abbildung 2.2 zeigt ein vereinfachtes Schema der Signaltransduktion.

### **2.1.3 Netzwerke ermöglichen einen systemischen Blick auf Signaltransduktion**

Während der Weiterleitung eines Signals vom Rezeptor zu den regulierten Genen durchquert es verschiedene Ebenen der Zelle (siehe Abb. 2.2). Schon die intrazelluläre Weiterleitung eines Signals durch katalysierte Reaktionen und Protein-Protein-Interaktionen entlang von Signalkaskaden bietet eine große Vielfalt an



Möglichkeiten, wie ein Signal in der Zelle weitergeleitet werden kann. Um Signaltransduktion jedoch ganzheitlich zu verstehen, müssen die vielfältigen Interaktionen verschiedener Signale untereinander sowie die Interaktionen mit ihren Rezeptoren und die Regulation der verschiedenen Gene betrachtet werden. Dies bildet ein kompliziertes und feingliedriges Geflecht an Interaktionen, welches als Netzwerk, bzw. mathematisch als Graph, aufgefasst wird. In der Biologie helfen Netzwerke dabei, die Verbindungen zwischen verschiedenen Komponenten eines Systems darzustellen, wodurch diese strukturiert und gezielt graphentheoretisch analysiert werden können (Mason und Verwoerd, 2007; Bortfeldt, Schuster und Koch, 2010). Besonders bei einer unüberschaubaren Anzahl an Komponenten und Interaktionen, wie bei der Signaltransduktion, helfen Netzwerke bei der Erlangung von Kenntnissen über generelle Struktur sowie Funktion oder Wichtigkeit einzelner Komponenten (Barabási, 2002; Kriete *et al.*, 2006; Palsson, 2006). Dadurch erzeugen Netzwerke ein besseres Verständnis für die Funktionsweise der gesamten Zelle und der in ihr stattfindenden Signaltransduktion (Palsson, 2006; Kriete *et al.*, 2006).

Neben Signaltransduktion werden Netzwerke beispielsweise auch für die Modellierung und Analyse von Genregulation und Metabolismus eingesetzt (Kriete *et al.*, 2006). Bei ersteren will man wissen, wie verschiedene Gene interagieren und welche Effekte ausgelöst werden. Dadurch lassen sich evolutionäre Zusammenhänge bestimmen oder es werden Aussagen über Zellentwicklung und Regulation verschiedener Gene abgeleitet (Babu und Teichmann, 2003; Teichmann und Babu, 2004; Medina *et al.*, 2004; Barabási und Oltvai, 2004; Brasier, 2006; Erwin und Davidson, 2009). Bei metabolischen Netzwerken<sup>12</sup> stehen Metabolite, ihre Interaktionen und Folgen im Zentrum des Interesses. All diese Netzwerke und ihre Interaktionen sind unter anderem in der Medizin von Interesse, wenn Ursachen für Krankheiten oder ihre Entwicklung aufgeklärt werden.

Zentrale Elemente eines Netzwerks sind von besonderem Interesse, da sie starken Einfluss auf Struktur und Zusammenhalt des Netzwerks besitzen. Die Identifikation solcher Elemente und ihre Bewertung stellt eine Hauptaufgabe bei der Untersuchung von Netzwerken dar (Zhao *et al.*, 2006; Potapov, Goemann und Wingender, 2008; Opsahl, Agneessens und Skvoretz, 2010). Bei genregulatori-

---

<sup>12</sup>Metabolische Netzwerke stellen die Beziehungen von Metaboliten untereinander dar, Metabolite sind sämtliche am Stoffwechsel beteiligte Elemente.

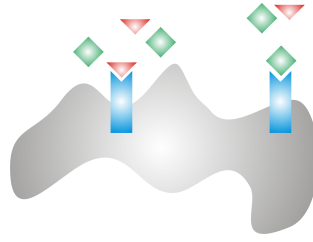


Abbildung 2.3: Schema einer antagonistischen Interaktion. Zwei verschiedene Signalmoleküle (Dreiecke, Vierecke) konkurrieren um die Bindung an einer Proteindomäne (blau). (Zum besseren Verständnis wurde die Bindedomäne doppelt dargestellt.)

schen Netzwerken sind Schlüsselkomponenten interessant, da sie Gene darstellen, die besonders wichtige Rollen in der Regulation biologischer Prozesse, wie Immunantwort, Kommunikation, Zellentwicklung etc., einnehmen (Groot, Coffer und Koenderman, 1998; Gaur und Aggarwal, 2003; Li, Assmann und Albert, 2006; Lan, Selmi und Gershwin, 2008; Alberts *et al.*, 2008). Zur Detektion von Schlüsselkomponenten werden topologische, das heißt sich aus der Struktur ergebende, Parameter genutzt, die das jeweilige Netzwerk charakterisieren und beispielsweise zentrale Elemente bestimmen. Dadurch werden die Beziehungen zwischen Genen, ihren Funktionen und ihrem resultierenden Phänotyp untersucht (Barabási, 2002; Barabási und Oltvai, 2004; Palsson, 2006). In den Signaltransduktionsnetzwerken dieser Arbeit wird untersucht, wie und wo verschiedene Signale miteinander interagieren.

#### 2.1.4 Interferenz und Synergie zwischen verschiedenen Signalpfaden

Interaktionen verschiedener Signalpfade treten an allen Punkten der Signaltransduktion auf, also an Rezeptoren, während der Signalweiterleitung und auch auf der genomischen Ebene. Diese Interaktionen unterteilen sich wie die Signal-Rezeptor-Interaktion (siehe Unterabschnitt 2.1.2) in agonistisch, antagonistisch sowie synergistisch.

Durch Konkurrenzbeziehungen (antagonistisch) wird eine negative Kontrolle auf Signalpfade ausgeübt. Diese werden durch Binden eines nicht-spezifischen Partners deaktiviert oder partiell inhibiert. Bei Rezeptoren verhindert die Bindung

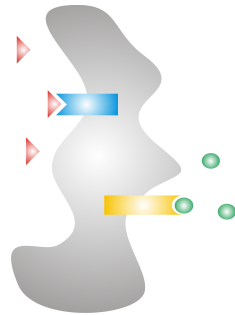


Abbildung 2.4: Schema einer agonistischen Interaktion. Zwei verschiedene Signalmoleküle (Dreiecke, Kugeln) verstärken die ausgelöste Wirkung gegenseitig durch die Bindung an ihren jeweiligen Domänen (blau, gelb). Die abgebildeten Interaktionen können alternativ auch als nicht-kompetitiv antagonistisch interpretiert werden. Dabei wird durch Bindung der Signalmoleküle an unterschiedliche Proteindomänen die Effizienz des ausgelösten Signalpfades reduziert.

eines unspezifischen Liganden die Aktivierung der Signalkaskade. Ein Beispiel für antagonistische Interaktionen, unabhängig von Signaltransduktion, ist die kompetitive Bindung von Kohlenstoffmonooxid (CO) an Hämoglobin. Durch die höhere Affinität von CO zu Hämoglobin wird der Sauerstoff verdrängt (siehe dazu auch Abb. 2.3) (Hall, 2011). Beim Zusammenspiel zweier Signalpfade kommt es zu inhibitorischen Ereignissen, wenn beispielsweise eine Kinase durch einen Prozess deaktiviert wird und der originäre Signalweiterleitungsprozess nicht mehr initiiert wird. Konkurrenzsituationen innerhalb einer Zelle sind eine weitere Möglichkeit inhibitorischer Effekte (Choi *et al.*, 2004b). Dabei greifen verschiedene Prozesse auf ein gemeinsames Edukt zu, welches nur in beschränkter Menge zur Verfügung steht. Es kommt es zur gegenseitigen Hemmung der chemischen Reaktionen, wobei jene im Vorteil ist, deren Reaktionspartner zum Edukt die höhere Affinität besitzen. Neben kompetitiven antagonistischen Interaktionen treten auch nicht-kompetitive Situationen auf. Beispielsweise wird ein Protein an verschiedenen Stellen modifiziert, wodurch sich die Konformation ändert, sodass die Signalkaskade unterdrückt oder negativ reguliert wird (siehe Abb. 2.4).

Bei agonistischen Signalpfadinteraktionen wird ein Signal verstärkt oder ausgelöst. Abbildung 2.4 zeigt, wie unterschiedliche Signale (Dreiecke, Kugeln) an unterschiedliche Bindedomänen eines Proteins andocken. Dabei wird durch positive Beeinflussung das Signal verstärkt. Ein Signal (z.B. Dreiecke) wird auch

ohne Anwesenheit des zweiten Signalmoleküls weitergeleitet, jedoch ist die Weiterleitung unter Anwesenheit des zweiten Signals (Kugeln) effizienter. Parallel ablaufende Prozesse können sich so gegenseitig verstärken (agonistische Interaktionen). Dabei wirken zwei oder mehr Signalfade auf ein Protein ein, wodurch es an verschiedenen Stellen modifiziert (z.B. phosphoryliert) wird. Dadurch können beispielsweise Adaptoren, die weitere Kinasedomänen zur Verfügung stellen oder die Effizienz der Kinase erhöhen, gebunden werden. Die Möglichkeit, dass Proteine an verschiedenen Stellen modifiziert werden oder sich durch Binden von Adaptoren etc. zu Komplexen zusammenschließen, liefert ein unüberschaubares Spektrum möglicher Interaktionen und zeigt die Komplexität der Signaltransduktion.

Ein Beispiel für eine agonistische Interaktion ist die Gabe eines Acetylcholin-Agonisten (beispielsweise Acetylcarnitine) bei Alzheimer oder Parkinson (Spagnoli *et al.*, 1991). Dies erhöht die Menge des Neurotransmitters Acetylcholin bzw. von Acetylcholin-analogen im synaptischen Spalt, wodurch Nervenzellen die Signale schneller weiterleiten.

### **Synergistische Interaktion von Signalpfaden**

Neben agonistischer und antagonistischer Interaktion und ihren Spielarten existieren noch synergistische Interaktionen. Das Oxford Dictionary of Biochemistry and Molecular Biology beschreibt Synergismus (Synergie) als „*Phänomen, bei dem der gemeinsame Effekt, welcher von zwei oder mehr Agonisten ausgelöst wird, größer ist als die Summe der Effekte, welche jeder einzelne (Agonist) auslöst.*“ (Smith *et al.*, 2003). Deshalb stellen synergistische Interaktionen das Gegenteil zu Antagonistischen dar, da implizit von einer positiven Beeinflussung ausgegangen wird (Smith *et al.*, 2003).

In der biologischen Literatur wird Synergie quantitativ begriffen (Gouwy *et al.*, 2004; Shelton *et al.*, 2004; Anastassiou, 2007; Doreau *et al.*, 2009). Dies bedeutet, dass synergistische Effekte messbar sind. In der systembiologischen Sichtweise, welche die gesamte Zelle gleichzeitig betrachtet, wird Synergie somit anhand von Proteinabundanz oder Genexpressionsraten detektiert. Dabei ist die Bestimmung der Genexpression jedoch zuverlässiger als die Messung der Proteinabundanz

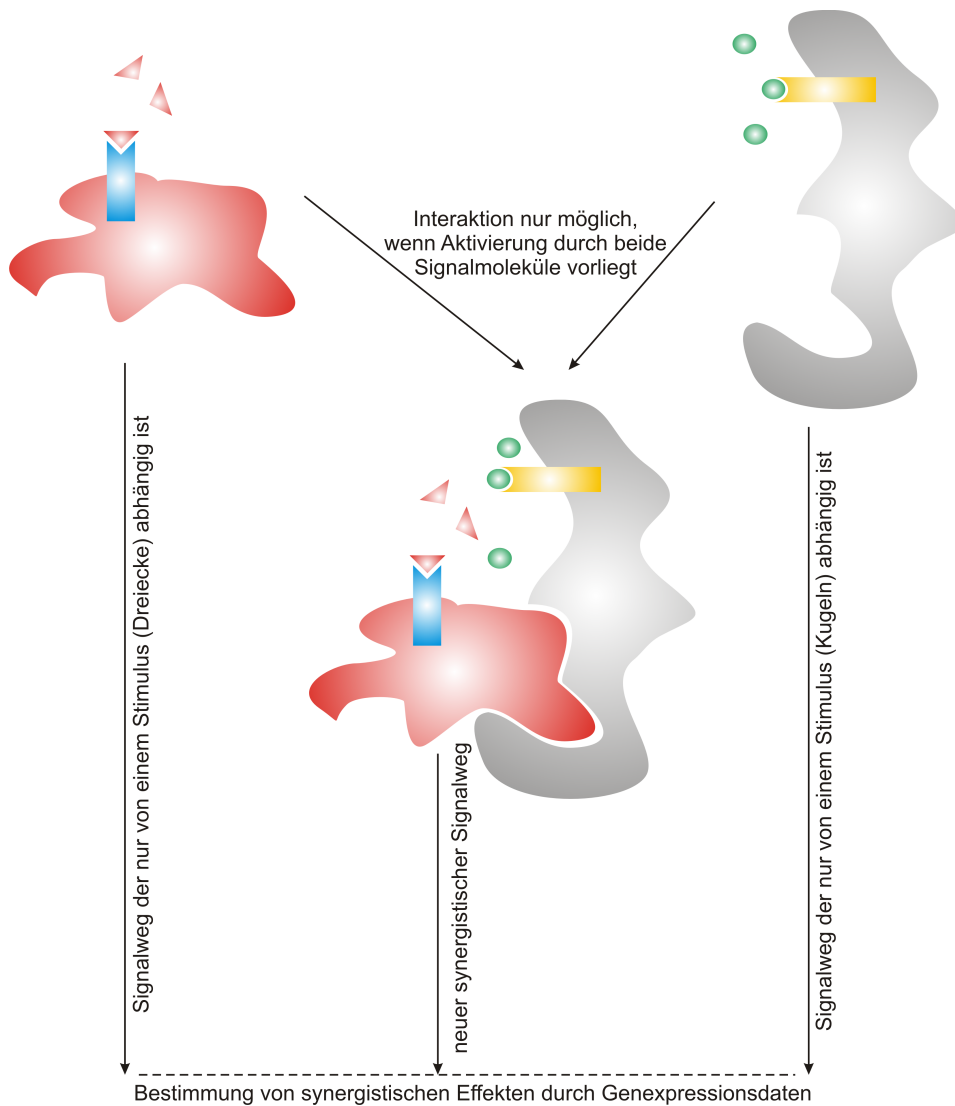


Abbildung 2.5: Schema einer synergistischen Signalpfadinteraktion. Jedes einzelne Protein (rot, grau) löst signalspezifische Interaktionen aus. Sind beide Signale vorhanden, interagieren die Proteine miteinander und aktivieren einen neuen synergistischen Signalpfad. Die ausgelösten synergistischen Effekte werden, unabhängig vom Ort des Auftretens, experimentell, beispielsweise durch Messung der Genexpression, bestimmt.

in der Zelle. Es ist zwar zu erwarten, dass in Zukunft auch Proteine mit einem Hochdurchsatzverfahren, ähnlich zur Bestimmung der Genexpression durch Microarrays (siehe Abschnitt 3.4), gemessen werden können. Bislang sind diese Verfahren, wie beispielsweise die massenspektrometrische Bestimmung des kompletten Proteoms einer Zelle mit den jeweils spezifischen Modifikationen (z.B. Phosphorylierungen) an einem Protein, nicht ausgereift genug (Cravatt, Simon und Yates, 2007; Cox und Mann, 2007; Han, Aslanian und Yates, 2008; Nilsson *et al.*, 2010).

Die Betrachtung von synergistischen Effekten alleine auf der Grundlage von Genexpressionsdaten liefert jedoch weder Informationen über die Ursachen der gemessenen Synergien noch über die Prozessierung synergistischer Signale während der Signalweiterleitung. Die in Abbildung 2.5 dargestellte Komplexierung findet erst nach Aktivierung der beiden unabhängigen Signalpfade statt. Wie das synergistische Signal weitergeleitet wird, ist nicht bestimmbar, ebenso wie nicht garantiert werden kann, dass auf Genexpressionsebene durch die dargestellte Interaktion ein positiver, also agonistischer, Effekt ausgelöst wird.

Um Synergien systemweit zu verstehen, müssen deshalb alle Ebenen der Zelle betrachtet werden und die ausgelösten Signalpfade verfolgbar sein. Für die Bestimmung der Ursachen von Synergien ist eine rein quantitative Messung unzureichend, da sie weder Informationen über die Funktionsprinzipien bereithält noch darüber, an welchen Orten (Proteinen) Synergien auftreten. Die Aufklärung dieser Fragen und die Bestimmung von Synergie im gesamten Signaltransduktionsnetzwerk ist eine wichtige Aufgabe im Rahmen dieser Arbeit.

## 2.2 Der Einfluss verschiedener Signale auf plasmacytoide dendritische Zellen im menschlichen Immunsystem

### 2.2.1 Angeborenes und adaptives Immunsystem des Menschen

Das Immunsystem dient der Bekämpfung von Infektionen, welche durch Pathogene ausgelöst werden. Dabei unterscheidet es körperfremde von körpereigenen Substanzen und Lebewesen, wie beispielsweise Viren, Bakterien, Pilze und andere krankheitsauslösende eukaryotische Organismen (Janeway *et al.*, 2001; Alberts *et al.*, 2008; Campbell *et al.*, 2009). Diese dringen als Pathogene in den Organismus ein, wodurch sie eine Immunantwort auslösen. Das Immunsystem besteht aus zwei wesentlichen Komponenten, dem angeborenen (innaten) und dem erworbenen (adaptiven) Immunsystem.

Das angeborene Immunsystem existiert in allen rezenten Tieren und Pflanzen und ist phylogenetisch sehr alt (Campbell *et al.*, 2009). Die Detektion der Pathogene durch spezialisierte Immunzellen basiert auf der Perzeption allgemeiner Oberflächeneigenschaften die verschiedene Pathogene teilen. So werden beispielsweise Einzelstrang-RNA-Viren (ssRNA-Viren) detektiert, indem durch den intrazellulär vorliegenden *toll-like*-Rezeptor (TLR7) GU-reiche Regionen<sup>13</sup> der RNA<sup>14</sup> detektiert werden (Cao und Liu, 2007). Bakterien-DNA oder DNA-Viren werden durch andere *toll-like*-Rezeptoren anhand von CpG-Regionen detektiert. Dies sind Wiederholungen an linearen Sequenzen von Cytidin, Guanodin-Paaren in DNA, die bei Bakterien häufiger unmethyliert vorliegen als beispielsweise in tierischen Zellen (Janeway *et al.*, 2001; Lenert, 2006). Das angeborene Immunsystem reagiert auf perzipierte Pathogene sofort und unspezifisch. Die als körperfremd erkannten Pathogene werden markiert und durch spezialisierte Zellen, in einem Vorgang der Phagozytose genannt wird, verdaut (Janeway *et al.*, 2001). Erste

---

<sup>13</sup>In GU-reichen Regionen häuft sich das Auftreten der Nucleobasen Guanin und Uracil.

<sup>14</sup>Die RNA ist, wie die DNA, ein Makromolekül, welches im Allgemeinen nicht Träger der Erbinformation ist. Sie übernimmt Kommunikationsaufgaben zwischen genomischer und proteomischer Ebene, indem sie die Aminosäuresequenz der Proteine kodiert. Außerdem übernimmt sie genregulatorische Aufgaben.

Arbeiten zum angeborenen Immunsystem gehen auf Elie Metchnikoff und Paul Ehrlicher zurück und wurden 1908 mit dem Nobelpreis für Medizin ausgezeichnet (Metchnikoff, 1967).

Das erworbene Immunsystem existiert nur in Wirbeltieren wie dem Menschen und ist stammesgeschichtlich eine neuartige Entwicklung (Campbell *et al.*, 2009). Gegenüber dem angeborenen Immunsystem geht die Funktionsweise des erworbenen Immunsystems auf Interaktionen mit Antikörpern zurück, die spezifisch Pathogene erkennen. Diese Spezifität ist jedoch nicht direkt verfügbar, sondern wird während der im Allgemeinen 4-7 Tage andauernden Immunreaktion erlernt. Dabei werden durch spezialisierte Zellen, wie beispielsweise plasmacytoide dendritische Zellen, kleine Teile des verdauten Pathogens an der Oberfläche der Zellen präsentiert (McKenna, Beignon und Bhardwaj, 2005; Gilliet, Cao und Liu, 2008). B- und T-Zellen, welche wesentliche Bestandteile des erworbenen Immunsystems sind, detektieren diese Pathogenteile als Antigene und entwickeln darauf basierend spezifische Antikörper (Janeway *et al.*, 2001). Die entwickelten Antikörper bleiben oft ein Leben lang erhalten, wodurch die Pathogene bei zukünftigen Infektionen schneller erkannt, gezielt angegriffen und vernichtet werden. Im Laufe ihres Lebens erwerben Wirbeltiere dadurch eine spezifische Immunität gegen eine Vielzahl unterschiedlicher Pathogene. Gleichzeitig dient das erworbene Immunsystem so als *Gedächtnis* für bereits erfolgreich bekämpfte Infektionen (Janeway *et al.*, 2001). Erste grundlegende Arbeiten zum adaptiven Immunsystem wurden von Emil von Behring und Katsato Shibasaburo veröffentlicht und 1901 mit dem ersten Nobelpreis für Medizin prämiert (Behring und Wernicke, 1892; Behring, 1967; Janeway *et al.*, 2001).

### **2.2.2 Plasmacytoide dendritische Zellen verbinden angeborenes und adaptives Immunsystem**

Wie alle Immunzellen leiten sich plasmacytoide dendritische Zellen (pDC) aus der hämatopoietischen Stammzelllinie ab. Diese pluripotenten Stammzellen entstammen dem Knochenmark und differenzieren sich unter anderem in eine lymphoide und eine myeloische Vorläuferzelle, aus denen sich die weiteren Immunzellen (Leukozyten) ableiten (Janeway *et al.*, 2001). Vereinfacht wird davon ausgegangen, dass Lymphozyten, welche sich aus der lymphoiden Vorläuferzelle differen-



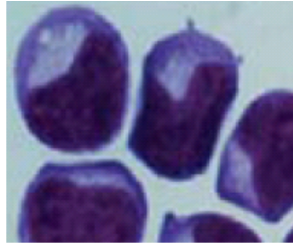


Abbildung 2.6: Mikroskopische Aufnahme einer eingefärbten, zu pDC morphologisch ähnlichen, Plasmazelle. Die Abbildung wurde verändert aus (Liu, 2005) entnommen.

zieren, in die adaptive Immunantwort involviert sind, wogegen Zelltypen, welche sich aus der myeloischen Vorläuferzelle ableiten, in angeborene und adaptive Immunantwort involviert sind.

Erstmalig wurden pDCs 1958 von K. Lennert und W. Remmele, als T-Zell-assoziierte, morphologisch ähnliche Plasmazellen (siehe Abb. 2.6) in Lymphknoten beschrieben, wodurch sie der lymphatischen Zelllinie zugeordnet wurden (Lennert und Remmele, 1958; Janeway *et al.*, 2001; Comeau *et al.*, 2002; Colonna, Trinchieri und Liu, 2004; Ghirelli, Zollinger und Soumelis, 2010). Ähnlich wie B-Zellen, ein ebenfalls lymphoider Zelltyp, präsentieren pDCs Teile der phagozytierten Pathogene an der Zelloberfläche (Antigenpräsentation), wodurch weitere Immunzellen aktiviert werden (Janeway *et al.*, 2001; Comeau *et al.*, 2002; Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007). Jedoch exprimieren pDC weder B-Zell-typische Antigenrezeptoren noch spezifische Immunglobuline oder einen T-Zell-Rezeptor, dafür aber verschiedene Zellmarker sowie MHC-II-Antigene<sup>15</sup>, welche das erworbene Immunsystem aktivieren, aber in Zellen der myeloischen Linie zu beobachten sind (Feller *et al.*, 1983; Facchetti *et al.*, 1988; Colonna, Trinchieri und Liu, 2004; McKenna, Beignon und Bhardwaj, 2005; Liu, 2005). Die Herkunft von pDCs ist daher ungeklärt, jedoch zeigten Grouard *et al.*, dass sich pDCs unter Zugabe von Interleukin-3 (IL-3) in dendritische Zellen (DC), ein Immunzelltyp der myeloischen Zelllinie, differenzieren (Grouard *et al.*, 1997; Janeway *et al.*, 2001). Die Existenz eines zweiten DC-Vorläuferzelltyps, mDC, verschärft Fragen nach der Herkunft von pDC sowie der Beziehung zwischen mDC und pDC (Vezzio *et al.*, 1996; Comeau *et al.*, 2002). Beide Zelltypen

---

<sup>15</sup>Eine genkodierte Familie an Glycoproteinen, deren Proteinprodukte an der Zelloberfläche lokalisiert sind und T-Helferzellen aktivieren (Smith *et al.*, 2003).

differenzieren sich nach Aktivierung in DC. Dabei sekretieren pDC wie mDC IL-6, was Entzündungen markiert sowie TNF- $\alpha$ , was an der Verdauung von als „fremd“ erkannten Pathogenen beteiligt ist und dem angeborenen Immunsystem zugeordnet wird (Locksley, Killeen und Lenardo, 2001; Comeau *et al.*, 2002; Chan *et al.*, 2005; Lenert, 2006; Ghirelli, Zollinger und Soumelis, 2010). Beide unterscheiden sich in ihren exprimierten Rezeptoren. Während mDCs alle *toll-like*-Rezeptoren außer TLR7 und TLR9 exprimieren, wird der IL-3-Rezeptor nur in geringem Maße exprimiert (Grouard *et al.*, 1997; Liu, 2005; Cao und Liu, 2007). Dem gegenüber exprimieren pDCs den IL-3-Rezeptor in hohem Maße sowie TLR7 und TLR9, dafür jedoch kaum andere TLRs (Grouard *et al.*, 1997; Colonna, Trinchieri und Liu, 2004; Liu, 2005). Beide Zelltypen scheinen deshalb aus unterschiedlichen Zelllinien zu entstammen, während pDC lymphoider Herkunft sind, scheinen mDC myeloischer Herkunft zu sein. Dagegen zeigten Comeau *et al.*, dass sich pDCs in mDCs umdifferenzieren können, wodurch pDCs zwischen lymphoider und myeloischer Zelllinie stehen (Comeau *et al.*, 2002; Colonna, Trinchieri und Liu, 2004; Liu, 2005). Die Unterschiede in den exprimierten Zellmarkern sowie die Möglichkeit der Umwandlung von pDC zu mDC legen die Existenz einer gemeinsamen Vorläuferzelle nahe, welche beide Zelltyp-Populationen bildet (siehe Abb. 2.7) (Comeau *et al.*, 2002; Hoyo *et al.*, 2002; Zuniga *et al.*, 2004; Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007). Dabei ist es unsicher, ob sich diese Zelle direkt aus der Stammzelllinie oder aus der lymphoiden Vorläuferzelle ableitet (Hoyo *et al.*, 2002; Colonna, Trinchieri und Liu, 2004). Abbildung 2.7 zeigt die mögliche Abstammung von der hämatopoietischen Stammzelllinie, wobei gestrichelte Linien Unsicherheiten darstellen.

Die Mediatorrolle von pDC zwischen angeborenem und erworbenem Immunsystem wird jedoch nicht bestritten, da pDCs auf Signale beider Immunsystemteile reagieren. Beispielsweise detektieren TLR7 und TLR9 verschiedene Viren, wodurch sich die Rolle im innaten Immunsystem zeigt (Takeda und Akira, 2004; Liu, 2005; Chan *et al.*, 2005; Lenert, 2006; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008; Ghirelli, Zollinger und Soumelis, 2010). Dabei werden durch TLR9 doppelsträngige DNA-Viren detektiert, wie beispielsweise das *Herpes simplex Virus I* (HSV-I) (Lund *et al.*, 2003; Krug *et al.*, 2004; Zuniga *et al.*, 2004; Guiducci *et al.*, 2006; Kawai und Akira, 2006; Finberg, Wang und Kurt-Jones, 2007; Guiducci *et al.*, 2008; Gilliet, Cao und Liu, 2008). Durch TLR7 hingegen werden

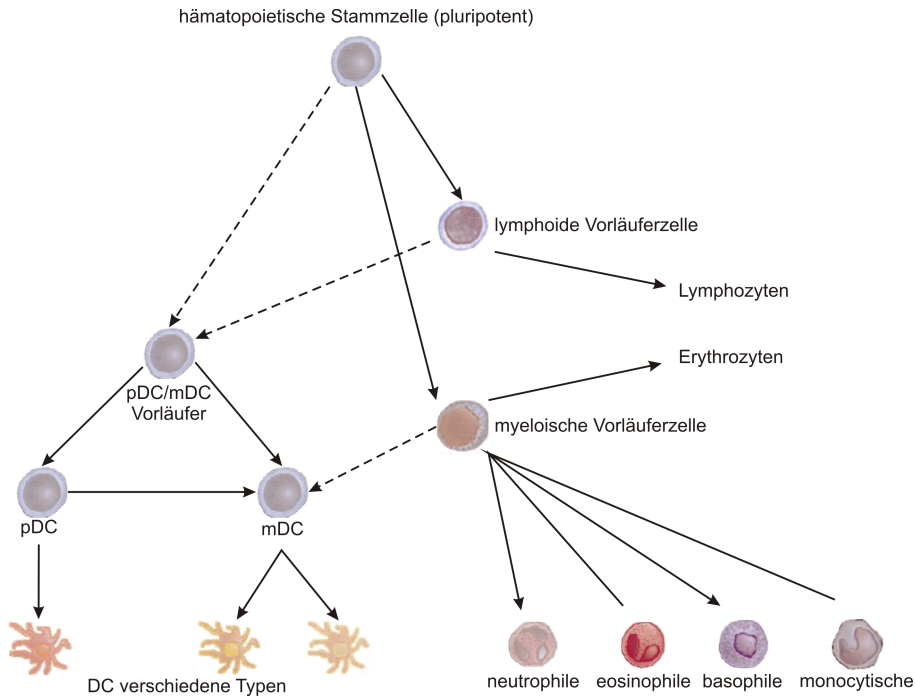


Abbildung 2.7: Ableitung verschiedener Zelltypen aus der hämatopoietischen Stammzelllinie. Die Entwicklung von pDC ist unklar, wie die gestrichelten Linien zeigen, ebenso wie die Existenz der Vorläuferzelle unklar ist (Colonna, Trinchieri und Liu, 2004). Existiert diese nicht, leiten sich pDCs entweder direkt von der Stammzelle oder von der lymphoiden Vorläuferzelle ab, eine myeloische Herkunft ist unwahrscheinlich. Durch Zelltypkonversion sind pDCs in der Lage sich in mDCs umzuwandeln (Comeau *et al.*, 2002). Die aus den Subpopulationen abgeleiteten DCs erfüllen hauptsächlich Aufgaben im erworbenen Immunsystem (siehe dazu Martinez-Moczygemba und Huston, 2003). Die Darstellung der Zelltypen wurde aus Martinez-Moczygemba und Huston, 2003 entnommen und verändert.

einzelsträngige RNA-Viren wie etwa das Influenza-Virus detektiert (siehe Unterabschnitt 2.2.4) (Chan *et al.*, 2005; Seth, Sun und Chen, 2006; Kawai und Akira, 2006; Finberg, Wang und Kurt-Jones, 2007; Zhang *et al.*, 2007; Gilliet, Cao und Liu, 2008; Hoeve *et al.*, 2012). IL-3 hingegen ist ein Signal, welches eine Rolle bei Entzündungsreaktionen und der adaptiven Immunantwort einnimmt (siehe Unterabschnitt 2.2.3). Beide Signalformen aktivieren pDC, wodurch die Zellen zu DC heranreifen (Grouard *et al.*, 1997; Liu, 2005; Cao und Liu, 2007). Dabei werden die Interferone  $\alpha$  und  $\beta$  in hohen Mengen ausgeschüttet, wodurch pDCs auch den Namen Interferon-produzierende Zellen (IPC) erworben haben (Siegal *et al.*, 1999; Colonna, Trinchieri und Liu, 2004; Zuniga *et al.*, 2004; Liu, 2005; Gilliet, Cao und Liu, 2008). Die zusätzliche Sekretion von Cyto- und Chemokinen als inflammatorische Signale markieren die Infektionsstellen, wodurch das Immunsystem reguliert und beispielsweise weitere Immunzellen rekrutiert werden (Colonna, Trinchieri und Liu, 2004; Cao und Liu, 2007; Kool *et al.*, 2009).

### 2.2.3 Interleukin-3, ein aktivierendes inflammatorisches Signal

Interleukin-3 (IL-3) ist ein 1985 durch Ihle *et al.* erstmals beschriebenes Cytokin sowie ein hämatopoietischer Wachstumsfaktor (Pirog, Stabinsky und Goldman, 2010). Als Signal wirkt IL-3 sowohl in erworbener als auch in angeborener Immunantwort (Ihle *et al.*, 1985; Ihle, 1995; Guthridge *et al.*, 1998; Mroczko und Szmitkowski, 2004; Zhu und Bertics, 2011). Es beeinflusst dabei Zelldifferenzierung, Apoptose, Proliferation und Reifeprozesse sowie die Lokalisation von Immunzellen (Peso *et al.*, 1997; Songyang *et al.*, 1997; Groot, Coffey und Koenderman, 1998; Guthridge *et al.*, 1998; Shelton *et al.*, 2004; Lopez *et al.*, 2010; Zhu und Bertics, 2011). Produzenten für IL-3 sind dabei T-Zellen, Mastzellen oder NK-Zellen (Pirog, Stabinsky und Goldman, 2010). Durch die vielfältigen Wirkungen in der Signaltransduktion ist es mit verschiedenen Krankheiten assoziiert, beispielsweise allergischen Immunreaktionen, Asthma und Krebs sowie mit Alzheimer, Schizophrenie und weiteren neurodegenerativen Erkrankungen (Arai *et al.*, 1990; Groot, Coffey und Koenderman, 1998; Yoshimura *et al.*, 1996; Martinez-Moczygemba und Huston, 2003; Mroczko und Szmitkowski, 2004; Chen und Kendler, 2008; Dey *et al.*, 2009; Zambrano *et al.*, 2010).

Über den IL-3-Rezeptor perzipieren die Zellen IL-3 als Liganden, wobei der Rezeptor alleine nicht in der Lage ist, das extrazelluläre Signal in die Zelle zu transferieren. Durch eine gemeinsame  $\beta$ -Untereinheit, die sich der IL-3-Rezeptor zusammen mit den Rezeptoren für IL-5 und GM-CSF teilt, wird das Signal in den intrazellulären Raum transferiert (Ihle, 1995; Adachi und Alam, 1998; Martinez-Moczygamba und Huston, 2003). Über die Struktur dieses Rezeptors besteht Uneinigkeit, wobei *Dey et al.* ihn zuletzt als Dodecamer beschreiben, in dem vier  $\beta$ -Untereinheiten eine IL-3-spezifische  $\alpha$ -Untereinheit binden und zwei dieser Teilstrukturen dimerisieren (*Dey et al.*, 2009).

Außerdem fehlt dem Rezeptorkomplex die intrinsische Kinaseaktivität, weshalb er nicht in der Lage ist, die Signalkaskade zu aktivieren. Erst die Bindung zweier Janus-Kinasen (JAK) an den Rezeptor und ihre gegenseitige Phosphorylierung führen zur Aktivierung der Signalkaskade (*Huang et al.*, 2005*b*; Huang, Lee und Chang, 2006; *Dey et al.*, 2009; *Lopez et al.*, 2010). Welche JAKs miteinander interagieren und ob sie unterschiedliche Signalpfade anstoßen (beispielsweise einen signalspezifischen für IL-3 und einen generellen, der durch die gemeinsame  $\beta$ -Untereinheit ausgelöst wird), ist nicht vollständig geklärt (*Watanabe, Itoh und Arai*, 1996; *Guthridge et al.*, 1998; *Adachi und Alam*, 1998; *Woodcock, Bagley und Lopez*, 1999; *Huang et al.*, 2005*b*; *Huang, Lee und Chang*, 2006; *Lopez et al.*, 2010).

In dieser Arbeit ist IL-3 relevant, da es zum einen die Ausdifferenzierung von pDC aktiviert. Zum anderen werden durch IL-3 die erworbene Immunantwort sowie inflammatorische Ereignisse, beispielsweise die Freisetzung großer Mengen an Interferonen, ausgelöst. Dadurch wird der Infektionsherd markiert, weitere Immunzellen aktiviert bzw. rekrutiert und die Immunantwort gesteuert (*Korpelainen et al.*, 1996; *Bosscher, Berghe und Haegeman*, 2000; *Colonna, Trinchieri und Liu*, 2004; *Cao und Liu*, 2007).

## 2.2.4 Das Influenza-A-Virus im Immunsystem

Die Influenza oder Grippe, ausgelöst durch das Influenza-A-Virus (Flu), ist eine der häufigsten Viruserkrankungen bei Menschen (Hoeve *et al.*, 2012). Sie betrifft hauptsächlich den respiratorischen Trakt und wird begleitet durch z.B. Fieber, Kopf- und Gliederschmerzen, Husten, Schweißausbrüche und Schnupfen (Cox und Subbarao, 1999). Sie tritt saisonal epidemisch oder pandemisch auf und wurde schon von Hippokrates um etwa 500 v. Chr. beschrieben (Cox und Subbarao, 1999; Martin und Martin-Granel, 2006; Hoeve *et al.*, 2012).

Das 1933 erstmalig isolierte Influenza-A-Virus gehört zur Gruppe der Orthomyxoviridae und ist ein einzelsträngiges RNA-Virus (ssRNA) (Smith, Andrewes und Laidlaw, 1933; Regenmortel *et al.*, 2000; Tong *et al.*, 2012). Anhand der ausgelösten Antigen-Antikörper-Reaktion wird das Virus weitergehend unterschieden nach zwei vorkommenden Glykoproteinen, die an der Virushülle anhaften, dem Hämagglutinin (H) und der Neuraminidase (N) (Cox und Subbarao, 1999; Hoeve *et al.*, 2012; Tong *et al.*, 2012). Bislang wurden 17 Subtypen an Hämagglutinin (H1-H17) und 9 Typen an Neuraminidase (N1-N9) bestimmt, wodurch sich die klassische Bezeichnung für Grippeviren ableitet, wie beispielsweise H1N1 für Schweine- und spanische Grippe oder H5N1 für Vogelgrippe (Cox und Subbarao, 1999; Garten *et al.*, 2009; Hoeve *et al.*, 2012; Tong *et al.*, 2012). Die Vermehrung der Viren geschieht im Flimmerepithel der Atemwege, anschließend werden sie vom Wirt ausgeschieden und verbreiten sich über Tröpfcheninfektion (Cox und Subbarao, 1999).

Flu beeinflusst das angeborene Immunsystem und wird als ssRNA-Virus durch TLR7 in den Endosomen<sup>16</sup> der Zelle perzipiert (Takeda und Akira, 2004; Seth, Sun und Chen, 2006). Dadurch werden in pDC verschiedene TLR-abhängige Signaltransduktionswege aktiviert (Takeda und Akira, 2004; Liu, 2005; Takeda und Akira, 2005; Guiducci *et al.*, 2006; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008; Ghirelli, Zollinger und Soumelis, 2010). Flu induziert ebenfalls die Ausdifferenzierung von pDC zu DC, wobei große Mengen an Interferonen sowie weitere Cytokine ausgeschüttet werden (Siegal *et al.*, 1999; Jarrossay *et al.*, 2001; Akira,

---

<sup>16</sup>Ein Endosom ist eine von einer Membran umschlossene Zellorganelle (Vesikel), welche unter anderem durch Endozytose aufgenommenes Material zu spezifischen Organellen transportiert (Smith *et al.*, 2003).

2003; Krug *et al.*, 2004; Takeda und Akira, 2004; Chan *et al.*, 2005; Liu, 2005; Lenert, 2006; Hoeve *et al.*, 2012). Neben Interferon- $\beta$  werden besonders oft die Signalproteine RANTES und IL-6 ausgeschüttet, die eine immunverstärkende Wirkung besitzen und weitere Entzündungsmediatoren und Immunzellen rekrutieren bzw. aktivieren (Dawson *et al.*, 2000; Chan *et al.*, 2005; Pirog, Stabinsky und Goldman, 2010; Hoeve *et al.*, 2012).

Das Virus ist insgesamt ein starker und unspezifischer Aktivator besonders des angeborenen Immunsystems, weshalb es in dieser Arbeit als Stimulus eingesetzt wird.





# 3 Material und Methoden

In Abschnitt 3.1 werden die verwandten Datenbanken vorgestellt. Diese bilden die Grundlage, um Synergien in Netzwerken zu identifizieren. Ebenso sind sie notwendig, um ausgelöste Effekte funktional charakterisieren zu können. Abschnitt 3.2 stellt die eingesetzten Programme kurz vor. Abschnitt 3.3 beschreibt wie experimentelle Genexpressionsdaten gewonnen werden und wie diese methodologisch analysiert werden. In 3.4 wird der durch den französischen Partner generierte experimentelle Datensatz vorgestellt, dabei wird dargestellt unter welchen Gegebenheiten und wie pDCs und die beiden eingesetzten Stimuli IL-3 und Flu untersucht worden sind. Der letzte Abschnitt liefert einen kurzen Überblick, wie Daten abstrahiert werden können, um biologische Signaltransduktionsnetzwerke möglichst vollständig zu rekonstruieren.

## 3.1 Datenbanken

### 3.1.1 TRANSFAC<sup>®</sup>

Die TRANSFAC<sup>®</sup>-Datenbank repräsentiert enzyklopädische Informationen über die transkriptionelle Regulation von Genen in Eukaryoten, ausgelöst durch Interaktionen von Transkriptionsfaktoren mit DNA-Bindestellen (Wingender *et al.*, 1996; Wingender *et al.*, 2000; Matys *et al.*, 2006; Wingender, 2008). Transkriptionsfaktoren binden an spezifische Stellen der DNA, die sogenannten Transkriptionsfaktorbindestellen, wodurch die Expression der Zielgene und somit die Anpassung an die Umgebung reguliert wird.

Bei den aus TRANSFAC<sup>®</sup> stammenden Daten handelt es sich um aus einschlägiger Fachliteratur entnommene Interaktionen, bei denen experimentell ei-

ne Bindung zwischen einem TF und der DNA bzw. einer spezifischen DNA-Bindestelle nachgewiesen wurde (Wingender *et al.*, 2001; Matys *et al.*, 2003). Ein TF reguliert das Zielgen, wobei die Produkte des Gens selbst wieder genregulatorische Funktionen erfüllen oder andere zelluläre Prozesse steuern. Durch Anbindung externer Datenbanken, wie etwa *SwissProt* (Bairoch und Boeckmann, 1992) oder *PDB* (Berman *et al.*, 2000), wird der enzyklopädische Charakter von TRANSFAC<sup>®</sup> komplettiert, beispielsweise durch Hinzunahme von Protein- und Strukturinformationen (Heinemeyer *et al.*, 1999). Die TRANSFAC<sup>®</sup>-Datenbank wird von der Biobase GmbH vertrieben, wobei in der vorliegenden Arbeit die kommerzielle *Professional* Version 2009.4 verwendet wird. Eine in der abrufbaren Datenmenge deutlich reduzierte öffentliche Version der Datenbank ist unter <http://www.gene-regulation.com/pub/databases.html> verfügbar.

#### 3.1.2 TRANSPATH<sup>®</sup>

In der Datenbank TRANSPATH<sup>®</sup> werden Informationen über Signaltransduktionsprozesse, Signalmoleküle, Modifikationen und Komplexe aus verschiedenen eukaryotischen Spezies zusammengefasst (Schacherer *et al.*, 2001; Krull *et al.*, 2003; Choi *et al.*, 2004b; Krull *et al.*, 2006). Der Fokus von TRANSPATH<sup>®</sup> liegt in der Repräsentation regulatorischer Signaltransduktionspfade und ihren Interaktionen mit TFs (Schacherer *et al.*, 2001). Die abgebildeten Interaktionen und Moleküle (Proteine, TFs, Signalmoleküle etc.) wurden experimentell beobachtet und in Fachpublikationen veröffentlicht (Schacherer *et al.*, 2001; Choi *et al.*, 2004a).

Verschiedene experimentelle Methoden unterliegen jedoch einer unterschiedlichen Auflösung oder Granularität in ihren Ergebnissen. Um diesem Problem zu begegnen nutzt TRANSPATH<sup>®</sup> hierarchische und orthologe Abstraktionen zur Repräsentation experimenteller Informationen (siehe Abb. 3.1 und Abschnitt 3.5) (Choi *et al.*, 2004b; Krull *et al.*, 2006). Beispielsweise werden verschiedene Proteine zu Proteinfamilien zusammengefasst oder verschiedene Spleiß-Varianten durch eine gemeinsame Isoform repräsentiert (siehe Abb. 3.1). Orthologe Abbildungen werden durch Verallgemeinerung der speziesspezifischen Informationen zu einer speziesspezifischen Sichtweise realisiert, sodass beispielsweise Proteine mit gleicher Funktionalität zusammengefasst werden (siehe Abb. 3.1). Durch or-

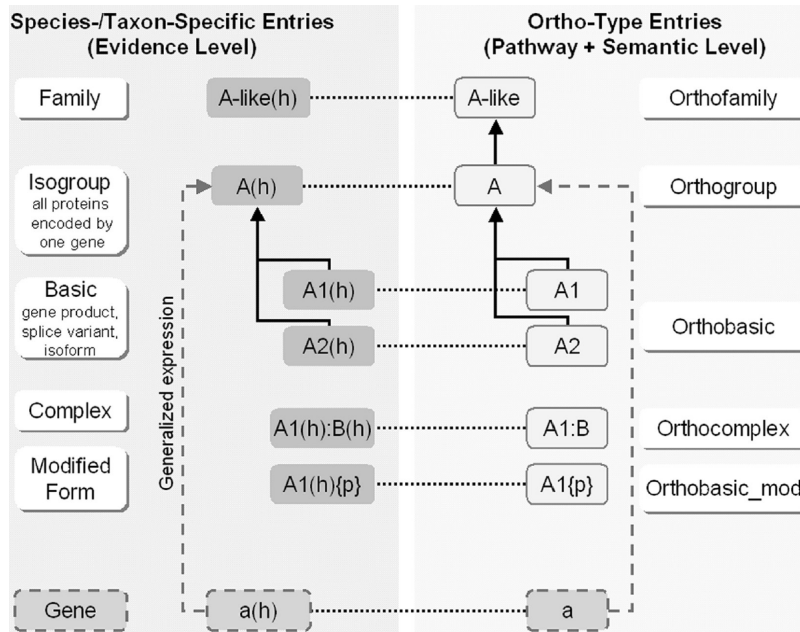


Abbildung 3.1: Orthologe und hierarchische Abstraktion in TRANSPATH<sup>®</sup> (aus Krull *et al.*, 2006). Links werden die speziesspezifischen Informationen dargestellt. Dabei besitzt die unterste Ebene die höchste Genauigkeit (Granularität), welche nach oben hin nachlässt. Rechts werden verallgemeinerte speziesspezifische Informationen dargestellt. Gepunktete Linien zeigen, wie speziesspezifische Informationen abstrahiert werden. Gestrichelte Elemente zeigen, wie genomische Informationen (beispielsweise aus TRANSFAC<sup>®</sup>) inkludiert werden.

thologe oder hierarchische Abbildungen und die damit verbundene Abstraktion von Wissen werden generelle Referenzpfade abgeleitet, die mit spezifischen Funktionen verbunden sind und in allen Spezies vorliegen (Choi *et al.*, 2004b; Krull *et al.*, 2006). Zum Beispiel ist der EGF-Signaltransduktionspfad in Säugerzellen generell für die Steuerung von Zellwachstum, Differenzierung und Überleben verantwortlich (Oda *et al.*, 2005).

Referenzpfade dienen dazu, einen generellen Überblick über die wichtigsten Signaltransduktionsschritte in einer sogenannten „semantischen Sichtweise“ zu erhalten. Dabei werden viele Informationen nicht repräsentiert, beispielsweise Moleküle, wie ATP, H<sub>2</sub>O oder Reaktionen (Choi *et al.*, 2004b). Die spezifischen, mechanistischen Informationen über einfache Signalmoleküle, Modifikationen, Katalysatoren und Reaktionen werden bei TRANSPATH<sup>®</sup> im sogenannten *evidence*

*level* hinterlegt (Choi *et al.*, 2004b). Dieses zeigt die höchste Genauigkeit hinsichtlich der ablaufenden Prozesse und repräsentiert die molekularen Grundlagen. Aufgrund der hohen Anzahl abgebildeter Komponenten ist die Darstellung des *evidence level* jedoch sehr unübersichtlich (Choi *et al.*, 2004b; Krull *et al.*, 2006).

Zusammen mit TRANSFAC<sup>®</sup> (siehe Unterabschnitt 3.1.1) ist es möglich, ein ganzheitliches Bild der Signaltransduktion zu skizzieren, welches von extrazellulär vorliegenden Signalen über die von ihnen angestoßenen molekularen Prozesse bis zu den regulierten Genen reicht (Choi *et al.*, 2004a). Im Rahmen dieser Arbeit wurde TRANSPATH<sup>®</sup> *Professional* in der Version 2009.4 benutzt.

### 3.1.3 Kyoto Encyclopedia of Genes and Genomes (KEGG)

Die KEGG-Datenbanken dienen dazu ganzheitlich die grundlegenden und höheren Funktionen von Zellen und Organismen zu verstehen (Kanehisa, 1997; Kanehisa *et al.*, 2002; Kanehisa *et al.*, 2006). Dabei werden genomische, chemische sowie systemische Informationen durch einen modularen Ansatz miteinander kombiniert, sodass alle wesentlichen Aspekte eines Organismus systembiologisch untersucht werden können (Kanehisa *et al.*, 2008).

Im Rahmen dieser Arbeit sind vor allem die Datenbankmodule KEGG-*GENES*, KEGG-*PATHWAY* und KEGG-*BRITE* relevant. KEGG-*GENES* ist eine Sammlung genomischer Informationen, beginnend mit einzelnen sequenzierten Genen bis hin zu Sammlungen kompletter Genome von Pro- und Eukaryoten (Kanehisa *et al.*, 2006). Wie TRANSPATH<sup>®</sup> oder TRANSFAC<sup>®</sup> komplettiert KEGG-*GENES* die repräsentierten Inhalte durch Anbindung weiterer öffentlicher Datenbanken (Kanehisa *et al.*, 2006; Kanehisa *et al.*, 2008). Mittels KEGG-*BRITE* werden Proteine und ihre Funktionen untersucht. Ähnlich zu TRANSPATH<sup>®</sup> (siehe Unterabschnitt 3.1.2 und Abb. 3.1) liegt eine hierarchische Strukturierung vor, die das Wissen kategorisiert. So werden Proteine nach ihrer Funktion unterschieden und in Proteinfamilien zusammengefasst. Außerdem werden Krankheiten oder die Wirkung einzelner Proteine auf einen Organismus abgebildet. So ist es möglich, Enzyme, TFs und Rezeptoren bezüglich assoziierter Krankheiten oder ihrer jeweiligen Funktion zu untersuchen (Kanehisa *et al.*, 2002; Kanehi-

sa *et al.*, 2006). Das Modul *KEGG-PATHWAY* zeigt anhand manuell erstellter und abstrakter Interaktionskarten, wie und wo Proteine in einer Zelle wirken und wie sie miteinander interagieren (Kanehisa *et al.*, 2002; Kanehisa *et al.*, 2004; Kanehisa *et al.*, 2006; Kanehisa *et al.*, 2008). Dabei werden metabolische, Signaltransduktions- und andere zelluläre Prozesse abgebildet, sodass generelle Funktionen von Genen und Proteinen bestimmt werden.

## 3.2 Programme

### 3.2.1 GraphViz

GraphViz ist eine Sammlung verschiedener Werkzeuge zum Visualisieren von Graphen (Ellson *et al.*, 2002; Ellson *et al.*, 2004). Mittels einer deskriptiven Sprache (*dot*), werden in Textdateien Ecken und Kanten eines Graphen sowie sämtliche stilistische Parameter, beispielsweise Form und Farbe von Ecken, definiert (Ganser und North, 1999). Der Vorteil von GraphViz liegt im einfachen Zugang zur Sprache und der vielfältigen Anwendbarkeit sowie der Einbindung von Graphen in Web-Applikationen. Durch verschiedene Layout-Algorithmen, wie *dot*, *neato*, *fdp* oder *circo* werden die Textdateien mithilfe eines Interpreters in statische Graphen umgewandelt, wobei die Ausgabe in gängigen Bildformaten, wie *jpg* oder *png* erfolgt (Ganser und North, 1999; Ellson *et al.*, 2002; Ellson *et al.*, 2004). Im Rahmen dieser Arbeit wird GraphViz zur Visualisierung von Netzwerken verwandt. GraphViz kann unter <http://www.graphviz.org> für verschiedene Plattformen bezogen werden.

### 3.2.2 R

R als Skriptsprache ermöglicht die Formulierung und Lösung statistischer Fragestellungen (Ihaka und Gentleman, 1996). Dank der Vielzahl vorimplementierter Methoden und weiterer zusätzlich installierbarer Pakete dient R der statistischen Auswertung und grafischen Analyse von umfangreichen Studien, beispielsweise durch statistische Tests, Histogramme etc. (Sachs und Hedderich, 2006; Krijnen,

2009). Die Definition eigener Funktionen sowie die Anpassbarkeit an jeweils spezifische Fragestellungen verleiht R eine hohe Flexibilität, weshalb es in der Bioinformatik zu einem der wichtigsten Werkzeuge gehört (Ihaka und Gentleman, 1996; Gentleman, 2008; Krijnen, 2009). R steht unter <http://www.r-project.org> zur Verfügung.

Die Bioconductor-Erweiterung für R fokussiert auf molekularbiologische Fragestellungen und sogenannte Hochdurchsatz-Experimente, wobei verschiedene unabhängige Experimente in hoher Anzahl parallel ausgeführt werden (Gentleman *et al.*, 2004; Gentleman *et al.*, 2005). Ein Beispiel dafür sind Microarray-Experimente (siehe Unterabschnitt 3.3.1), bei denen zeitgleich die Expression verschiedener Gene bestimmt wird. Das Bioconductor-Paket hält Methoden bereit, die das Verhalten der Gen-Sonden auf einem Chip auswerten und normalisieren. Beispielsweise können mRNAs anhand ihrer Sequenzen an verschiedenen Stellen eines Microarray-Chips binden und Fehlhybridisierungen oder unspezifische Bindungen auslösen. Dies führt zu einem Hintergrundrauschen, welches vor der Analyse der Daten entfernt werden muss (Quackenbush, 2001; Gautier *et al.*, 2004; Wu *et al.*, 2004). Bioconductor-Pakete, wie beispielsweise *affy*, stellen Normalisierungsmethoden, wie *RMA*<sup>17</sup> oder *GCRMA* zur Verfügung. Dadurch wird eine umfassende und hochwertige Prozessierung von Daten möglich und das Hintergrundrauschen wird aus den Daten entfernt (siehe Irizarry *et al.*, 2003; Wu *et al.*, 2004; Gautier *et al.*, 2004; Irizarry, Wu und Jaffee, 2006). Darüber hinaus bietet Bioconductor mit dem *limma*-Paket die Möglichkeit, die Daten ganzheitlich auszuwerten und zu interpretieren. Beispielsweise werden besonders interessante, statistisch signifikant unterschiedlich exprimierte Gene detektiert (siehe Unterabschnitt 3.3.2) (Smyth, 2005; Nitsche, Ram und Meyer, 2012). Bioconductor ist ein Open-Source Projekt, welches unter <http://www.bioconductor.org> heruntergeladen werden kann.

---

<sup>17</sup>Im Gegensatz zu *GCRMA* werden bei der *RMA* keine gesonderten Fehlhybridisierungen betrachtet. Deshalb werden nur echte Hybridisierungen auf verschiedenen Chips zur Normalisierung betrachtet (Irizarry *et al.*, 2003).

## 3.3 Bestimmung von Genexpression und Identifikation differentiell exprimierter Gene

### 3.3.1 Microarrays - globale und simultane Bestimmung von Genexpression

Um zu verstehen, wie verschiedene Umwelteinflüsse und Signale das Verhalten von Zellen beeinflussen, ist es notwendig zu verstehen, wie einzelne Gene beeinflusst werden. Informationen über die Zusammenhänge werden mittels Genexpressionsanalyse gewonnen. Vor etwa 20 Jahren wurde diese neue Technologie eingeführt, die es ermöglicht, viele Gene und ihre Expression simultan zu überwachen. Diese Technik misst die Anzahl, der an einer kurzen Oligonukleotidsequenz bindenden Transkripte. Im Allgemeinen werden dabei Gene und ihre zugehörigen mRNA-Transkripte<sup>18</sup> gemessen (Schena *et al.*, 1995). Microarrays stellen die industrielle Umsetzung dieser Technologie dar (Schena *et al.*, 1995; Brown und D. 1999). Sie sind in einer Matrix organisiert, wobei eine spezifische Kombination aus Zeile und Spalte eine Sonde definiert.

Jede Sonde besteht aus elf Wiederholungen, einer künstlich auf den Microarray aufgebracht und paarig angeordneten kurzen Oligonukleotidsequenz von 20-25bp. Diese kurzen Oligonukleotide sind der Lage entsprechende mRNA-Transkripte anhand von komplementären Sequenzen zu binden (Schena *et al.*, 1995; Lockhart *et al.*, 1996; Wu *et al.*, 2004). Bei einem solchen Paar weist die erste Oligonukleotidkette spezifische mRNA-Bindungen nach, wobei Gene anhand der Sequenz auf dem Microarray identifiziert werden. Die zweite Oligonukleotidsequenz ist mittig mutiert und dient dem Nachweis unspezifischer mRNA-Bindungen. Sie zeigt an, wie oft mRNA-Transkripte zufällig an einer Sonde binden.

Nach Bindung der mRNA an die Sonden werden durch Inkubation mit einem fluoreszierenden Agens die Oligonukleotidstränge, welche mRNA binden markiert und anschließend photometrisch gemessen (*GeneChip® Human Genome*

---

<sup>18</sup>Eine mRNA ist eine Kopie eines Gens bzw. eines Sequenzabschnittes der DNA, welches anschließend in ein Protein translatiert wird und so die im Gen kodierte Eigenschaft realisiert.



*U133 Arrays*; Wu *et al.*, 2004). Durch Verrechnung spezifischer und unspezifischer Hybridisierungen<sup>19</sup> wird die Expressionsrate einer Sonde bestimmt (Lockhart *et al.*, 1996; Brown und D. 1999; *GeneChip® Human Genome U133 Arrays*). Obwohl jede Sonde, durch ihre spezifische Sequenz, einem Gen zugeordnet ist, kann anhand verschiedener Sonden (mit verschiedenen Sequenzen) ein gleiches Gen mehrfach gemessen werden. Dadurch wird die statistische Relevanz verbessert, wodurch quantitative Rückschlüsse auf die Expressionsraten der Gene möglich werden (Gautier *et al.*, 2004). Der wesentliche Vorteil beim Einsatz von Microarrays liegt jedoch in der simultanen Messung verschiedenster Gene auf einem Chip<sup>20</sup> (Lockhart *et al.*, 1996; Brown und D. 1999; Canales *et al.*, 2006). Dadurch ermöglichen Microarrays erstmals die genomweite Untersuchung von Genexpression (Schena *et al.*, 1995).

In dieser Studie wurde als experimentelle Plattform der Affymetrix HG-U133 2.0+ Microarray Chip genutzt. Dieser deckt eine Bandbreite von mehr als 54.000 Sonden ab, was mehr als zwanzigtausend verschiedenen Genen entspricht, die parallel gemessen werden (*GeneChip® Human Genome U133 Arrays*). Somit deckt dieser Chip den Großteil des menschlichen Genoms ab und liefert einen globalen Überblick über die exprimierten Gene.

### 3.3.2 Identifikation differentiell exprimierter Gene mittels *limma*

Zur Analyse großer Datenmengen, wie sie bei Microarray-Untersuchungen anfallen, sind computergestützte Methoden unerlässlich. Die Bioconductor-Erweiterung für R (siehe Unterabschnitt 3.2.2) stellt verschiedene Methoden zur Analyse von großen molekularbiologischen Datensätzen zur Verfügung (Ihaka und Gentleman, 1996; Gentleman *et al.*, 2004; Gentleman *et al.*, 2005).

Das *limma*-Paket, ein Teil der Bioconductor-Erweiterung, bietet beispielsweise Methoden zur Analyse von Genexpressionsdaten. Dadurch wird verstanden, wie sich Gene unter verschiedenen physiologischen Bedingungen verhalten, z.B.

---

<sup>19</sup>Als Hybridisierung bezeichnet man das Binden einer exprimierten mRNA an einer auf dem Chip aufgetragenen komplementären Oligonukleotid-Sequenz.

<sup>20</sup>Chip = ein Microarray Chip, ist genau eine Messung für alle Sonden, die sich auf dem Chip befinden.



welche Krankheiten ausgelöst werden können und welche genomischen Ursachen dafür verantwortlich sind. Durch *limma* werden in einem Datensatz die Gene identifiziert die zwischen verschiedenen Bedingungen statistisch signifikant differentiell exprimiert sind. Heutzutage ist die Analyse von Genexpressionsdaten mit *limma* als Standardmethode etabliert, welche dabei hilft die kausalen Ursachen für die beobachteten Unterschiede aufzuklären (Mount, 2004; Smyth, 2005; Gentleman *et al.*, 2005; Selbach *et al.*, 2008; Ideker, Dutkowski und Hood, 2011; Nitsche, Ram und Meyer, 2012).

Mit Hilfe von *limma* werden lineare Modelle berechnet, die das Verhalten eines Gens in einem definierten Umfeld beschreiben (Smyth, 2005). Ursprünglich wurde diese Technik von R.A. Fisher eingeführt, um unvoreingenommene Vergleiche von Objekten durchzuführen, die mehreren Parametern unterliegen. Dabei sollte aber nur ein spezifischer Parameter verglichen werden (Kerr und Churchill, 2001). Eine Zelle unterliegt vielen Faktoren, wie etwa Lokalisation, Signale benachbarter Zellen oder Hormonwirkungen im gesamten Körper. Verglichen werden sollen hingegen nur gesunde mit krankhaften Zellen hinsichtlich ihrer Unterschiede in der Genexpression. Auf beide Bedingungen wirken daher alle anderen Parameter gleichermaßen ein. Ein lineares Modell muss in der Lage sein, diese Parameter in adäquater Form zusammenzufassen, sodass die relevanten Abweichungen sicher identifiziert werden.

Ein lineares Modell ist zum Beispiel gegeben als:

$$y_{i\dots k} = \mu + \underbrace{A_i + \dots + A_k}_{\text{Einzelparameter}} + \underbrace{A_i A_{i+1} + \dots + A_{k-1} A_k}_{\text{Kombination von Parametern}} + \epsilon_{i\dots k}$$

Dabei ist  $y_{i\dots k}$  die gemessene Größe, beispielsweise die gemessenen Expressionsraten und  $\mu$  ist das arithmetische Mittel aller Messungen.  $A_i \dots A_k$  beschreibt den Einfluss der einzelnen Parameter auf die beobachtete Expression und  $A_i A_{i+1}$  den gemeinsamen Einfluss zweier Parameter  $A_i$  und  $A_{i+1}$ .  $\epsilon_{i\dots k}$  ist ein allgemeiner Fehlerterm.

Bei Microarrays werden lineare Modelle für die einzelnen Sonden berechnet. Dazu muss die Variabilität der jeweiligen Beobachtungen berechnet werden (Kerr und Churchill, 2001). Zuerst wird deshalb die Variabilität einer Sonde unter gleichartigen Zuständen geschätzt. Dies geschieht über mehrfache Messungen mittels

verschiedener Chips unter gleichen Bedingungen. Als zweites wird die Variabilität aller Sonden auf jeweils einem Chip betrachtet. Beide Messungen zusammengenommen ergeben für jede Sonde ein lineares Modell, welches die Genexpression unter bestimmten Bedingungen (z.B. gesunde Zellen, frisch extrahiert) beschreibt (Kerr und Churchill, 2001). Formalisiert ergibt sich ein derartiges lineares Modell als:

$$y_{ijk} = \mu + \beta_1 C_i + \beta_2 S_j + \beta_3 (C_i S_j) + \beta_4 (B_k S_j) + \epsilon_{ijk}$$

Dabei beschreibt  $\beta_1 C_i$  den  $i$ -ten Chip,  $\beta_2 S_j$  die  $j$ -te Sonde und  $y_{ijk}$  das Signal einer Sonde  $j$  auf einem Chip  $i$  unter der Bedingung  $k$ .  $\mu$  ist die gemittelte Intensität aller gemessenen Expressionsraten über alle Chips und alle Sonden.  $\beta_4 (B_k S_j)$  ist die Beschreibung der Variabilität zwischen den Bedingungen  $k$  und den Sonden  $j$ . Da  $C$ ,  $S$ ,  $CS$  und  $BS$  als Zufallsvariablen angenommen werden, ist  $\beta_i$  die Beschreibung der Verteilung der Zufallsvariable anhand eines spezifischen Satzes an Hyperparametern. Dadurch wird beispielsweise beschrieben, welche Faktoren auf den  $i$ -ten Chip einwirken ( $\beta_1 C_i$ ).

Ein grundsätzliches Problem bei biologischen Microarray-Studien ist die oft kleine Menge an Replikaten. *limma* nutzt zur Verbesserung der Vorhersagegüte den empirischen Bayes-Schätzer, der sich der Daten verschiedener Sonden auf dem gleichen Chip bedient und so die Parameterschätzungen für das Modell verbessert (Robbins, 1956; Smyth, 2004; Smyth, 2005). Daher nimmt man vereinfachend an, dass die Parameter eines linearen Modelles einer fixen aber unbekanntem Verteilung folgen, welche mittels des empirischen Bayes-Schätzers bestimmt wird. Deshalb ist beispielsweise  $\beta_1$  für alle Chips  $C_i$  gleich und kann durch den Mittelwert bestimmt werden (Mount, 2004). Die Werte für jedes  $\beta_{1\dots n}$  werden unabhängig voneinander bestimmt (für weitergehende Erklärungen siehe Robbins, 1956; Kerr und Churchill, 2001; Mount, 2004).

Verschiedene lineare Modelle für eine Sonde, z.B. unter verschiedenen Bedingungen (etwa IL-3 anwesend vs. Flu anwesend oder gesund vs. krankhaft), können miteinander verglichen werden. Mittels einer F-Statistik werden so die statistisch signifikant differentiell exprimierten Sonden identifiziert (siehe auch Kerr und Churchill, 2001; Mount, 2004; Sachs und Hedderich, 2006; Smyth *et al.*, 2007).

Grundsätzlich läuft die *limma*-Analyse in mehreren Schritten ab. In den ersten Schritten erfolgt die Prä-Prozessierung, bei der die Rohdaten eingelesen, vorverarbeitet und normalisiert werden. Außerdem wird das Experiment mittels einer Experimentalmatrix beschrieben, in deren Zeilen die jeweiligen Sonden und in deren Spalten die gemessenen Intensitäten unter einer bestimmten Bedingung auf einem Chip dargestellt werden (Smyth *et al.*, 2007).

Die Normalisierung der Daten ist eine notwendige Voraussetzung zur Erstellung linearer Modelle, da die gemessene Genexpression jeder Sonde über die verschiedenen Replikate normalverteilt sein muss (Smyth, 2005; Smyth *et al.*, 2007). Nur wenn dies erfüllt ist, können verschiedene Modelle miteinander verglichen werden (Smyth *et al.*, 2007). Biologisch gesehen ist dies plausibel. Aufgrund technischer Fehler (z.B. Fehlmessungen, Fehlhybridisierungen) ist dies *a priori* nicht immer erfüllt.

Anschließend werden anhand der Experimentalmatrix die linearen Modelle berechnet und durch Definition einer Kontrastmatrix werden die interessanten Vergleiche bestimmt. Hierbei wird beschrieben, welche Bedingungen des Experiments miteinander verglichen werden. Nicht immer ist ein Vergleich aller Bedingungen notwendig oder sinnvoll. Zum Beispiel werden in einem Experiment drei verschiedene Bedingungen gemessen: eine krankhafte, eine gesunde und eine Negativkontrolle. Es ist jedoch nur der Vergleich zwischen krank und gesund interessant. Alle weiteren Vergleiche können vernachlässigt werden und müssen nicht definiert werden. Die Aufbereitung der erhaltenen Ergebnisse hängt stark vom Ziel der Analyse ab. Hier bietet *limma* verschiedene Methoden zur Präsentation der Ergebnisse (siehe Smyth *et al.*, 2007). Egal wie die Daten aufbereitet werden, am Ende existiert immer eine Liste mit identifizierten und signifikant differentiell exprimierten Sonden, welche dann zu den jeweils korrespondierenden Genen zusammengefasst wird.

## 3.4 Datensatz - Experimentell bestimmte Genexpression in pDC

In dieser Arbeit wurde durch Analyse eines experimentellen Datensatzes, welcher durch einen Projektpartner zur Verfügung gestellt wurde, der synergistische Einfluss von IL-3 und Flu auf plasmacytoide dendritische Zellen untersucht. Dieser 1958 von W. Remmele erstmalig beschriebene Zelltyp ist eng mit dem Funktionieren der mammalischen Immunantwort verbunden (Lennert und Remmele, 1958; Cao und Liu, 2007). Dabei beeinflussen pDCs sowohl die angeborene, wie auch die erworbene Immunantwort (siehe 2.2.2, Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008; Kool *et al.*, 2009). Nach Aktivierung durch ein externes immunrelevantes Signal beginnt die Differenzierung, wodurch letztlich eine Zelltypkonversion von pDC zu DC durchgeführt wird (O'Doherty *et al.*, 1994; Grouard *et al.*, 1997; Colonna, Trinchieri und Liu, 2004). pDCs entstammen der hämatopoietischen Stammzelllinie, das heißt sie werden im Knochenmark gebildet und als unreife Vorläuferzellen in lymphatische Gewebe (z.B. Lymphknoten) transportiert, wo sie heranreifen. Außerdem besitzen pDCs die Möglichkeit, sich weitgehend frei im Blutsystem zu bewegen, weshalb sie direkt aus Blut extrahiert werden können (Colonna, Trinchieri und Liu, 2004; Cao und Liu, 2007; Kool *et al.*, 2009).

Durchflussscytometrie ist ein in Medizin und Biologie häufig eingesetztes Verfahren, um Zellen direkt aus Körperflüssigkeiten, wie Blut oder Lymphe, zu extrahieren und anhand morphologischer Parameter, wie Form, automatisch voneinander zu trennen. Dabei wird eine Reinheit von etwa 98% je nach Zelltyp erreicht (Shapiro, 1995; Caron *et al.*, 2000; Riley und Idowu, 2003).

Mittels Durchflussscytometrie wurde das frisch extrahierte Blut dreier gesunder menschlicher Spender aufgereinigt. Dabei wurden die pDCs abgetrennt und in Kultur (Medium) überführt (siehe Abb. 3.2). Das Medium erhält die frisch extrahierten Zellen am Leben, stellt ihnen jedoch keine zusätzlichen, immunrelevanten Signale zur Verfügung, weshalb die Medium-Kultur als Negativkontrolle dient. Anschließend wurde die Medium-Kultur in vier Proben aufgeteilt, wobei eine Kontrolle weiterhin unbehandelt blieb (Abb. 3.2). Die anderen Proben wurden unabhängig voneinander jeweils 6h mit IL-3, Flu oder mit beidem (IL-3+Flu)

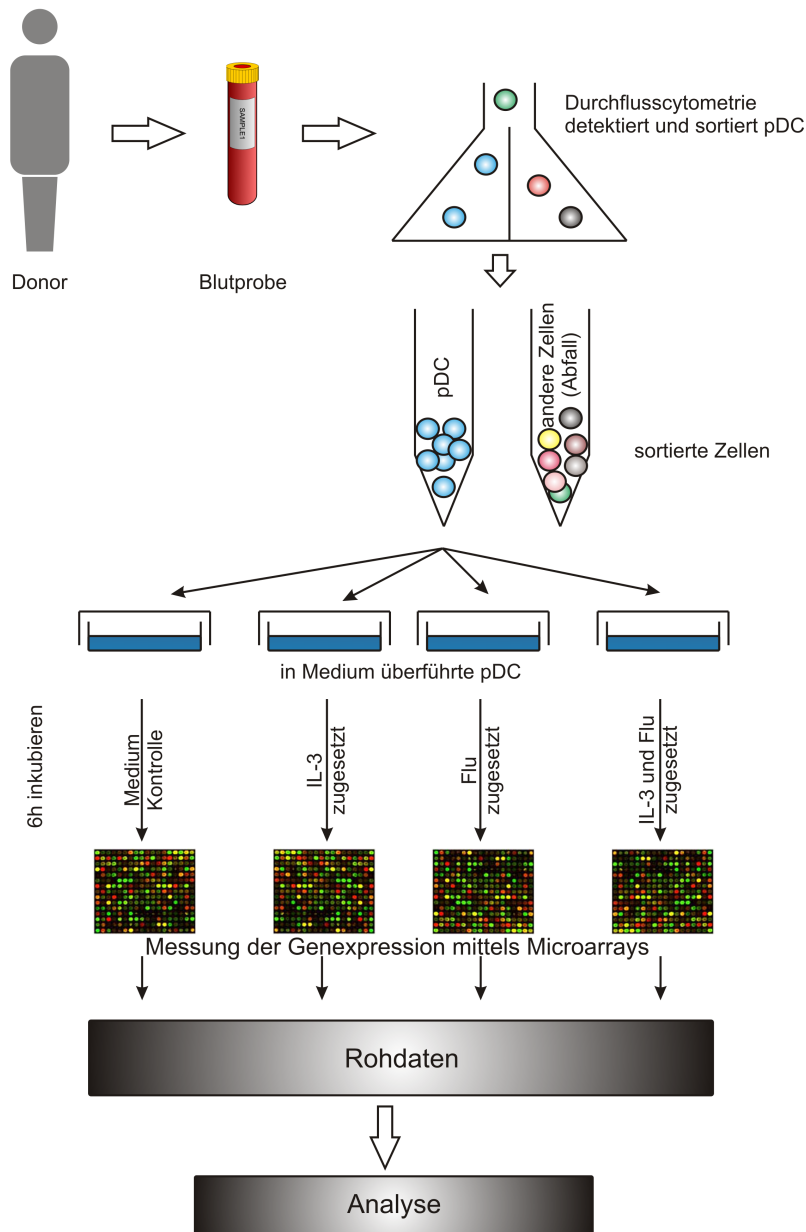


Abbildung 3.2: Schematischer Ablauf zur Gewinnung der Experimentaldaten. Aus einer Blutprobe wurden mittels Durchflusszytometrie pDCs extrahiert und in Medium überführt. Das Medium wurde auf 4 verschiedene experimentelle Ansätze aufgeteilt (Medium, Zusatz von IL-3, Zusatz von Flu, Zusatz von IL-3 und Flu) und 6h inkubiert. Anschließend wurden die Genexpressionsraten mittels Microarrays gemessen. Die erhaltenen Rohdaten wurden vorprozessiert und mittels *GCRMA* normalisiert, um anschließend umfassend analysiert zu werden.

inkubiert. Abschließend wurde die Genexpression jeder Probe mittels Affymetrix-Oligonukleotid-Chips gemessen (Lockhart *et al.*, 1996; *GeneChip® Human Genome U133 Arrays*). Die Messung der Genexpression wurde unter jeder Bedingung dreifach wiederholt.

Die gewonnen Rohdaten jedes Chips wurden mittels *GCRMA*<sup>21</sup> normalisiert, sodass sie untereinander vergleichbar waren (Wu *et al.*, 2004; Gautier *et al.*, 2004; Ritchie *et al.*, 2007; Lim *et al.*, 2007). Die überwiegend experimentellen Arbeiten, von der Extraktion der Zellen aus menschlichem Blut bis hin zur Vorprozessierung der Microarray-Daten mittels *GCRMA* wurden durch das Department of immunology and cell biology am Institut Curie, Paris durchgeführt und als Datensatz zur Verfügung gestellt.

Um die gemessene Genexpression der Sonden auf einem Chip zu beschreiben wurden lineare Modelle mittels *limma* abgeleitet. Dabei wurden jedoch nur exprimierte Sonden betrachtet. Zur Bestimmung der exprimierten Sonden wurde ein Schwellwert von 2.40 der gemessenen Intensitäten (log-ratios<sup>22</sup>) festgelegt (Chen, Dougherty und Bittner, 1997; Quackenbush, 2001; Shi *et al.*, 2006). Die Wahl des Wertes basiert auf Expertenwissen und wurde durch den französischen Partner<sup>23</sup> anhand der Besonderheiten von pDC sowie bekannter technischer Probleme bei der Messung von Genexpression mittels Microarrays bestimmt (Quackenbush, 2001; *GeneChip® Human Genome U133 Arrays*). Alle Sonden, die unter jeder Bedingung immer unterhalb des gewählten Schwellwertes von 2.40 lagen wurden aus den Daten entfernt. Dadurch wird garantiert, dass nur Gene, repräsentiert durch ihre zugehörigen Sonden, betrachtet werden, welche nachweislich in pDC exprimiert sind.

Auf einem Chip werden einige Gene durch verschiedene Sonden gemessen. Diese werden nicht direkt zusammengefasst. Die verwendeten linearen Modelle nutzen die Variabilität aller Sonden auf einem Chip, beziehungsweise die Variabilität der gleichen Sonde auf verschiedenen Chips aus (siehe 3.3.2 sowie Mount, 2004;

---

<sup>21</sup>Bei *GCRMA* werden Fehlhybridisierungen auf dem Gen-Chip betrachtet und mit den *echten* Bindungen verrechnet, sodass die erhaltenen Genexpressionswerte genauer sind und auch verschiedene Chips miteinander verglichen werden können (Wu *et al.*, 2004).

<sup>22</sup>log-ratios beschreiben die Genexpression eines Microarrays wie HG-U133 2.0+ als Verhältnis der gemessenen Intensitäten von spezifischer und unspezifischer Hybridisierung:

$$\text{Intensität} = \frac{\log_2(\text{gemessene Intensität spezifische Hybridisierung})}{\log_2(\text{gemessene Intensität unspezifische Hybridisierung})} \quad (\text{Quackenbush, 2001}).$$

<sup>23</sup>*Department of immunology and cell biology, Institut Curie, Paris*

Smyth *et al.*, 2007). Der Datensatz ist daher größer, was die Qualität der linearen Modelle durch Nutzung vieler Datenpunkte verbessert (Smyth, 2004; Smyth, 2005; Smyth *et al.*, 2007). Die für jede Bedingung abgeleiteten linearen Modelle wurden abschließend miteinander verglichen, wodurch die experimentell synergistisch regulierten Gene bestimmt wurden (siehe Abschnitt 4.1). Dieser Datensatz stellt die Basis für die weitere Analyse synergistischer Effekte dar.

## 3.5 Datenabstraktion bei der Netzwerkmodellierung

Netzwerke werden als Graphen repräsentiert und bestehen aus Knoten und Kanten, wobei Knoten über Kanten miteinander verknüpft sind (Diestel, 2000; Mason und Verwoerd, 2007). Die Knoten eines Netzwerks repräsentieren die *statischen* Komponenten, beispielsweise Gene, Moleküle, Metabolite, Proteine usw., während die Kanten Verbindungen repräsentieren, z.B. die Interaktionen zwischen zwei Genen oder zwischen Proteinen. In der Biologie sind die Beziehungen zwischen verschiedenen Komponenten oft gerichtet. Beispielsweise regulieren TFs verschiedene Gene, wobei die meisten dieser Gene keinen Einfluss auf die Expression der TFs besitzen, sondern andere Funktionen vermitteln (Wingender, 1993). Durch gerichtete Kanten im Graphen werden die aus der Biologie bekannten Beziehungen realisiert.

In Signaltransduktionsnetzwerken stellen die jeweiligen Signalkomponenten sämtliche an der Signalweiterleitung und Prozessierung beteiligten Moleküle und Proteine die Knoten des Graphen dar. Die Relationen zwischen den Knoten werden durch gerichtete Kanten repräsentiert. So kann beispielsweise dargestellt werden, wie Signale in einer Zelle weitergeleitet und verarbeitet werden. Im Allgemeinen werden für die Modellierung biologischer Netzwerke unipartite Graphen genutzt. In dieser Arbeit wird jedoch eine andere Modellierung verwandt, worauf in Abschnitt 4.2 eingegangen wird. Deshalb werden hier nur die generellen Abstraktionsprinzipien der Orthologie und der Wissensgranularität sowie ihre Abbildung in Graphen dargestellt. Dadurch gelingt es, Daten aus verschiedenen Quellen und verschiedener Qualität zu integrieren, wodurch die im Netzwerk repräsentierten Informationen komplettiert werden.

## Orthologien

Beim ganzheitlichen, systembiologischen Blick auf Signaltransduktion in menschlichen Zellen, also der Betrachtung möglichst vieler Aspekte, fällt auf, dass Wissen oft unvollständig ist und einige Aspekte schlechter untersucht sind als andere. Ein Grund sind beispielsweise die limitierten Möglichkeiten zur experimentellen Untersuchung beim Menschen. Deshalb werden Modellorganismen eingesetzt, deren Stoffwechsel, Signaltransduktion und genetische Ausstattung mit dem Menschen vergleichbar sind (Campbell *et al.*, 2009). Zwei Modellorganismen für den Menschen (*Homo sapiens*) sind Maus (*Mus musculus*) und Ratte (*Rattus spp.*). Sie sind detaillierter untersucht und stehen in großen Mengen als Versuchstiere zur Verfügung.

Verwandtschaft und Vergleichbarkeit zwischen Mensch und Modell ermöglicht es beide zusammenzufassen. Dieses Konzept nennt man Orthologie. Es beschreibt eine funktionale Ähnlichkeit, von zum Beispiel einem Organ, zwischen verschiedenen Spezies (Fitch, 2000). Aus einer gleichen Funktion in den Modellorganismen wird eine gleiche oder ähnliche Funktion im Menschen gefolgert (Fitch, 1970; Tatusov, Koonin und Lipman, 1997; Fitch, 2000; Smith *et al.*, 2003). Generell ist eine Orthologie eine Spezialform der 1843 durch Owen geprägten Homologie, die eine gemeinsame Funktion unter einer Vielzahl an Bedingungen beschreibt (Owen, 1843; Fitch, 1970; Fitch, 2000). Mithilfe von Orthologien wird unbekanntes Wissen einer Spezies durch verfügbares Wissen im Modell komplettiert.

Beispielsweise wird ein Protein in mehreren verwandten Arten identifiziert, in der Maus ist es jedoch unbekannt. Durch die Orthologe Beziehung zwischen den Spezies wird angenommen, dass sich ein entsprechendes Protein auch in der Maus finden lässt. Dadurch helfen Orthologien, das Wissen verschiedener Spezies gemeinsam zu betrachten, wie Abbildung 3.3a für das in der Signalweiterleitung beteiligte Vav-Protein<sup>24</sup> zeigt. Dabei werden die verschiedenen Spezies zu einem speziesunspezifischen Vav-Knoten verallgemeinert.

---

<sup>24</sup>Vav katalysiert bei spezifischen Reaktionen, besonders bei jenen an denen Rac oder Rho-Kinasen beteiligt sind, den Austausch des gebundenen GDPs durch GTP (Bustelo, 2001).



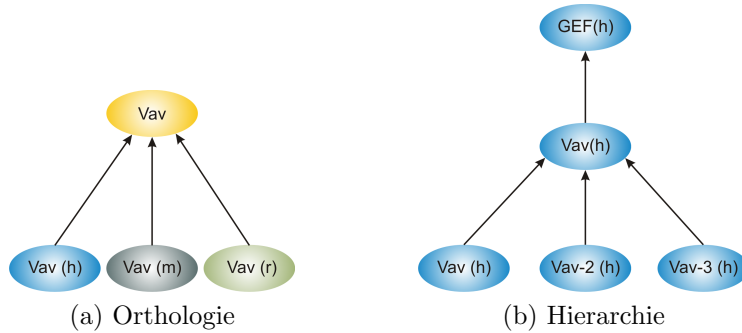


Abbildung 3.3: Orthologe und hierarchische Abstraktionsprinzipien. (a) Zusammenfassung verschiedener Modellorganismen (h= Mensch, m=Maus, r=Ratte) zu einem gemeinsamen speziesunspezifischen Knoten (gelb). (b) Hierarchie unterschiedlicher Wissensgranularitäten. Im gezeigten Beispiel bilden Vav - Vav-3 im Menschen (blaue Knoten) eine Proteinfamilie (Vav) die wiederum zu GEF verallgemeinert wird.(Abbildung nach Krull *et al.*, 2006)

### Wissensgranularität

Ein weiteres Problem ist die Granularität von vorhandenem Wissen. Unterschiedliche experimentelle Ansätze produzieren Ergebnisse in unterschiedlicher Qualität und Auflösung. In einem Organismus übernehmen beispielsweise zwei Proteine überwiegend gleiche Aufgaben. Dabei unterscheiden sie sich nur in wenigen Spezifika, die experimentell schwierig zu bestimmen sind. In einigen Experimenten liegt der Fokus auf der Unterscheidung der beiden Proteine, in anderen Experimenten ist diese Unterscheidung irrelevant. Die aus verschiedenen Experimenten gewonnenen Erkenntnisse unterliegen deshalb einer unterschiedlichen Auflösung und bilden eine implizite Hierarchie (siehe Abb. 3.3b). Durch Berücksichtigung dieser Unterschiede wird es möglich, unterschiedlichste Untersuchungsmethoden in einem Bild darzustellen. Ein gängiges Beispiel für Wissen unterschiedlicher Güte ist die RAS-Superfamilie, die in mehr als 150 einzelne Proteine zerfällt und eine Unterfamilie der GEFs<sup>25</sup> ist (Rojas *et al.*, 2012). Ebenfalls eine Teilgruppe der GEFs bildet die Vav-Familie, welche in Abbildung 3.3b dargestellt ist. Innerhalb der Spezies Mensch existieren drei unterschiedliche Vav-

<sup>25</sup>GEFs: Guanine nucleotide exchange factors, eine Proteinfamilie die GTPasen aktiviert, indem sie den Austausch von GDP zu GTP ermöglicht (Bourne, Sanders und McCormick, 1990; Alberts *et al.*, 2008)

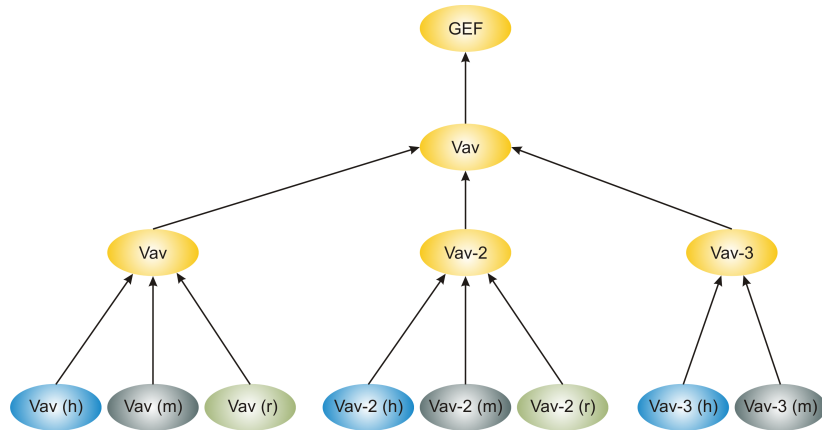


Abbildung 3.4: Zusammenfassung orthologer und hierarchischer Netzwerkabstraktionsprinzipien am Beispiel der Vav-Proteinfamilie. Die speziesspezifischen (h=Mensch, m=Maus, r=Ratte) Proteine werden ortholog zum speziesspezifischen Vav verallgemeinert, welches anschließend zur Vav-Familie und zu GEFs gruppiert. GEFs katalysieren den Austausch von GDP durch GTP während der Signaltransduktion (Bourne, Sanders und McCormick, 1990; Platanias und Fish, 1999; Alberts *et al.*, 2008). (Daten entnommen aus TRANSPATH<sup>®</sup> (3.1.2))

Formen, die zuerst zu einer Familie (Vav(h)) zusammengefasst werden. Anschließend werden sie zur Proteinfamilie GEF abstrahiert. Die Form der Abstraktion ist dabei auf den Menschen bezogen.

Wie man jedoch aus Abb. 3.3a sieht, werden durch orthologe Abbildungen auch Modellorganismen zu gemeinsamen, dann speziesspezifischen Gruppen zusammengefasst. Deshalb sind Orthologie und hierarchische Abbildungen kompatibel zueinander. Abbildung 3.4 zeigt, wie diese beiden Konzepte miteinander kombiniert werden. Dabei repräsentieren die farbcodierten Elemente (blau, grau, grün) die Spezieszugehörigkeit, während in gelb speziesspezifische Gruppierungen dargestellt werden. Das Wissen um den Menschen wird vervollständigt sowie kompakt und verständlich im Netzwerk abgebildet. Orthologe und hierarchische Abbildungen werden beispielsweise von TRANSPATH<sup>®</sup> und TRANSFAC<sup>®</sup> genutzt (Krull *et al.*, 2006).

## 4 Ergebnisse und Diskussion

Synergien werden im Folgenden über zwei verschiedene Methoden detektiert. Die erste Methode in Abschnitt 4.1 zeigt, wie Synergien grundsätzlich experimentell bestimmt werden. Der Fokus liegt dabei auf der Auswertung und Interpretation eines vorliegenden, durch den französischen Projektpartner erzeugten, experimentellen Datensatzes (siehe 3.4). Die Abschnitte 4.2 und 4.3 behandeln die Detektion von Synergien mittels einer algorithmischen Methode auf Basis eines rekonstruierten Netzwerks. Dazu wird in 4.2 dargestellt, wie das Netzwerk rekonstruiert wird und warum hierfür eine bipartite Modellierung zu bevorzugen ist (siehe Unterabschnitt 4.2.1). Außerdem wird beschrieben, wie verschiedene Wissensquellen integriert werden, wodurch ein genereller Überblick über die Signalweiterleitung gewährleistet werden konnte (siehe Unterabschnitte 4.2.2 und 4.2.3). Weiterhin wird gezeigt, wie spezifische Informationen dabei halfen, Zelltyp- und signalspezifische Netzwerke abzuleiten (siehe 4.2.4 und 4.2.5).

In Abschnitt 4.3 wird sukzessive eine allgemeine algorithmische Methode entwickelt, die auf Basis eines Netzwerks Synergien detektiert. Anschließend wird diese Methode auf das in 4.2.5 erstellte Netzwerk angewandt.

Zum Abschluss werden in Abschnitt 4.4 die beiden Methoden bezüglich ihrer generellen Herangehensweise (Unterabschnitt 4.4.1) sowie im Bezug der konkreten Ergebnisse (Unterabschnitt 4.4.2) miteinander verglichen.

### 4.1 Detektion von Synergien in pDC mittels Genexpressionsanalyse

Die Identifikation synergistischer Interaktionen stellt einen Spezialfall bei der Detektion differentiell exprimierter Gene dar (siehe 3.3.2). Dabei erzeugen verschie-

dene Stimuli zusammen Effekte, die sich von denen durch jeweils einen Stimulus erzeugten unterscheiden. Dies bedeutet, dass die Stimuli miteinander interagieren und einen neuen Effekt erzeugen, oder einen schon vorhandenen in unerwarteter Weise, abweichend von den Effekten jedes einzelnen Stimulus, modulieren.

In dieser Arbeit wurden synergistische Effekte untersucht, welche durch IL-3 und Flu (Influenza-A-Virus) in pDC vermittelt werden (siehe Unterabschnitt 2.2.3 und 2.2.4). Diese Zellen sind Bestandteil des Immunsystems und verbinden das angeborene mit dem adaptiven Immunsystem (Siegal *et al.*, 1999; Colonna, Trinchieri und Liu, 2004; Cao und Liu, 2007). Plasmacytoid dendritische Zellen dienen der Identifikation mikrobieller Pathogene und Stressoren. Außerdem zeigen sie Infektionsherde mittels inflammatorischer Signale an (siehe Unterabschnitt 2.2.2) (Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007; Kool *et al.*, 2009).

Jeder Stimulus alleine ist in der Lage, eine Immunantwort zu induzieren. IL-3 wirkt hauptsächlich als inflammatorisches Signal über TLR-unabhängige Signaltransduktionswege, wogegen Flu über TLR-abhängige Wege wirkt (Ghirelli, Zollinger und Soumelis, 2010). Dabei ist die Aktivierung von TLR-Signaltransduktionswegen Bestandteil der angeborenen Immunantwort, wogegen die durch IL-3 aktivierten Signalfade auf das erworbene Immunsystem einwirken (siehe 2.2.3 und 2.2.4).

Durch eine *limma*-Analyse (3.3.2) wurde anhand von Experimentaldaten untersucht, wie Signalintegration in pDC *en détail* funktioniert und wie angeborenes und adaptives Immunsystem miteinander verknüpft sind. Es war von besonderem Interesse, die synergistischen Interaktionen von TLR- mit Cytokin-abhängigen (IL-3) Signaltransduktionspfaden zu verstehen.

Im Experiment wurden, mittels Microarrays, Genexpressionsdaten für vier verschiedene Bedingungen bestimmt (siehe Unterabschnitt 3.4). Dabei wurden die beiden eingesetzten Stimuli IL-3 und Flu allein gemessen. In einer dritten Bedingung wurde untersucht, wie sich pDCs verhalten, wenn beide Stimuli parallel gegeben werden. Zuletzt wurde kontrolliert, wie sich die Zellen gänzlich ohne Stimulus (Medium) verhalten. Die Messung der Genexpression wurde in jeder Bedingung dreimal wiederholt (siehe Abschnitt 3.4).

### 4.1.1 Durch IL-3- und Flu- induzierte Synergie in pDC

Da Microarrays in Sonden unterteilt sind (siehe Unterabschnitt 3.3.1), wurden diese auch zuerst betrachtet. Zu Beginn wurden alle Sonden aus dem Datensatz entfernt, welche keine ausreichende Menge an mRNA binden konnten und deren korrespondierende Gene somit als nicht exprimiert galten (siehe Abschnitt 3.4). Anschließend wurden mittels *limma* (siehe 3.3.2) die differentiell exprimierten Gene<sup>26</sup>, anhand ihrer Sonden, in jeder der einzelnen Bedingungen (IL-3, Flu und Medium) im Vergleich zu IL-3+Flu bestimmt (siehe Abb. 4.1 (a) - (c)).

Ein beobachteter Effekt in der IL-3+Flu-Bedingung kann prinzipiell synergistischer Natur sein, oder er wird durch die Wirkung eines Stimulus bzw. einer nicht synergistischen Interaktion induziert. Um synergistische Interaktion der beiden Stimuli experimentell sicher zu detektieren und von den Effekten der einzelnen Stimuli abzugrenzen, mussten die Gene nicht nur in den einzelnen Bedingungen differentiell exprimiert sein, sie mussten zusätzlich im Vergleich der Bedingungen untereinander differentiell exprimiert sein (siehe Abb. 4.1d). Dadurch wurden Effekte, die nur durch einen Stimulus erzeugt worden, aber nicht synergistisch waren, ausgeschlossen. Um echte synergistische Effekte nachzuweisen war es außerdem notwendig, dass sich die Genexpression im Medium signifikant von IL-3+Flu unterschied. Erzeugten beide Stimuli ein Signal, welches sich während gleichzeitiger Gabe von IL-3 und Flu nicht mehr vom Medium unterschied, so wurde dies als nicht synergistisch gewertet. Dadurch wurde einem gemessenen Effekt im Vergleich von IL-3+Flu und Medium eine Bedeutung in der Signalweiterleitung zugeordnet.

Mittels Venn-Diagrammen wurden die Beziehungen der Bedingungen untereinander untersucht. Dazu wird in Abbildung 4.1 (d) zuerst die Beziehung der beiden Stimuli (IL-3, Flu) zueinander illustriert. In Abb. 4.2 werden zusätzlich noch die Beziehungen zum Medium dargestellt. Ein kompletter Kreis repräsentiert das gemessene Verhalten aller Sonden in einem bestimmten Vergleich. Der obere linke Kreis in Abb. 4.2 zeigt den Vergleich zwischen IL-3+Flu und IL-3 und entspricht somit der Einzelbetrachtung in Abbildung 4.1a. Die Anzahl der Sonden

---

<sup>26</sup>Die Gene werden über ihre spezifischen Sonden bestimmt. Diese messen über eine für ein Gen spezifische, kurze Basenpaarsequenz, anhand der komplementären Bindung von mRNA-Transkripten an dieser Sequenz, die Genexpression. Sonden dienen dadurch als Identifikatoren für Gene (siehe 3.3.1)

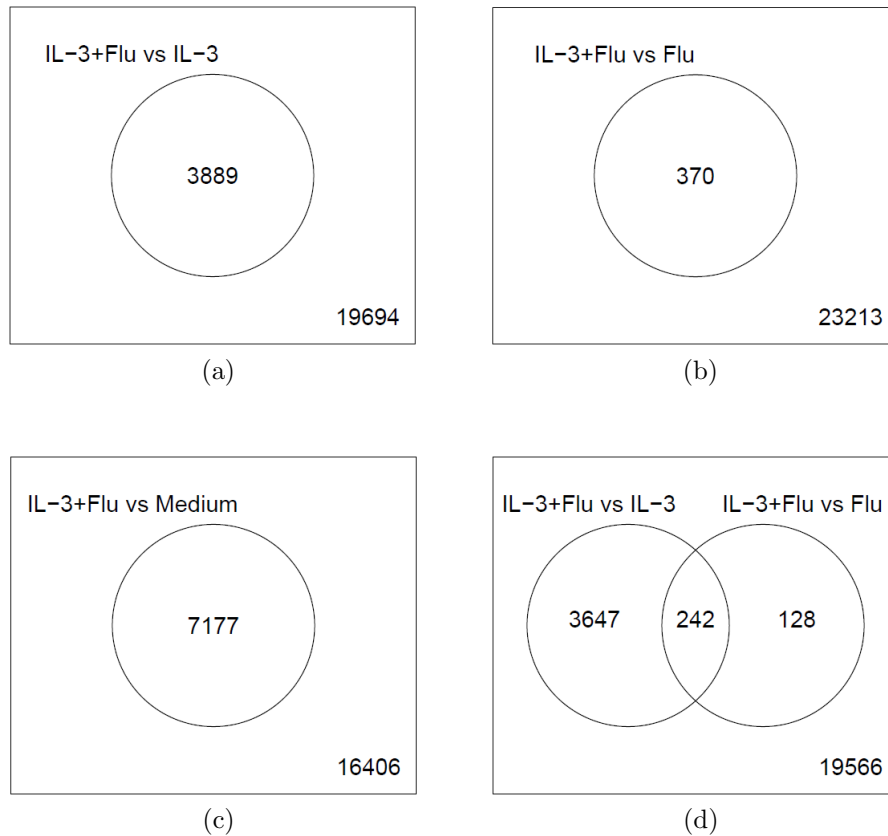


Abbildung 4.1: Darstellung der differentiellen Genexpression anhand der auf einem Microarray aufbrachten Sonden zwischen verschiedenen experimentellen Bedingungen: (a) zwischen IL-3+Flu und IL-3; (b) zwischen IL-3+Flu und Flu; (c) zwischen IL-3+Flu und Medium. (d) Sonden, die differentiell exprimierte Gene detektieren, welche sich von IL-3+Flu in beiden Einzelbedingungen (IL-3 (a), Flu (b)) unterscheiden. Die Überlappung zeigt Sonden, welche Gene messen, die in beiden Vergleichen unterschiedlich exprimiert sind.

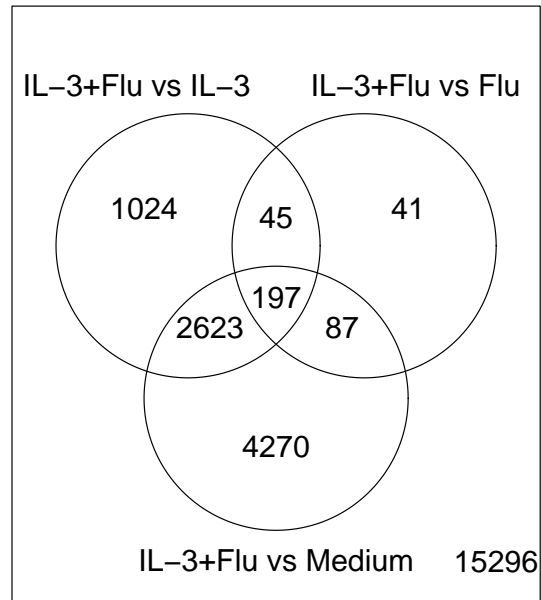


Abbildung 4.2: Darstellung der Verteilung der signifikant differentiell exprimierten Gene durch ihre Sonden in den verschiedenen Vergleichen mittels eines Venn-Diagramms. Jeder Kreis zeigt den Vergleich von IL-3+Flu mit einer weiteren experimentellen Bedingung (oben links: mit IL-3). Am rechten unteren Rand ist angegeben, wie viele Sonden im Datensatz nie differentiell exprimierte Gene nachwies. Schnittmengen geben an, wie viele Sonden in den jeweiligen Vergleichen differentiell exprimierte Gene bestimmen. Das Zentrum zeigt die 197 Sonden, anhand derer synergistisch regulierte Gene detektiert wurden. Die insgesamt kleine Menge an Sonden, die differentiell exprimierte Gene im Vergleich von IL-3+Flu und Flu identifizieren, zeigt, dass Flu einen stark dominanten Einfluss gegenüber dem IL-3 Stimulus besitzt (siehe auch Abb. 4.1).

innerhalb des kompletten Kreises stellt die Menge der als differentiell exprimiert induziert anzusehenden Sonden dar. Abbildung 4.1a zeigt, dass im Vergleich von IL-3+Flu und IL-3 durch 3889 Sonden differentiell exprimierte Gene gemessen wurden. Der Vergleich von IL-3+Flu und Medium (Abb. 4.1c) detektierte an 7177 Sonden differentielle Genexpression, wogegen der Vergleich von IL-3+Flu mit Flu (Abb. 4.1b) nur an 370 Sonden differentielle Genexpression im Vergleich zu IL-3+Flu detektierte. 15296 Sonden zeigten beim Vergleich aller Bedingungen keine Reaktion auf mindestens einen der beiden Stimuli (siehe Abb. 4.2).

Die Schnittmengen stellen die Mengen an Sonden dar, an welchen differentiell exprimierte Gene zwischen beiden Vergleichsbedingungen gemessen worden. Zum Beispiel konnten 242 Sonden identifiziert werden, die sich signifikant zwischen den beiden Einzel-Stimulusbedingungen unterschieden (siehe Abb. 4.1d). Berücksichtigt man zusätzlich noch den Unterschied zur IL-3+Flu vs. Medium-Bedingung (Abb. 4.2) so zeigten 45 Sonden keine Änderung im Medium, wogegen 197 Sonden eine Änderung der Genexpression in allen drei Vergleichen anzeigten. Diese 197 Sonden im Zentrum von Abbildung 4.2 sind für die synergistisch induzierten Effekte von besonderem Interesse, da sie in jeder Bedingung differentiell exprimiert waren und sich in jedem der durchgeführten Vergleiche unterschieden. Das heißt einerseits, dass jeder Stimulus alleine einen Effekt erzeugt, andererseits sich dieser aber vom Effekt bei gleichzeitiger Gabe der Stimuli unterscheidet. Folglich existieren Schnittpunkte, an denen die aktivierten Signaltransduktionskaskaden miteinander interferieren. Dies führt zur Modulation des vermittelten Signals, was zur Änderung der Genexpression führt.

Überraschenderweise konnte im Vergleich von IL-3+Flu und Flu (Abb. 4.1b) nur an 370 Sonden differentielle Expression nachgewiesen werden. Dies bedeutet, dass zwischen den beiden Bedingungen nur kleine Unterschiede vorlagen. Es ist bekannt, dass Flu die Immunabwehr stark aktiviert, wodurch sich die Mehrzahl der messbaren Effekte in IL-3+Flu auf Flu vermittelte Signale zurückführen lässt. Anders ausgedrückt heißt es, dass Flu auf pDC einen stärkeren Einfluss ausübt als IL-3. Wenn beide Stimuli gemeinsam gegeben werden, ist Flu dominant gegenüber IL-3. Es hat also einen größeren Einfluss auf die Änderung der Genexpression als IL-3. Der Grund dafür ist, dass Flu als Stimulus des angeborenen Immunsystems auf viele verschiedene Signalpfade wirkt. Interaktionen mit den Signalpfaden, ausgelöst durch IL-3, sind spezifischer und oftmals *downstream*



der Flu-Signalfade. Dadurch wird durch IL-3 in vielen Fällen nur eine modulierende Wirkung auf bereits durch Flu-aktivierte Signalfade realisiert (Guiducci *et al.*, 2006; Guiducci *et al.*, 2008).

Die identifizierten 197 synergistisch beeinflussten Sonden wurden auf 73 unterschiedliche Gene abgebildet. Dabei wurden Gene, die durch verschiedene Sonden repräsentiert waren, durch den Mittelwert aller Sonden zusammengefasst. Diese Gene zeigten eine signifikante Veränderung ihrer Expressionsraten nach gemeinsamer Gabe von IL-3 und Flu (IL-3+Flu). Abb. 4.3 zeigt die Differenzen dieser 73 Gene im Vergleich zur Negativkontrolle (Medium). Es ließ sich feststellen, dass innerhalb der synergistisch beeinflussten Gene weder ein Trend zur Hoch- noch zur Herunter-Regulation vorlag. Durch Betrachten der generellen Regulationsmuster wurde aufgeklärt, ob die Stimuli beispielsweise aktivierend auf die Zelle einwirkten, wobei z.B. viele Gene stärker exprimiert wurden. Als Maß für eine starke Regulation wurde ein Schlauch angenommen. Der Mittelpunkt des Schlauchs wurde durch die mittlere Änderung aller 73 Gene definiert. Die Außenseiten wurden durch die Standardabweichung (Mittelwert  $\pm$  Standardabweichung) definiert. Alle Gene, die außerhalb dieses Schlauchs lagen, wurden als stark hoch- (oberhalb des Schlauchs) oder stark herunter- (unterhalb des Schlauchs) reguliert angenommen und sind besonders interessant, da sie vermutlich die wesentlichen immunrelevanten Effekte vermitteln. Insgesamt wurden jeweils 11 Gene stark hoch- oder herunterreguliert. Diese Gene werden in Abb. 4.3 mit ihren Namen dargestellt. Eine Präferenz für einen Regulationstyp ließ sich nicht ableiten, weder unter Berücksichtigung der Gene außerhalb des Schlauchs, noch unter der Betrachtung aller Gene. Innerhalb des kompletten Satzes (73 Gene) waren insgesamt 40 Gene hoch- und 33 herunterreguliert. Gegen die Existenz eines generellen Trends spricht auch der errechnete Mittelwert. Dieser lag alle Gene zusammengekommen bei -0,020. Sollte ein Trend existieren, so wäre zu erwarten, dass er stärker in eine Richtung abweicht<sup>27</sup>. Da die beiden verwendeten Stimuli (IL-3, Flu) in der Literatur als Aktivatoren des Immunsystems beschrieben werden, war dies auch für die beeinflussten Gene zu erwarten (Shelton *et al.*, 2004; Chan *et al.*, 2005; Liu, 2005; Takeda und Akira, 2005; Cao und Liu, 2007; Hoeve *et al.*, 2012). Für die vorliegenden Daten konnte die Erwartung einer generell gesteigerten Expression von Genen jedoch nicht bestätigt werden.

---

<sup>27</sup>Ein zweiseitiger Ein-Stichproben-t-test gegen  $\mu = 0$  zeigt, mit  $p > 0,05$ , dass der Mittelwert nicht von 0 abweicht.

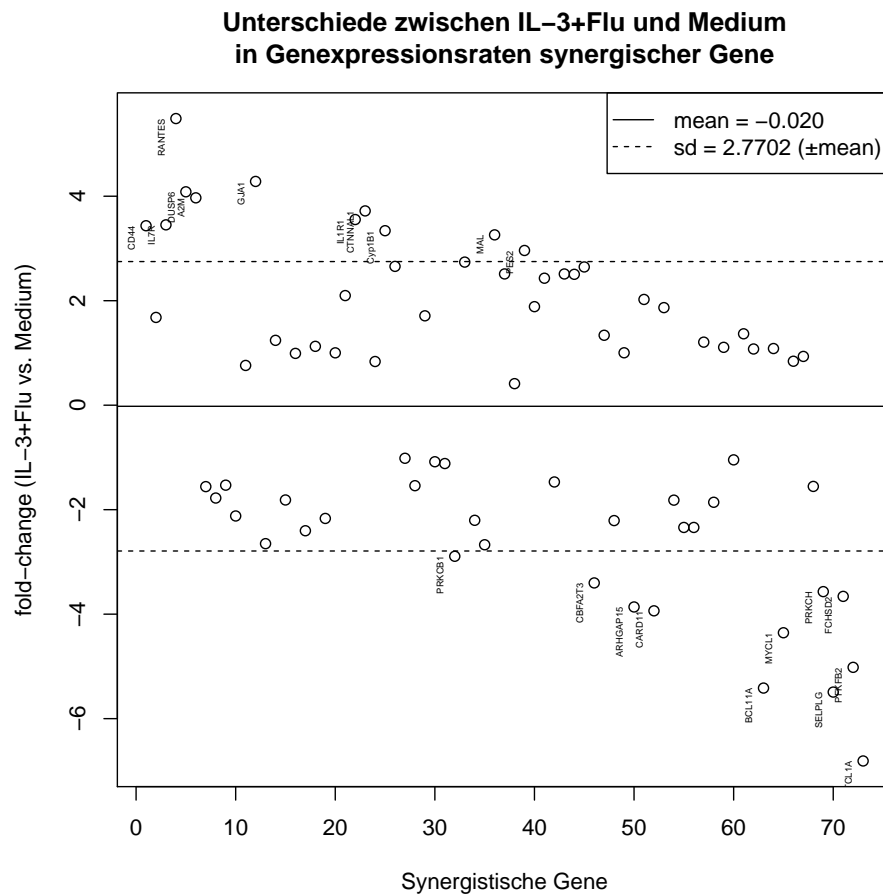


Abbildung 4.3: Die 73 detektierten synergistisch regulierten Gene (197 Sonden) werden nummeriert an der x-Achse dargestellt. Die y-Achse zeigt die Differenz der experimentell bestimmten Genexpression zwischen IL-3+Flu und Medium (Kontrolle). Werte größer 0 zeigen hochregulierte Gene; Werte kleiner 0 zeigen herunterregulierte Gene. Der Mittelwert (mean) der Differenzen aller Gene wird durch eine geschlossene Linie und die Standardabweichung (sd) durch gepunktete Linien dargestellt. Die 22 Gene, die besonders stark zwischen beiden Bedingungen variieren ( $\leq$  (Mittelwert  $\pm$  sd)), werden durch ihre Namen dargestellt. Dabei sind 11 verschiedene Gene hoch- wie auch herunterreguliert.

### 4.1.2 Funktionen synergistischer Gene

Die 73 als synergistisch identifizierten Gene verhalten sich signifikant unterschiedlich, wenn beide Stimuli (IL-3+Flu) anwesend sind, im Vergleich zu jeder anderen Bedingung (IL-3, Flu, Medium). Dies kann durch Interaktionen zwischen IL-3 und Flu erklärt werden. Dabei spielt jeder Stimulus alleine eine wichtige Rolle bei der Immunantwort in pDC (siehe 2.2.3 und 2.2.4). IL-3 wie auch Flu aktivieren die Zelltypkonversion von pDC zu DC. Über die Freisetzung von Interferonen sowie weiteren Cytokinen wird die Immunantwort gesteuert und es werden weitere Zellen für die Bekämpfung des Pathogens rekrutiert (Janeway *et al.*, 2001; Martinez-Moczygema und Huston, 2003; Liu, 2005; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008). Für die identifizierten synergistischen Gene ist deshalb zu erwarten, dass sie mit weiteren Immunprozessen oder spezifischer Immunantwort verschränkt sind. Durch Mappierung der identifizierten Gene an ihre *in-vivo*-Funktionen wurde dieser Zusammenhang überprüft.

KEGG-Datenbanken bieten eine gute Quelle um die Funktionen von Genen in einem Organismus zu überprüfen (siehe Unterabschnitt 3.1.3 sowie Kanehisa *et al.*, 2008; Zhang und Wiemann, 2009). Dazu wurde das Datenbankmodul KEGG-*GENES* (siehe 3.1.3) ausgewertet. Eine definierte Liste an Beschreibungen, welche mit Immunfunktionen assoziiert sind, wurde mit verschiedenen Feldern innerhalb des Datenbankmoduls wie dem *pathway*-<sup>28</sup>, dem *class*-<sup>29</sup> oder dem *definition*-Eintrag<sup>30</sup> abgeglichen. Die Liste enthielt immunrelevante Begriffe, wie „immune“, „B-cell“ oder „T-cell“ die eine direkte Beziehung zum Immunsystem herstellen (Janeway *et al.*, 2001). Außerdem wurde eine Immunfunktion angenommen, wenn Signalfade aktiviert werden, von denen bekannt ist, dass sie durch einen der beiden Stimuli beeinflusst werden. Dies erweiterte die Suchliste

<sup>28</sup> „Der *pathways*-Eintrag beschreibt funktionale Beziehungen für ein Protein im Kontext molekularer Netzwerke.“ Aus: [http://www.genome.jp/kegg/document/help\\_bget\\_genes.html](http://www.genome.jp/kegg/document/help_bget_genes.html)

<sup>29</sup> „Der *class*-Eintrag beschreibt eine hierarchische Klassifikation von Gen-Funktionen nach dem in KEGG genutzten allgemeinen KO (KEGG-Orthology) Klassifikationssystem“ (siehe Kanehisa *et al.*, 2006; Kanehisa *et al.*, 2008) „und verbindet die Module KEGG-*PATHWAY* mit KEGG-*BRITE* und generalisierten Funktionen, beispielsweise anhand von Protein Familien.“ Aus: [http://www.genome.jp/kegg/document/help\\_bget\\_genes.html](http://www.genome.jp/kegg/document/help_bget_genes.html)

<sup>30</sup> „Der *definition*-Eintrag liefert für ein Gen oder sein Produkt die Annotation aus den originalen, gegebenenfalls externen Datenbanken.“ Aus: [http://www.genome.jp/kegg/document/help\\_bget\\_genes.html](http://www.genome.jp/kegg/document/help_bget_genes.html)

um Terme wie: „JAK-STAT“, „Toll-like“ und „cytokine“. Konnte mittels KEGG-*GENES* zwischen einem detektierten Gen und einem der Terme aus der Liste ein Zusammenhang hergestellt werden, so wurde für das Gen eine Bedeutung in der Immunantwort gefolgert. Alle anderen Gene wurden als nicht immunrelevant betrachtet.

Das Anbinden der Gene an ihre Funktion (siehe Tab. 4.1 und Tab. 5.1 (Appendix)) zeigt, dass von insgesamt 73 identifizierten Genen nur 22 ( $\approx 30,14\%$ ) an eine Immunfunktion gekoppelt werden konnten. Die Mehrheit der Gene (51,  $\approx 69,86\%$ ) war nicht mit einer stimulusspezifischen Immunreaktion verbunden. Für die ebenfalls 22 am stärksten regulierten Gene und benannten Gene aus Abb. 4.3 wird ihre zugehörige Funktion in Tab. 4.1 dargestellt, wobei nur 9 immunrelevante Gene detektiert wurden. Bei diesen neun positiv identifizierten immun-relevanten Genen in Tab. 4.1 sind drei verschiedene Gene (IL1R1, IL7R, RANTES) an weiteren Cytokin-Signalwegen beteiligt. Cytokin-abhängige Signaltransduktionswege werden von verschiedenen Stimuli aktiviert, unter anderem von IL-3 (Arai *et al.*, 1990; Martinez-Moczygemba und Huston, 2003; Lopez *et al.*, 2010). Jedoch werden über derartige Signalpfade auch andere Interleukine perzipiert, wie etwa IL-6, was eine wichtige Rolle in der Reaktion auf inflammatorische Ereignisse spielt (Martinez-Moczygemba und Huston, 2003; Chan *et al.*, 2005; Pirog, Stabinsky und Goldman, 2010; Ghirelli, Zollinger und Soumelis, 2010; Gupta *et al.*, 2011).

Die Gene BCL11A, CARD11, PRKCB1 spielen eine Rolle in der Kommunikation mit B-Zellen. Diese Zellen sind Bestandteil des adaptiven Immunsystems und wichtig, um Pathogene langfristig identifizieren und bekämpfen zu können (Janeway *et al.*, 2001; Lenert, 2006). Hier zeigt sich, dass angeborene und erworbene Immunabwehr miteinander verschränkt sind und sich gegenseitig regulieren.

Der Begriff „hematopoietic cell lineage“ war nicht Bestandteil der Suchbegriffe, konnte aber auf detektierte Gene wie IL1R1, IL7R und CD44 mappiert werden. Immunzellen wie DC und pDC stammen von dieser Zelllinie ab (Martinez-Moczygemba und Huston, 2003; Shelton *et al.*, 2004; Lopez *et al.*, 2010). Dieser Term ist somit die Bestätigung der Beziehungen zwischen den Stimuli (IL-3 und Flu) und pDC sowie ihrer Rolle in der Vermittlung der Immunantwort (Martinez-Moczygemba und Huston, 2003; Lopez *et al.*, 2010). Außerdem wurde durch

Gen	IL-3+Flu	Medium	IL-3+Flu - Medium	immun- relevant	Funktion
TCL1A	1.4985	8.3081	-6.8096	+	T-cell leukemia/lymphoma 1A
SELPLG	2.2146	7.7057	-5.4911	-	Cell adhesion molecules (CAMs) Staphylococcus aureus infection
RANTES	7.43847	1.95127	5.4872	+	Cytokine-cytokine receptor interaction Toll-like receptor signaling pathway Immune System
BCL11A	2.8585	8.27262	-5.41412	+	B-cell CLL/lymphoma 11A (zinc finger protein)
PFKFB2	1.9497	6.9672	-5.0175	-	Fructose and mannose metabolism
MYCL1	2.5793	6.9348	-4.3555	-	v-myc myelocytomatosis viral oncogene homolog 1
GJA1	6.1997	1.9161	4.2836	-	Cell Communication; Gap junction Human Diseases; Cardiovascular Diseases;
DUSP6	6.91447	2.8293	4.08517	-	MAPK signaling pathway Transcriptional misregulation in cancer
A2M	6.7297	2.76	3.9697	+	Immune System Complement and coagulation cascades
CARD11	3.5858	7.5216	-3.9358	+	T cell receptor signaling pathway B cell receptor signaling pathway
ARHGAP15	3.6276	7.4894	-3.8618	-	Rho GTPase activating protein 15
CTNNA1	5.4314	1.712	3.7194	-	catenin (cadherin-associated protein), alpha-like 1
FCHSD2	2.1038	5.763	-3.6592	-	FCH and double SH3 domains 2
PRKCH	2.3179	5.886	-3.5681	-	Vascular smooth muscle contraction Tight junction Cell Communication;
IL1R1	5.4976	1.9435	3.5541	+	Hematopoietic cell lineage Cytokine-cytokine receptor interaction MAPK signaling pathway Immune System
IL7R	7.9177	4.4649	3.4528	+	Cytokine-cytokine receptor interaction Jak-STAT signaling pathway Hematopoietic cell lineage
CD44	8.1477	4.712	3.4357	+	ECM-receptor interaction Hematopoietic cell lineage Immune System
CBFA2T3	3.9316	7.3333	-3.4017	-	core-binding factor, alpha subunit 2
Cyp1B1	5.3606	2.02	3.3406	-	Steroid hormone biosynthesis Tryptophan metabolism Metabolism of xenobiotics by cytochrome P450
MAL	4.8198	1.5592	3.2606	-	not found
PES2 (PTGS2)	4.639	1.6771	2.9619	-	Metabolic pathways Pathways in cancer VEGF signaling pathway
PRKCB1	4.9478	7.83945	-2.89165	+	B cell receptor signaling pathway Pathways in cancer MAPK signaling pathway

Tabelle 4.1: Namentliche Darstellung der 22 Gene (Abb. 4.3), deren gemessenes Expressionslevel sich am stärksten zwischen IL-3+Flu und Medium unterscheidet. Ein „+“ in der Spalte „immunrelevant“ bedeutet, dass ein Gen mit einer Bedeutung in der Immunantwort annotiert ist. Die letzte Spalte zeigt über welche Signalwege und Funktionen das Gen in die Immunantwort involviert ist. Ein Gen ist in seiner Funktion immunrelevant, wenn ihm mindestens ein Term aus einer Liste mit immunrelevanten Begriffen zugeordnet wurde. Diese Liste enthält Begriffe wie: „immune“, „cytokine“, „Toll-like“, „JAK-STAT“, „B-cell“ or „T-cell“. Werden einem Gen verschiedene Funktionen zugeordnet, so werden bis zu drei verschiedene Funktionen dargestellt. Von den 22 benannten Genen (Abb. 4.3) sind neun mit der Immunantwort verbunden ( $\approx 40,91\%$ ). Im gesamten Datensatz (73 Gene) sind  $\approx 30.14\%$  der Gene (22) mit der Immunantwort verbunden (siehe dazu Tab. 5.1 im Appendix).

die Verknüpfung zwischen immunrelevanten Genen und ihrer Herkunft aus der hämatopoietischen Zelllinie gezeigt, dass durch die Stimuli eine Antwortkaskade in pDC angestoßen wird, die schließlich die Zelltypkonversion zu DC induziert (Grouard *et al.*, 1997; Colonna, Trinchieri und Liu, 2004; Cao und Liu, 2007; Ghirelli, Zollinger und Soumelis, 2010; Hoeve *et al.*, 2012).

Gene, die nicht an eine Funktionalität in der Immunantwort gebunden werden konnten, zeigen eine große Diversität in ihrer Funktion. SELPG zum Beispiel wird als Gen beschrieben, welches das Haften von Zellen aneinander beeinflusst. PFKFB2 ist annotiert als relevant in der Verstoffwechslung von Fruktose und Mannose, wogegen Cyp1B1 eine Rolle in der Steroid-Biosynthese und dem Tryptophan-Metabolismus zukommt. GJA1 und PRKCH werden generell als Bestandteil der Zellkommunikation gesehen, sind aber an das vaskuläre System anstatt an die Immunantwort gekoppelt. Freigesetzte Cytokine und Interferone beeinflussen andere Immunzellen (z.B. B-Zellen, T-Zellen) und rekrutieren sie zum Infektionsherd (Arai *et al.*, 1990; Janeway *et al.*, 2001; Martinez-Moczygemba und Huston, 2003; Cao und Liu, 2007; Lopez *et al.*, 2010). Eine Beeinflussung des vaskulären Systems (Blutgefäßsystems) kann sinnvoll sein, um einen schnelleren Transport von Signalen oder Zellen zum Infektionsort zu gewährleisten. Eine direkte Verknüpfung mit der Immunabwehr konnte jedoch nicht hergestellt werden.

### **Synergien und Krankheiten**

Interessanterweise sind IL1R1, PRKCB1 und DUSP6 in KEGG-*GENES* mit einer Beziehung zu MAPK-Signalpfaden annotiert. In der Literatur ist ebenfalls eine Verschränkung von IL-3- oder TLR-aktivierten Signalkaskaden mit PI3K-abhängiger Signaltransduktion beschrieben, wobei PI3K-abhängige Pfade auf MAPK-Pfade einwirken (Rodriguez-Viciano *et al.*, 1997; Shelton *et al.*, 2004; Saldeen und Welsh, 2004; Guiducci *et al.*, 2008; Zhu und Bertics, 2011). Die Verschränkung von Pfaden kann deshalb eine Fernwirkung der Immunantwort darstellen. Die genauen Zusammenhänge sind jedoch nur wenig erforscht. DUSP6 beispielsweise ist in der transkriptionellen Fehlregulation verschiedener Krebsarten, wie etwa Lungenkrebs und ihrer Suppression involviert (und annotiert) (Furukawa *et al.*, 2003; Okudela *et al.*, 2009). Andere identifizierte Gene,

wie TCL1A, BCL11A, PRKCB1, oder auch detektierte und nicht immunrelevante Gene, wie MYCL1 und PES2, werden ebenfalls mit Krebs in Verbindung gebracht (siehe Tab. 4.1).

Die Literatur beschreibt für Gene, die direkt oder indirekt mit der Immunantwort assoziiert werden, verschiedene Erkrankungen. Die Dysregulation solcher Gene löst Allergien, Alzheimer, andere Autoimmunerkrankungen oder auch Krebs aus (Martinez-Moczygemba und Huston, 2003; Zambrano *et al.*, 2007; Gilliet, Cao und Liu, 2008; Kool *et al.*, 2009; Cohen, 2009; Swardfager *et al.*, 2010). Veränderungen in den PI3K-Signaltransduktionspfaden durch IL-3 oder Flu (TLR-Signaltransduktionspfad) beeinflussen Apoptose sowie weitere Überlebensprogramme in Zellen (cell survival) (Songyang *et al.*, 1997; Dudek *et al.*, 1997; Gaur und Aggarwal, 2003; Shelton *et al.*, 2004; Youle und Strasser, 2008; Chipuk *et al.*, 2010). Fehlfunktionen, besonders in der Regulation von Apoptose, sind häufig eine Ursache für die Entstehung von Krebs (Frank, 1999; Wakefield und Roberts, 2002; Shelton *et al.*, 2004; Pikarsky *et al.*, 2004). Zum Beispiel ist für Leukämie bekannt, dass IL-3 die Tumorentwicklung beeinflusst (Daley und Baltimore, 1988; Wakefield und Roberts, 2002; Shelton *et al.*, 2004; Pikarsky *et al.*, 2004; Youle und Strasser, 2008; Liu *et al.*, 2011; Gupta *et al.*, 2011).

Die Verknüpfungen zwischen Immunsystem und z.B. Alzheimer sind noch nicht gut erforscht (Zambrano *et al.*, 2007; Cohen, 2009; Zambrano *et al.*, 2010). In jüngerer Zeit mehren sich Hinweise, dass das Immunsystem bei der Entstehung von Alzheimer, aber auch anderen psychischen Erkrankungen, wie Schizophrenie, beteiligt ist (Zambrano *et al.*, 2007; Chen und Kendler, 2008; Cohen, 2009; Zambrano *et al.*, 2010).

Die Vielzahl an Verknüpfungen zu anderen Funktionalitäten und Krankheiten zeigt, dass (a) in einem Organismus immer alles mit allem vernetzt ist, und dass (b) zur Identifikation von Synergien in der Immunantwort die klassische, rein experimentelle Methodik zu ungenau ist. Permanente Anpassungsprozesse in Zellen erfordern auch eine kontinuierliche Anpassung der Expressionsraten. Dabei werden, beispielsweise im Rahmen der Zellzyklusregulation, auch Gene reguliert, die nicht direkt von IL-3 und Flu beeinflusst werden. Deshalb ist die Untersuchung von Synergien nur mit Hilfe von quantitativen Genexpressionsdaten zu ungenau und liefert nur einen ersten Überblick. Somit erklärt sich auch, wieso von 73 detektierten Genen nur 22 ( $\approx 30.14\%$ ) in der Immunantwort relevant sind.

## Proteomdaten verbessern das Verständnis von Synergien

Für die Prozessierung von Signalen spielt jedoch die Proteinebene die wichtigste Rolle. Hier interagieren Signale miteinander, werden moduliert und integriert (Cox und Mann, 2007; Han, Aslanian und Yates, 2008). Erst am Ende einer Signalkaskade befinden sich Gene. Diese werden als Ergebnis der Signaltransduktion beeinflusst und induzieren die Anpassung der Zelle an die Umgebung. Dies geschieht durch die Änderung der Menge von Proteinen oder durch Produktion neuer Proteine (Alberts *et al.*, 2001). Es ist jedoch schwer anhand der Menge gemessener mRNA-Transkripte auf die Menge an gebildetem Protein zu schließen (Pratt *et al.*, 2002; Ingolia *et al.*, 2009). Posttranslationale Modifikationen<sup>31</sup> und verschiedene Protein-turnover-Raten<sup>32</sup> sind nur zwei Beispiele, die zeigen, wie schwierig es ist, von gemessener genomischer Genexpression auf die Abundanz von Proteinen zu schließen (Pratt *et al.*, 2002; Cox und Mann, 2007; Han, Aslanian und Yates, 2008; Selbach *et al.*, 2008; Shyu, Wilkinson und Hoof, 2008; Ingolia *et al.*, 2009). Obwohl in jeder somatischen Zelle eines biologischen Organismus gleiches genomisches Material vorliegt, variiert das Proteom<sup>33</sup> einer Zelle nach verschiedenen Parametern wie Zelltyp und Funktion (Han, Aslanian und Yates, 2008). Um Synergien im proteomischen Signaltransduktionsnetzwerk zu verstehen und ihre Funktionen bei der Immunantwort zu untersuchen, ist ein besseres Verständnis der zugrunde liegenden Prozesse erforderlich.

Ein proteomischer Ansatz gewährt tiefere Einblicke, wie Signale interagieren, wie sich Synergie entwickelt und welche Prozesse auf Proteinebene ablaufen. Eine Methode zur Untersuchung ist die Massenspektrometrie (Aebersold und Mann, 2003; Cravatt, Simon und Yates, 2007; Han, Aslanian und Yates, 2008; Nilsson *et al.*, 2010). Diese Methodik und ihre Anwendung bei der Aufklärung des gesamten Proteoms einer Zelle hat sich in den letzten Jahren fortwährend verbessert. Weitere Fortschritte werden eine flächendeckende Anwendung auf systembiologische Fragestellungen ermöglichen (Nilsson *et al.*, 2010). Dennoch ist die Reprodu-

---

<sup>31</sup>Posttranslationale Modifikationen werden am fertigen Proteinprodukt vorgenommen. Sie können aktivierender wie inaktivierender Natur sein oder einen Einfluss auf die Sekundär- und Tertiärstruktur des Proteins haben.

<sup>32</sup>Turnover-Raten bestimmen, wie schnell ein Protein durch ein Neues ersetzt wird und wie schnell es verbraucht wird. Sie variieren je nach Notwendigkeit des Proteins oder auch nach Wiedergebrauchsfähigkeit.

<sup>33</sup>Das Proteom ist die Gesamtheit der in einer Zelle verfügbaren Proteine.



zierbarkeit bei der Aufklärung des Gesamtproteoms von Zellen nicht in jedem Fall gegeben. Oft sind Ergebnisse nur innerhalb eines Labors reproduzierbar. Diese Variabilität und Unsicherheit der gewonnenen Erkenntnisse ist noch zu hoch und muss weiter verbessert werden (Cravatt, Simon und Yates, 2007; Addona *et al.*, 2009; Nilsson *et al.*, 2010). In Zukunft können massenspektrometrische Ansätze in Form von Hochdurchsatzexperimenten bei der proteomweiten Aufklärung von Signaltransduktion helfen. Dadurch wird Signalweiterleitung und Signalverarbeitung insgesamt besser verstanden und es wird aufgeklärt, wie einzelne Prozesse miteinander interagieren.

### 4.1.3 Zusammenfassung

Mit Hilfe einer *limma*-Analyse (siehe Abschnitt 3.3.2) wurden 197 verschiedene Sonden (73 Gene) identifiziert, die in synergistischer Art und Weise von den Stimuli IL-3 und Flu beeinflusst werden. IL-3, wie auch Flu, dienen als Signalgeber für das Immunsystem (Martinez-Moczygema und Huston, 2003; Lenert, 2006; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008). Durch die angestoßenen Signalpfade werden in pDC Infektionsstellen markiert sowie eine direkte Immunantwort initiiert. Außerdem werden weitere Zellen des Immunsystems zur Infektionsstelle rekrutiert (Liu, 2005; Cao und Liu, 2007). Die Untersuchung der hoch- oder herunter-Regulation der experimentellen Ergebnisse zeigte kein spezifisches Muster. Bei der Genregulation existierte keine Tendenz in eine bestimmte Richtung, die meisten Gene waren weder besonders stark hoch- oder herunterreguliert.

Im Vergleich von IL-3+Flu und Flu wird signifikant unterschiedliche Expression nur von 370 Sonden detektiert (siehe Abb. 4.1b). Im Gegensatz dazu zeigen im Vergleich von IL-3 mit IL-3+Flu 3889 Sonden differentielle Expression. Daraus wird gefolgert, dass Flu im direkten Vergleich mit IL-3 der stärkere und somit dominante Stimulus ist.

Beide Stimuli lösen die Immunantwort und die Zelltypkonversion von pDC zu DC aus. Dennoch kann in vielen Fällen aufgrund der Dominanz von Flu davon ausgegangen werden, dass IL-3 die Flu-abhängigen Signale moduliert und wenige neue Signale hinzufügt. Somit ist Flu auch deshalb der dominante Stimulus, weil die von IL-3 regulierten Signalpfade *downstream* liegen und von IL-3 sowie

Flu aktiviert werden. Dadurch entfaltet IL-3 nur noch modulierende Wirkungen (Guiducci *et al.*, 2006; Guiducci *et al.*, 2008).

Flu als virales Pathogen aktiviert das angeborene Immunsystem. Das Cytokin IL-3 markiert Infektionsstellen wobei durch IL-3-abhängige Effekte weitere Immunzellen zum Infektionsort rekrutiert werden. Beides ist für ein funktionsfähiges Immunsystem unerlässlich und stark miteinander verschränkt. Dadurch wird eine effiziente Gegenwehr gegen mikrobielle Pathogene gewährleistet (Jarrossay *et al.*, 2001; Janeway *et al.*, 2001; Colonna, Trinchieri und Liu, 2004; Shelton *et al.*, 2004; Chan *et al.*, 2005; Liu, 2005; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008).

Eine Verknüpfung von synergistisch beeinflussten Genen mit ihrer Funktion in der Immunantwort war erwartbar. Jedoch zeigt die Mappierung von identifizierten Genen auf ihre Funktionen nur in 22 Fällen ( $\approx 30,14\%$ ) eine Verbindung zu immunrelevanten Aufgaben. Tabelle 4.1 stellt die Funktionen für die Gene, die am stärksten synergistisch beeinflusst werden, dar. Dabei ist die Mehrheit der identifizierten Gene nicht in die Immunantwortkaskade eingebunden (siehe Tab. 4.1). Diese Gene zeigen jedoch eine hohe Diversität in ihren Funktionen. Sie sind beispielsweise im Tryptophan- (Cyp1B1) oder im Fruktose-/Mannose-Metabolismus (PFKFB2) involviert. Außerdem werden durch die identifizierten Gene auch verschiedene Krankheiten, wie Krebs oder Alzheimer, beeinflusst.

Die *limma*-Analyse ist zur Identifikation synergistischer Interaktionen ungeeignet, da sie nicht genau genug ist. Durch Integration weiterer Datenquellen kann das Auffinden immunspezifischer Effekte verbessert werden. Beispielsweise kann durch Einbindung proteomischer Daten besser aufgeklärt werden, wo Signale miteinander interagieren, da Signalintegration, wie auch Signalprozessierung, auf proteomischer Ebene stattfinden. So ist die beobachtete Änderung der Genexpression auch als Resultat der ablaufenden bzw. beeinflussten proteomischen Prozesse zu werten.

## 4.2 Erstellung eines IL-3- und Flu-spezifischen Signaltransduktionsnetzwerks in pDC

Die Prozessierung von extrazellulären Informationen stellt für Zellen eine Herausforderung dar. Signale müssen aufgenommen und entsprechend ihrer Bedeutung verarbeitet werden. Dabei interagieren verschiedene biologische Entitäten<sup>34</sup> miteinander und bilden ein komplexes Geflecht von Beziehungen, welches als Netzwerk abgebildet werden kann (Mason und Verwoerd, 2007; Erwin und Davidson, 2009; Bortfeldt, Schuster und Koch, 2010; Janowski *et al.*, 2010). Dadurch werden die Beziehungen der Entitäten zueinander strukturiert dargestellt (Mason und Verwoerd, 2007; Bortfeldt, Schuster und Koch, 2010).

Im rekonstruierten Signaltransduktionsnetzwerk werden die Interaktionen zwischen verschiedenen Signalkomponenten abgebildet. Ziel ist es, ein genaueres Verständnis über ablaufende Prozesse zu erhalten, um später synergistische Interaktionen zwischen verschiedenen Signalen zu detektieren und mögliche Effekte vorherzusagen (siehe Abschnitt 4.3).

Für die Modellierung des Netzwerkgraphen wurde ein systembiologischer Ansatz, bei dem proteomische und genomische Ebene miteinander verbunden werden, gewählt. Ein Signal ist somit vom Rezeptor bis zu den regulierten Genen verfolgbar. Die Signalinteraktion und die Signalintegration verschiedener Stimuli sowie ihre Wirkungen werden dadurch besser verstanden. Besonders die synergistische Interaktion der Signale IL-3 und Flu auf proteomischer Ebene und die Wirkung auf die genomische Ebene in pDC sind hier von Interesse. Die zur Modellierung notwendigen proteomischen Informationen bezüglich der bekannten Signalübertragungswege wurden aus TRANSPATH<sup>®</sup> sowie zusätzlicher aktueller Literatur entnommen (Krull *et al.*, 2003; Krull *et al.*, 2006). Zusammen mit Informationen über Transkriptionsfaktoren und die durch sie beeinflussten Gene (extrahiert aus TRANSFAC<sup>®</sup>; Matys *et al.*, 2003; Matys *et al.*, 2006) wurden sie zu einem integrierten Netzwerk verschmolzen. Dadurch werden beide Ebenen miteinander verbunden, können aber gleichzeitig auch separat analysiert werden.

---

<sup>34</sup>Entitäten sind hier sämtliche Objekte oder Komponenten der biologischen Welt, etwa einzelne Proteine oder auch Gene.

Die in TRANSPATH<sup>®</sup> (siehe 3.1.2) und TRANSFAC<sup>®</sup> (siehe 3.1.1) repräsentierten Daten sind von hoher Qualität, da sie experimentell verifiziert und manuell aus Fachpublikationen extrahiert wurden (Krull *et al.*, 2003; Choi *et al.*, 2004a; Krull *et al.*, 2006; Matys *et al.*, 2006). Beide Datenquellen sind vom Umfang her größer und aktueller als vergleichbare öffentliche Ressourcen, wie beispielsweise Reactome (siehe Aktualisierungen für Immune System [http://www.reactome.org/cgi-bin/toc?DB=gk\\_current](http://www.reactome.org/cgi-bin/toc?DB=gk_current), 2006). Für die Rekonstruktion eines möglichst vollständigen Netzwerkgraphen stellen sie eine gute Datenbasis zur Verfügung. Damit wurden sämtliche Prozesse beginnend mit der Perzeption eines Signals durch einen adäquaten Rezeptor und endend bei den durch das Signal beeinflussten Genen abgebildet.

### 4.2.1 Netzwerkmodellierung mittels bipartiter Graphen

Das extrahierte Netzwerk wird durch einen Graphen repräsentiert. Die relevanten biologischen Entitäten (Objekte) wurden identifiziert und durch die Knoten des Graphen dargestellt. Dabei wurden sowohl Entitäten der proteomischen als auch der genomischen Ebene abgebildet. Auf Ebene der Proteine basiert Signaltransduktion auf Protein-Protein-Interaktionen sowie Modifikation von Proteinen und Komplexbildungen<sup>35</sup>. Die Verbindung zwischen proteomischer und genomischer Ebene bilden Transkriptionsfaktoren, die verschiedene Gene beeinflussen.

Bei Signaltransduktionsprozessen auf Protein-Ebene spielen Komplexe und ihre Modifikationen eine wichtige Rolle. Verschiedene Modifikationen oder auch Konformationsänderungen innerhalb eines Komplexes können unterschiedliche Signalwege aktivieren, so beeinflussen sie die Signalweiterleitung ebenso wie die weitere Prozessierung (Hlavacek *et al.*, 2003). In gängigen Projekten<sup>36</sup>, z.B. bei InnateDB (Lynn *et al.*, 2008; Lynn *et al.*, 2010) oder NetPath (Kandasamy *et al.*, 2010), wird auf eine vollständige Abbildung von Modifikationen und Komplexbildungen verzichtet, weshalb Signaltransduktion nicht in allen Aspekten nachvoll-

---

<sup>35</sup>Komplexe sind Agglomerate zwischen Proteinen und ggf. anderen Molekülen zu einer gemeinsamen Struktur, die aus verschiedenen oder gleichen Untereinheiten besteht. Ein Beispiel sind Liganden-Rezeptor-Komplexe, die Signale aus der Umwelt an die Zelle weitergeben (Janeway *et al.*, 2001).

<sup>36</sup>Eine Übersicht vieler verschiedener Projekte unter anderem zur Abbildung von Signaltransduktionsprozessen findet man unter: <http://pathguide.org> (Bader, Cary und Sander, 2006).

zogen werden kann. Die korrekte Abbildung von Komplexen und ihren Untereinheiten ist schwierig und das verfügbare Wissen lückenhaft. Ein Problem ist die experimentelle Aufklärung der Struktur eines Komplexes, insbesondere die seiner Untereinheiten und Modifikationen. So wird beispielsweise bis heute darüber diskutiert, wie genau der IL-3-Rezeptor aufgebaut ist und wie die Signalkaskade aktiviert wird (Guthridge *et al.*, 1998; Martinez-Moczygamba und Huston, 2003; Huang *et al.*, 2005b; Dey *et al.*, 2009; Lopez *et al.*, 2010). In der vorliegenden Arbeit werden alle zur Verfügung stehenden Informationen über Komplexe und ihre Modifikationen detailliert repräsentiert.

Relationen zwischen verschiedenen Knoten werden durch Kanten realisiert. Beispielsweise kann so abgebildet werden, wie durch eine chemische Reaktion ein Stoff modifiziert wird oder wie Signale zwischen zwei Signalträgern ausgetauscht werden. Besteht ein Netzwerkgraph nur aus einem Knotentyp, so heißt er unipartit. Diese Graphen stellen heutzutage das gängigste Modellierungskonzept für biologische Netzwerke dar (siehe Barabási und Oltvai, 2004; Zhao *et al.*, 2006; Mason und Verwoerd, 2007; Opsahl, Agneessens und Skvoretz, 2010). Analysen, welche die Wichtigkeit von Knoten im Netzwerk bestimmen, wie die zum Grad eines Knotens<sup>37</sup> (z.B. Diestel, 2000; Zhao *et al.*, 2006; Mason und Verwoerd, 2007) oder der Pairwise Disconnectivity Index<sup>38</sup> (Potapov, Goemann und Wingender, 2008), basieren auf unipartiten Graphen und liefern dort eine biologische Interpretationsmöglichkeit. Biologische Prozesse sind jedoch inhärent bipartit (Heiner, Gilbert und Donaldson, 2008). Dies bedeutet, dass außer den biologischen Entitäten (z.B. Molekülknoten) die zwischen ihnen ablaufenden Prozesse als eigenständige Objekte anzusehen sind. An diesen Prozessen nehmen oft mehrere Entitäten teil, gegebenenfalls werden aber nicht alle verändert, wie beispielsweise Katalysatoren. Prozesse benötigen mehrere Vorbedingungen zum Start, welche alle erfüllt sein müssen. Dies ist auch einer der wesentlichen Unterschiede zu unipartiten Netzwerken, denn durch die Verbindung von genau zwei Knoten können keine weiteren Vorbedingungen dargestellt werden. Im folgenden Beispiel werden beide Modellierungsformen miteinander verglichen.

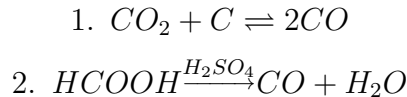
---

<sup>37</sup>Der Grad eines Knotens ist die Menge der mit ihm verbundenen Kanten (Diestel, 2000).

<sup>38</sup>Der Pairwise Disconnectivity Index ist eine Metrik zur Bestimmung der Wichtigkeit eines Knotens im Graphen, dem eine hohe Wichtigkeit zuordnet wird, wenn nach Entfernung des Knotens viele Paare an Knoten nicht mehr gegenseitig erreichbar sind.

## Unipartite und bipartite Modellierung

Anhand zweier chemischer Reaktionen zur Darstellung von Kohlenstoffmonooxid (CO) werden unterschiedliche Modellierungskonzepte illustriert. In einer Zelle spielen diese Reaktionen keine Rolle. Signaltransduktionsprozesse folgen den gleichen Grundregeln wie chemische Reaktionen, daher kann aus dem Beispiel ein Rückschluss auf die zu präferierende Modellierung gezogen werden. Zuerst wird eine unipartite und anschließend eine bipartite Modellierung diskutiert.



Chemische Substanzen (Moleküle<sup>39</sup>) bilden Knoten des Modells. Die exakten stöchiometrischen Verhältnisse<sup>40</sup> werden im Modell vernachlässigt. Bei der Modellierung einer Zelle sind die genauen Konzentrationen für bestimmte Moleküle oft unbekannt. Deshalb wird immer von einer ausreichenden Konzentration ausgegangen.

Im klassischen unipartiten Modell werden zwei Knoten miteinander verbunden. Dies führt zu Unklarheiten bei der Repräsentation von Reaktionen. Im Modell der beiden Reaktionen (Abb. 4.4) kann nicht unterschieden werden, welche Moleküle miteinander interagieren um CO herzustellen. Es ist unklar, dass  $CO_2$  und C zur Bildung von CO notwendig sind, ebenso wie die alternative Herstellung durch HCOOH nicht erkennbar ist. Offen bleibt außerdem wie  $H_2SO_4$  als Katalysator modelliert werden soll. Als reine Kantenmarkierung steht es nicht für weitere Reaktionen zur Verfügung, falls das Modell erweitert werden soll. Bei einer Modellierung von  $H_2SO_4$  als Knoten ist unbekannt, wie es mit den anderen Knoten verbunden werden muss, um als Katalysator identifizierbar zu sein.

Die Interpretation der unipartiten Darstellung erfordert implizit chemisches Expertenwissen, um die Zusammenhänge zwischen einzelnen Knoten zu verstehen und auflösen zu können. Bei der Modellierung eines großen Netzwerks,

---

<sup>39</sup>Zur Vereinfachung und für einen einheitlichen verständlichen Sprachgebrauch werden hier sämtliche Entitäten als Moleküle bezeichnet, auch wenn es sich um Ionen oder Atome wie im Fall von Kohlenstoff (C) handelt.

<sup>40</sup>„Die Stöchiometrie ist die Lehre der Mengenverhältnisse von Elementen in Verbindungen und der mengenmäßigen Beziehungen zwischen Elementen und Verbindungen bei chemischen Reaktionen“ (entnommen aus: <http://netchemie.de/lexikon/29.html>). Die Stöchiometrie stellt dabei ein mathematisches Hilfskonstrukt dar, welches hier ignoriert wird, da zu keinem Zeitpunkt Berechnungen an einzelnen Komponenten des Systems bezüglich ihrer Konzentrationen vorgenommen werden.

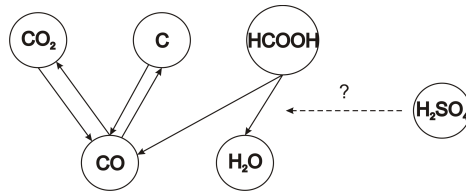


Abbildung 4.4: Klassische unipartite Darstellung der beiden Beispielreaktionen zur Synthese von CO. Die Knoten werden mit ihren Bezeichnern dargestellt. Die Richtung einer Kante ist angegeben durch ihren Pfeil. Z.B. bei  $\text{CO}_2$  und  $\text{CO}$  stellen die beiden Pfeile die Hin- und Rückreaktionen dar.  $\text{H}_2\text{SO}_4$  ist ein Katalysator, der die Reaktion beeinflusst, aber nicht *aktiv* an ihr teilnimmt. Wie er in den Graph eingebunden werden soll, ist ungewiss und wird durch ein „?“ and der Kante symbolisiert. Eine Möglichkeit ist eine Kantenmarkierung an den beiden von  $\text{HCOOH}$  ausgehenden Kanten. Wenn  $\text{H}_2\text{SO}_4$  an weiteren Reaktionen teilnimmt, jedoch nicht als Katalysator, so muss es zusätzlich als Knoten repräsentiert werden, wodurch Doppelrepräsentationen entstehen.

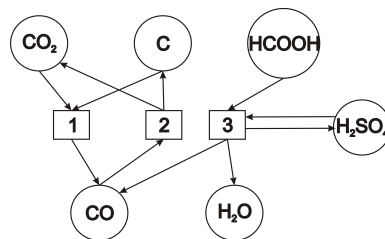


Abbildung 4.5: Bipartite Darstellung der Beispielreaktionen. Der zweite Knotentyp (Prozesse oder Reaktionen) wird durch nummerierte Rechtecke symbolisiert. Die Knoten 1 und 2 sind Bestandteil von Reaktion 1. Dabei werden Hin(1)- und Rück(2)-Reaktion getrennt dargestellt. Ein Katalysator wie  $\text{H}_2\text{SO}_4$  kann direkt an eine Reaktion gebunden werden. Die jeweiligen Interaktionspartner sind in jeder Reaktion unterscheidbar. Voraussetzungen oder Edukte werden an einer Reaktion durch eingehende Kanten dargestellt, z.B.  $\text{CO}_2$  und  $\text{C}$  für Reaktionsknoten 1. Die Produkte werden durch ausgehende Kanten mit einem Reaktionsknoten verbunden wie  $\text{CO}$  bei der Anbindung an Reaktionsknoten 1.

wie dem einer Zelle, kann aufgrund der Menge an Interaktionen dieses Wissen nicht durch einen Menschen bereitgestellt werden (Palsson, 2006). Informationen, die in den Reaktionsgleichungen noch enthalten sind, gehen verloren. Die Einführung von Prozessknoten und die daraus resultierende bipartite Darstellung (siehe Abb. 4.5), welche sich aus den in der Informatik bekannten Petri-Netzen motiviert (siehe Petri, 1962; Heiner, Gilbert und Donaldson, 2008), verhindert den Informationsverlust. Der neue Knotentyp repräsentiert die Reaktionen als abstrakte Prozesse. Die Kanten sind die Verbinder zu diesen Prozessen. In Abb. 4.5 sind beide Prozesse sowie die Hin- und Rückreaktion unterscheidbar. Dabei werden Edukte<sup>41</sup> und Produkte<sup>42</sup> unterschieden. Katalysatoren oder Enzyme, die Bestandteil einer Reaktion sind, werden dargestellt und stehen gleichzeitig für weitere Prozesse zur Verfügung. Im Vergleich zur unipartiten Modellierung (siehe Abb. 4.4) wird für die Interpretation kein implizites chemisches Expertenwissen benötigt. Eingehende Kanten an einem Molekülknoten markieren diesen als Produkt eines Prozesses, ausgehende Kanten beschreiben ihn als Edukt für einen Prozess. Für jeden Prozess gilt, dass die Anwesenheit aller Edukte notwendig ist, um ihn zu aktivieren. Nur bei einem aktiven Prozess werden alle mit ihm verbundenen Produkte erreicht. Bei einer bipartiten Modellierung werden die biologischen Gegebenheiten korrekt abgebildet. Für die Repräsentation des rekonstruierten Netzwerks wurde daher die bipartite Modellierungsform genutzt.

### 4.2.2 Inkorporation speziesübergreifender Informationen

Mithilfe des Netzwerks wurde die Signalintegration verschiedener Stimuli in menschlichen Zellen untersucht. Obwohl das Wissen um Signalweiterleitungsprozesse und ihre Verarbeitung im Menschen immer weiter wächst, sind einige Teile schlechter untersucht als andere, wodurch noch viele Wissenslücken existieren. Deshalb wurden die Informationen des Menschen (*Homo sapiens*) mit denen aus den Modellorganismen Maus (*Mus musculus*) und Ratte (*Rattus spp.*) anhand einer Orthologie im Netzwerk (siehe Abschnitt 3.5) zusammengefasst. Dadurch wird unbekanntes Wissen einer Spezies durch verfügbares Wissen einer anderen komplettiert.

---

<sup>41</sup>Edukte: Die Gesamtheit an Reaktionspartnern, die für den Ablauf einer Reaktion notwendig sind.

<sup>42</sup>Produkte: Gesamtheit der Ergebnisse einer Reaktion.



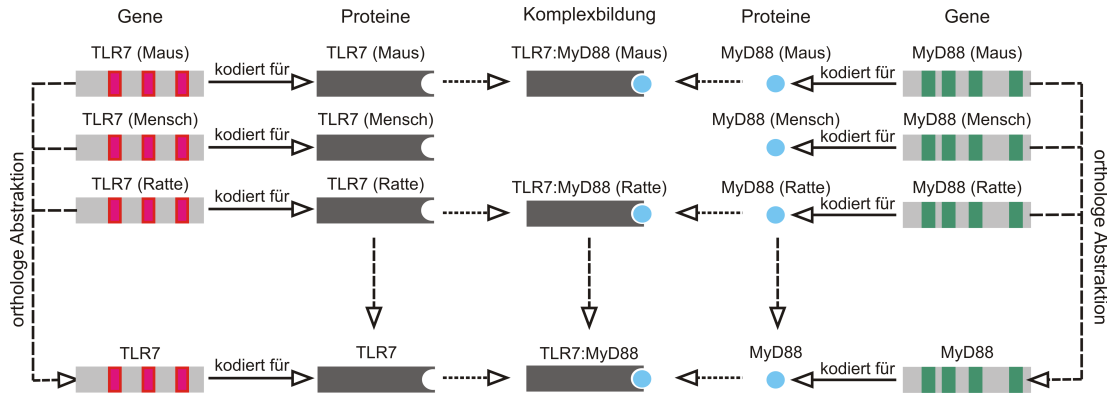


Abbildung 4.6: Schematische Darstellung der orthologen Integration von Informationen aus verschiedenen Spezies. Links und rechts außen werden Gene für TLR7 und MyD88 in den Spezies Maus, Mensch und Ratte dargestellt. Diese Gene kodieren für ihr spezifisches Protein. Für MyD88 und TLR7 ist in Maus und Ratte bekannt, dass sie zusammen einen Komplex bilden (gepunktete Linie). Für den Menschen ist dieses Wissen im Beispiel nicht präsent. Durch orthologe Abstraktion (unten) werden die Informationen verschmolzen, sodass die Gültigkeit für alle drei Spezies angenommen wird. Somit wird die Komplexbildung von MyD88 und TLR7 auch im Menschen möglich und Wissenslücken werden geschlossen.

Abbildung 4.6 zeigt exemplarisch, wie Informationen aus Mensch, Maus und Ratte in einem gemeinsamen Knoten gebündelt werden. Im Beispiel werden die kodierenden Gene für TLR7 und MyD88 betrachtet, welche in allen Spezies vorkommen. Das Gen wird jeweils in sein spezifisches Protein übersetzt. Über die Interaktionen von MyD88 mit TLR7 wird in Maus und Ratte ein Komplex gebildet. Beim Menschen fehlt dieses Wissen. Durch die Integration der drei Spezies in einem gemeinsamen Knoten wurde die Komplexbildung auch für den Menschen angenommen. Mittels orthologer Abbildungen wird nicht nur das Wissen über mögliche Reaktionen komplettiert (siehe Abbildung 4.6), es wird auch das Wissen über die Existenz von Genen und Proteinen vervollständigt. Die für die Netzwerkrekonstruktion zusätzlich notwendigen Informationen über Maus und Ratte wurden ebenfalls aus TRANSPATH<sup>®</sup> und TRANSFAC<sup>®</sup> extrahiert (Matys *et al.*, 2003; Choi *et al.*, 2004a; Choi *et al.*, 2004b; Krull *et al.*, 2006).

### 4.2.3 Einbindung unterschiedlich granularen Wissens

Unterschiedliche experimentelle Methoden bedingen eine unterschiedliche Qualität und Granularität ihrer Ergebnisse (siehe Abschnitt 3.5). Mittels hierarchischer Abbildung wird im rekonstruierten Netzwerk dieser Fakt berücksichtigt und das Wissen dennoch miteinander verknüpft. Im Folgenden wird am Beispiel von Sos die Umsetzung einer hierarchischen Abbildung illustriert.

Sos ist eine Zusammenfassung von zwei Proteinen, welche in der Signaltransduktion Zellwachstum und Zelldifferenzierung beeinflussen und kontrollieren. Es existiert in zwei verschiedenen Formen, Sos1 und Sos2, die beide an diesen Aufgaben beteiligt sind. Sos1 ist dabei an lang- und kurzfristiger Kontrolle beteiligt, wogegen Sos2 vornehmlich kurzfristige Effekte vermittelt (Qian *et al.*, 2000). Die Speziesebene in Abb. 4.7 beinhaltet die größte Informationsdichte sowie die feinste Granularität, da beide Formen einzeln dargestellt werden. Mit jeder Zusammenfassung nimmt die Grobkörnigkeit zu und spezifische Informationen gehen verloren. Jedoch sind die spezifischen Informationen nicht immer relevant. Zum Verständnis der Regulation von Zellwachstum mag es ausreichend sein zu wissen, dass ein Protein der Sos-Gruppe beteiligt ist. Aus unterschiedlichem Fokus bei Experimenten folgt eine unterschiedliche Granularität an Wissen.

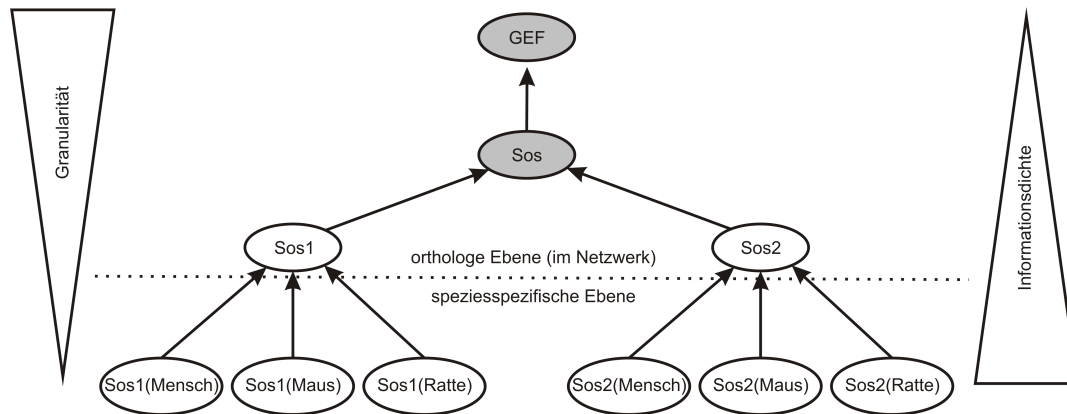


Abbildung 4.7: Schematische Abbildung der Hierarchien im Netzwerk am Beispiel von Sos1 und Sos2. Die speziesspezifische Ebene verdeutlicht die orthologe Abbildung (siehe Unterabschnitt 4.2.2). Durch die ungefüllten Pfeile wird diese dargestellt und zeigt, wie die Modellorganismen zusammengefasst werden. Links wird die Granularität des Wissens wiedergegeben, rechts die Informationsdichte. Je weiter man sich dem Boden nähert, desto größer ist die Informationsdichte und umso feinkörniger wird sie aufgelöst. Die in grau gehaltenen Knoten (Mitte) zeigen die implizite Hierarchisierung durch Gruppenbildung. Sos1 und Sos2 werden zu Sos verallgemeinert, was wiederum zu GEF verallgemeinert wird. GEF integriert noch verschiedene andere Proteingruppen, wie beispielsweise Vav (siehe Abb. 3.4 und Abschnitt 3.5).

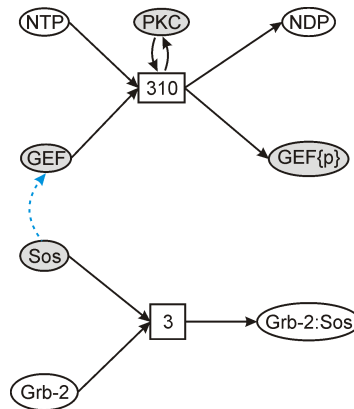


Abbildung 4.8: Wissen unterschiedlicher Granularität wird durch hierarchische Beziehungen (blau gestrichelt) repräsentiert. Aus Abb. 4.7 kann entnommen werden, dass GEF eine Übergruppe von Sos ist. Somit sind Reaktionen, die für GEF möglich sind, auch für Sos möglich. Die Verfolgung der hierarchischen Kanten führt dazu, dass dieses Schema eingehalten werden kann. Bei grau unterlegten Knoten handelt es sich um eine Gruppierung (Sos ist eine Gruppierung von Sos1 und Sos2, siehe Abb. 4.7). Die Informationen sind aus TRANSPATH<sup>®</sup> entnommen und zeigen, dass an einer Reaktion auch Moleküle unterschiedlicher Auflösung teilnehmen können.

Daraus ergeben sich Folgen für die repräsentierten Prozesse. Denn auch diese unterliegen, je nach Fokus der Untersuchung, einer unterschiedlichen Auflösung und klären spezifische Sachverhalte auf. Abbildung 4.8 verdeutlicht dies. So wird einerseits eine Komplexbildung von Sos mit Grb-2 beschrieben, welche beispielsweise bei der Signalweiterleitung für die Kontrolle des Zellwachstums relevant ist (Qian *et al.*, 2000). Andererseits wird ein Grundprinzip beschrieben, wobei GEF durch PKC phosphoryliert werden kann. Aus Abb. 4.7 ist zu entnehmen, dass Sos in einer hierarchischen Beziehung zu GEF steht. Es muss demnach im Sinne einer auf verallgemeinernde Eigenschaften fokussierten Modellierung gefolgert werden, dass Sos, wie auch Sos1 und Sos2 durch PKC phosphoryliert werden können. Unvollständiges Wissen, Fehlannotationen in den Datenbanken oder die experimentelle Unschärfe bei der Aufklärung spezifischer Prozesse erzeugen hier Unsicherheiten, welche nur durch weitere Experimente aufgelöst werden können und ggf. zur Überarbeitung der Gruppierungen führen. Auf den grundsätzlichen Modellierungsansatz hat diese Problematik jedoch keinen Einfluss.

Durch die Einführung von hierarchischen Kanten (blaue Kante in Abb. 4.8) in den Netzwerkgraphen wurden Ebenen verschiedener Granularität und Informationsdichte miteinander verbunden. Diese Kanten repräsentieren einen unterschiedlichen Fokus in den Experimenten, mit denen die Entitäten oder Prozesse untersucht werden. Deshalb wurden diese Ebenen miteinander verknüpft, wie beispielsweise die Verallgemeinerung von Sos zu GEF (Abb. 4.7 und 4.8) zeigt. Die Navigation entlang dieser Kanten ermöglicht es, ein globales Schema der Signaltransduktion abzubilden, trotz unterschiedlicher Qualität der zugrunde liegenden Daten. Dadurch werden zusätzlich Gruppierungen und Gemeinsamkeiten zwischen verschiedenen Knoten im Netzwerk erkennbar.

### 4.2.4 Zellspezifität für pDC durch Entfernung nicht exprimierter Gene und Proteine

Der Netzwerkgraph beinhaltet sämtliche Informationen über Signaltransduktion und die beeinflussten Gene, welche in TRANSPATH<sup>®</sup> und TRANSFAC<sup>®</sup> verfügbar sind oder zusätzlich aus der Literatur inkorporiert wurden. In einem komplexen Organismus sind jedoch nicht alle Prozesse zu jeder Zeit möglich. Unterschiedliche Funktionen werden durch spezialisierte Zellen, Gewebe oder Organe realisiert. Diese unterscheiden sich unter anderem anhand der in ihnen exprimierten Gene. Nach der Übersetzung der transkribierten Gene in Proteine verfügt jeder Zelltyp über eine charakteristische Ausstattung an Proteinen. Dies bedeutet, dass einige Signaltransduktionsprozesse nicht mehr stattfinden, da notwendige Voraussetzungen in der Zelle nicht mehr erfüllt sind. Für die spezifische Untersuchung von pDC mussten diese Signalpfade aus dem Netzwerkgraphen entfernt werden.

Experimentelle Datensätze geben Aufschluss über die Verfügbarkeit von Proteinen und die Expression von Genen in einem Gewebe und somit auch über mögliche und nicht mögliche Signalpfade. Beispielsweise enthält die Datenbank *Gene Expression Omnibus* (GEO) Genexpressionsdaten aus verschiedenen Experimenten, die angeben, in welchen Zelltypen oder Geweben welche Gene exprimiert werden (Edgar, Domrachev und Lash, 2002; Barrett *et al.*, 2005; Maglott *et al.*, 2007). Für die Untersuchung von pDC bietet GEO jedoch keinen adäquaten Datensatz. Daher wurde der hier untersuchte Datensatz selbst zur

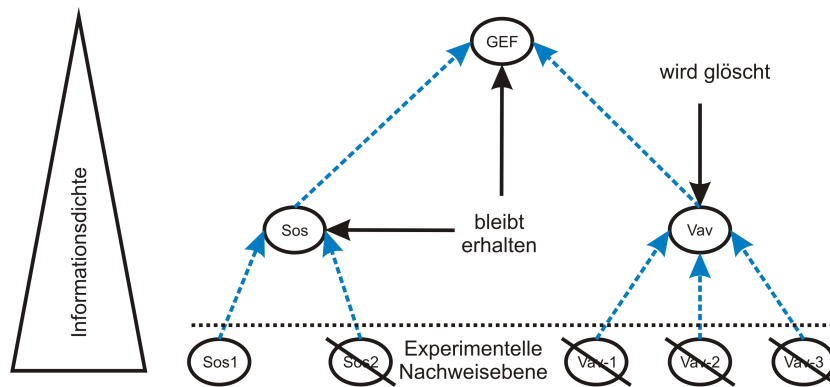


Abbildung 4.9: Zelltypspezifität durch Filterung und Entfernung von nachweislich nicht-exprimierten Knoten sowie ihren hierarchischen Verbindungen (blaue Kanten). Durchgestrichene Moleküle werden entfernt. Die Präsenz von Sos1 impliziert die Existenz von Sos sowie von GEF. Für Vav sind alle untergeordneten Proteine (Vav-1, Vav-2, Vav-2) nicht verifiziert, daher wird auch Vav entfernt.

Detektion differentiell exprimierter Gene verwandt. Die als nicht anwesend, nicht exprimiert identifizierten Gene (siehe Abschnitt 3.4) und die durch sie kodierten Proteine werden aus dem Netzwerk entfernt.

Problematischer ist der Nachweis der (Nicht-)Existenz von gruppierenden Knoten, wie beispielsweise von GEF (siehe Abb. 4.7), welche aufgrund der Nutzung hierarchischer und orthologer Informationen entstehen (siehe dazu 3.5 bzw. 4.2.2 und 4.2.3). Abbildung 4.9 verdeutlicht dies. Dazu werden die in Unterabschnitt 4.2.3 beschriebenen hierarchischen Kanten (blau, gestrichelt) verfolgt. Die durchgestrichenen Knoten werden im Beispiel entfernt. Aus dem Nachweis der Existenz eines Knotens auf einer Ebene mit höherer Informationsdichte wird auf die Existenz von gruppierenden Knoten auf Ebenen mit niedrigerer Dichte geschlossen. Dies führt dazu, dass gruppierende Knoten nur dann aus dem Netzwerk entfernt werden, wenn alle Knoten höherer Informationsdichte nicht exprimiert werden. Im Beispiel ist dies für Vav der Fall. Dort sind Vav-1, Vav-2 und Vav-3 als nicht-exprimiert anzusehen, weshalb auch der gruppierende Knoten Vav aus dem Graphen entfernt wird. So wird sichergestellt, dass nur nachweislich nicht vorhandene Entitäten entfernt werden. Das Löschen von Knoten führt zu Inkonsistenzen im Netzwerk und betrifft auch Prozesse. Durch die entfernten Entitäten, die in einigen Prozessen als Edukte dienen, sind sie nicht mehr aktivierbar. Diese Prozesse, sowie ihre Produkte, so sie nicht über alternative Wege angebunden

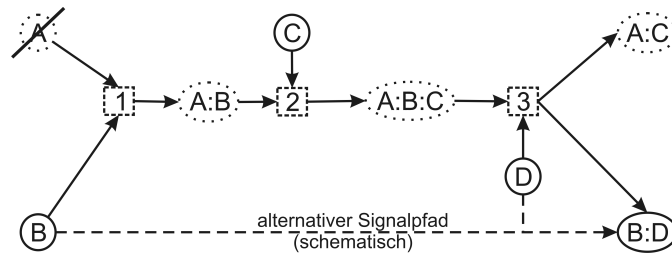


Abbildung 4.10: Anpassung der Interaktionen im Netzwerkgraphen nach Filterung von nicht-exprimierten Genen und Proteinen. Die Komponente A soll aus dem Netzwerk entfernt werden. Alle gepunkteten Entitäten und Prozesse sind davon betroffen. Die Reaktionen sind Komplexbildungsreaktionen. Reaktion 1 kann nicht stattfinden, da A nicht mehr im Graph enthalten ist. Es wird der Prozessknoten entfernt sowie alle Produktknoten. Abhängige Reaktionen und Moleküle werden solange iterativ entfernt, bis es zu einer Entität eine Alternative gibt, die nicht direkt von A abhängig ist. Der Netzwerkgraph kann nach Entfernen aller nicht mehr möglichen Prozesse isolierte Entitäten zurücklassen (hier: C).

sind, wurden ebenfalls aus dem Graphen entfernt. Dadurch wurden die nicht in pDC möglichen Signaltransduktionspfade entfernt.

In Abb. 4.10 werden die Folgen einer gelöschten Entität auf stattfindende Prozesse anhand einer Reihe konsekutiver Komplexreaktionen verdeutlicht. Zuerst wird ein gemeinsamer Komplex aus den Komponenten A und B gebildet (A:B). Da jedoch A markiert (durchgestrichen) ist und aus dem Netzwerkgraphen entfernt wird, kann Reaktion 1 nicht stattfinden. Diese sowie ihre Produkte werden daher aus dem Graphen entfernt. Mittelbar sind auch die Reaktionen 2 und 3 von A abhängig. Sie sowie ihre Produkte, die nicht über einen alternativen (von A unabhängigen) Pfad verfügen, werden ebenso entfernt. Die Präsenz des Komplexes B:D ist von A unabhängig, da ein alternativer Pfad existiert. So werden zwar Reaktion 3 sowie die ein- und ausgehenden Kanten entfernt, aber B:D bleibt erhalten. Diese Remodellierung kann dazu führen, dass einzelne Teile im Netzwerkgraphen isoliert sind. Ein Zerfall des Netzes in verschiedene Module ist möglich, wobei nicht verbundene Elemente disjunkte Signaltransduktionspfade darstellen. Nach erfolgter Remodellierung war der Netzwerkgraph konsistent bezüglich der repräsentierten Prozesse und stellt die für pDC spezifische Signaltransduktion dar.

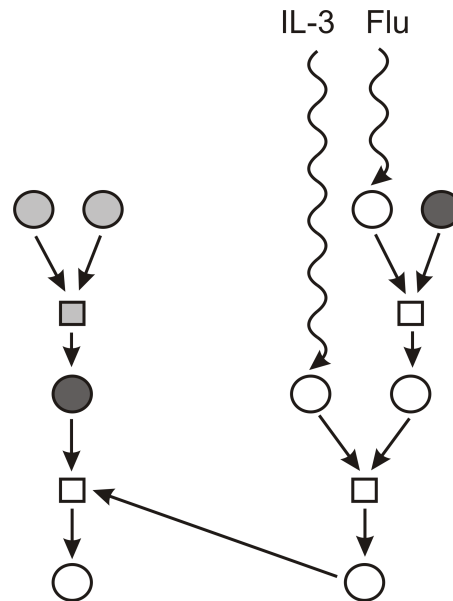


Abbildung 4.11: Schematische Ableitung des signalabhängigen Netzwerkgraphen. In weiß gehaltene Knoten (Kreise) und Prozesse (Quadrate) werden von IL-3 oder Flu beeinflusst. Sie sind für das Verständnis der Signalweiterleitung essentiell. Prozesse in hellgrau sowie ihre Edukte werden von keinem der Signale beeinflusst und deshalb aus dem Netzwerkgraphen entfernt. Sie liefern keine Informationen über die stimuliabhängige Signaltransduktion. Die in dunkelgrau gehaltenen Knoten sind notwendig zur Aktivierung von IL-3- oder Flu-abhängigen Prozessen. Diese Knoten bleiben erhalten.

#### 4.2.5 IL-3- und Flu-spezifische Signaltransduktion in pDC

Für die signalspezifische Untersuchung der durch IL-3- und Flu-induzierten Weiterleitungsprozesse in pDC müssen sämtliche Prozesse und Moleküle von mindestens einem Stimulus erreichbar sein. Nur auf diesen Pfaden werden die Signale verarbeitet. Deshalb lassen sich Synergien nur dort finden, wo von IL-3 oder Flu abhängige Signalpfade miteinander interagieren. Mittels einer Downstreamsuche beginnend mit den für IL-3- und Flu-spezifischen Rezeptoren (IL3R, TLR7) wurden diese Signalpfade identifiziert.

Nur diese sowie ihre Edukte und Produkte wurden abgebildet. Alle weiteren Prozesse sind isoliert oder bieten keinen Informationsgewinn über die Signaltransduktion von IL-3 oder Flu. Deshalb wurden sie aus dem Netzwerkgraphen



entfernt, ebenso ihre Edukte und Produkte, so sie nicht mit weiteren, von IL-3- und Flu-abhängigen Prozessen verbunden waren. Abbildung 4.11 verdeutlicht die Situation. Die in hell- und dunkelgrau markierten Knoten werden nicht durch die beiden Signale beeinflusst. Jedoch sind die in dunkelgrau gehaltenen Knoten Edukte für einen Prozess, der von mindestens einem der Signale erreicht wird. Da für den Ablauf eines Prozesses oder einer Reaktion immer alle Voraussetzungen erfüllt sein müssen, also alle notwendigen Edukte verfügbar sein müssen, werden die dunkelgrauen Knoten nicht aus dem Netzwerkgraphen entfernt. Die in hellgrau gehaltenen Knoten werden jedoch entfernt. Der daraus resultierende Netzwerkgraph fokussiert sich ausschließlich auf die durch IL-3- und Flu-induzierte Signaltransduktion in pDC.

In Tab. 4.2 wird die ursprüngliche Datengrundlage mit dem resultierenden signal- und gewebespezifischen Netzwerk verglichen. Moleküle, die neben Prozessen (Reaktionen) den zweiten Knotentyp im bipartiten Netzwerk darstellen, werden hinsichtlich ihrer annotierten Funktion aufgeteilt. Es fällt auf, dass viele Komplexe enthalten bleiben, wobei im rekonstruierten Netzwerk die Bedeutung von Komplexen ( $\approx 67\%$  aller Moleküle) größer ist als in TRANSPATH<sup>®</sup> insgesamt ( $\approx 35\%$  aller repräsentierten Moleküle). Dies ist ein Hinweis für die generelle und bisher, insbesondere bei Modellierungsansätzen, oft unterschätzte Wichtigkeit von Komplexen in Signaltransduktionsprozessen. So spielen Komplexe bei enzymgekoppelten Rezeptoren eine wichtige Rolle. Der IL-3-Rezeptor etwa besitzt keine intrinsische Kinasedomäne, welche nach Bindung des IL-3-Liganden den Rezeptor direkt aktiviert und dadurch die weitere Signalkaskade anstößt (Martinez-Moczygemba und Huston, 2003; Dey *et al.*, 2009; Lopez *et al.*, 2010). Erst die Bindung weiterer Komponenten (JAK) zu einem Multikomponentenkomplex ermöglicht die Aktivierung des IL-3-Rezeptors (Guthridge *et al.*, 1998; Martinez-Moczygemba und Huston, 2003; Huang *et al.*, 2005b; Dey *et al.*, 2009; Lopez *et al.*, 2010). Für ein genaues Verständnis der Signaltransduktion sind mehr Informationen über Komplexe und Modifikationen notwendig, da beides Einfluss auf die Aktivität eines spezifischen Signalpfades hat (Hlavacek *et al.*, 2003).

Auffällig in Tab. 4.2 ist die geringe Menge an Inhibitoren. Dies zeigt, dass Signalinteraktionen generell noch nicht vollständig verstanden sind. Am Anfang der Aufklärung von spezifischen Signalwegen liegt das Hauptaugenmerk auf der Ver-

Parameter	Datenbasis (TP,TF)	IL-3-, Flu- und pDC-spezifisches-NW
<b>Moleküle</b>	98197	15843
Katalysatoren	1512	632
Inhibitoren	24	19
Komplexe	33938	10614
TFs	7570	1259
Edukte	17946	4475
Produkte	39364	11496
<b>Reaktionen</b>	42158	8901
<b>NW Kanten</b>	nicht bestimmt	44990
<b>Gene (Mensch, Maus, Ratte)</b>	17642	
<b>orthologe Gene</b>	7707	2382

Tabelle 4.2: Darstellung wichtiger Parameter des rekonstruierten IL-3-, Flu-signalspezifischen Netzwerks in pDC. Die zweite Spalte zeigt die Menge der aus TRANSPATH<sup>®</sup> (TP) und TRANSFAC<sup>®</sup> (TF) extrahierten Informationen für Mensch, Maus und Ratte. Diese Informationen bilden die Datenbasis für die Netzwerkrekonstruktion. Das resultierende Netzwerk (signal- und gewebespezifisch) ist in der letzten Spalte dargestellt. Es enthält  $\approx 16\%$  (15843) der Molekülknoten aus der Datengrundlage (98197). Jedoch sind  $\approx 67\%$  (10614) der enthaltenen Knoten Komplexe, im Gegensatz zur gesamten Basis, wo nur etwa  $\approx 35\%$  (33938) aller Molekülknoten Komplexe sind. Bei der Rekonstruktion bleiben  $\approx 17\%$  (1259) der Transkriptionsfaktoren erhalten sowie  $\approx 79\%$  (19) an Inhibitoren. Die insgesamt kleine Anzahl an Inhibitoren ist ein Indiz für Wissenslücken. Bei der Abbildung des signalspezifischen Netzwerks bleiben  $\approx 21\%$  (8901) der Reaktionen erhalten. Somit werden tendenziell mehr Reaktionen als Moleküle erhalten. Auf genomischer Ebene werden im rekonstruierten Netzwerk  $\approx 31\%$  (2382) der verfügbaren orthologen Gene (7707) abgebildet. Diese Gene werden in pDC exprimiert und werden potentiell durch IL-3 und Flu beeinflusst.

vollständigkeit des Wissens über diesen Signalpfad. Interaktionen oder mögliche inhibitorische Wirkungen entlang des Pfades werden erst später interessant und erst dann untersucht. Die relativ geringe Menge an Edukten im rekonstruierten Netzwerk (siehe Tab. 4.2) ist ein Beleg für diese Form der experimentellen Aufklärung. Besonders der Vergleich zu den Produkten (etwa  $2,5\times$  so viele Produkte wie Edukte) impliziert, dass nur wenige Moleküle Produkt und Edukt gleichzeitig sind. Für die Netzwerkstruktur ergeben sich somit überwiegend lineare Pfade, bei denen viele Produkte nicht weiter verknüpft sind und nur die wichtigsten Teile der Signalpfade aufgeklärt wurden.

Interessant ist außerdem die große Menge an TFs (1259), welche durch IL-3 und Flu erreicht werden. Dies zeigt, dass viele proteomische und genomische Funktionen durch die beiden Stimuli beeinflusst werden. Wie aus unipartiten Graphen bekannt ist, existieren auch in bipartiten Graphen sogenannte Hubknoten. Dies sind Knoten, die sehr gut vernetzt sind und viele verschiedene Funktionen regulieren. So wird beispielsweise NF- $\kappa$ B von vielen unterschiedlichen Signalwegen erreicht und bildet als Molekülknoten und TF eine zentrale Schlüsselstelle für die Verteilung der Signale (Pikarsky *et al.*, 2004; Tripathi und Aggarwal, 2006; Gupta *et al.*, 2011). Obwohl theoretisch beide Knotentypen Hubknoten ausbilden können, trifft dies bei der hier vorgestellten Modellierung nicht zu. Nur die Molekülknoten bilden Hubknoten aus, da sie die biologischen Entitäten darstellen. Die Prozessknoten repräsentieren die spezifischen biochemischen Reaktionen bei denen nur eine überschaubare Anzahl an Molekülen miteinander interagiert.

Im rekonstruierten Netzwerk werden etwa doppelt so viele Gene wie vorhandene TFs erreicht. Das heißt, dass ein TF mehrere Gene beeinflusst und ist konsistent mit heutigem Wissen, wonach die Kombination verschiedener TFs die Expression eines Gens steuert (Alberts *et al.*, 2008). Die recht große Menge an erreichbaren Genen (2382) im signal- und gewebespezifischen Netzwerk hebt die Bedeutung der beiden Stimuli hervor. Es unterstreicht die Beobachtungen aus der Literatur, wonach durch die beiden Stimuli in pDC eine Immunantwort induziert wird. Als Folge differenziert sich pDC beispielsweise in DC aus oder induziert über freigesetzte inflammatorische Signale die Rekrutierung zusätzlicher Immunzellen (Grouard *et al.*, 1997; Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007). Gerade die Umwandlung von pDC in DC erfordert Änderungen in vielen genetischen Programmen, wodurch sich die große Anzahl an erreichbaren Genen erklärt.

### 4.2.6 Zusammenfassung

Für das Verständnis von Signaltransduktion und Signalintegration im Menschen wurde ein systembiologischer Ansatz gewählt. Dadurch werden die verschiedenen an der Signalverarbeitung beteiligten Ebenen zusammengeführt. Mittels Protein-Protein-Interaktionen werden Signale perzipiert und an die entsprechenden Organellen<sup>43</sup> der Zelle, insbesondere den Zellkern, weitergeleitet. Durch Interaktionen von Transkriptionsfaktoren mit der DNA wird die Aktivität von Genen, die als Blaupausen für Proteine dienen, gesteuert. Die Übersetzung von mRNA-Transkripten in neue Proteine versetzt die Zelle in die Lage, auf eine sich ändernde Umwelt zu reagieren. Es existiert daher eine permanente Kopplung von proteomischer und genomischer Ebene.

Die Datenbanken TRANSPATH<sup>®</sup> und TRANSFAC<sup>®</sup> (siehe Unterabschnitt 3.1.2 und 3.1.1) liefern die Grundlage für die Untersuchung von IL-3- und Flu-spezifischer Signaltransduktion in pDC. Da der Signaltransduktion biochemische Prozesse zugrunde liegen und das rekonstruierte Netzwerk die vorhandenen Informationen ohne Informationsverlust abbilden sollte, wurde eine bipartite Modellierung verwandt. Diese Modellierungsform entspricht der biologischen Realität, da verschiedene Moleküle in Prozessen miteinander interagieren (siehe Unterabschnitt 4.2.1 und Heiner, Gilbert und Donaldson, 2008). Bei der Modellierung wurden Proteine (inklusive TFs), Komplexe, Modifikationen, Katalysatoren, Inhibitoren und Gene berücksichtigt und im Netzwerk repräsentiert. Wesentlich für die Signaltransduktion sind Komplexe und Modifikationen. Da eine genaue (experimentelle) Aufklärung von Komplexen und Modifikationen schwierig oder unnötig beziehungsweise ungenau ist (je nach Fokus der Untersuchung), existiert Wissen unterschiedlicher Granularität. Viele andere Projekte (für eine Übersicht siehe Bader, Cary und Sander, 2006) vernachlässigen Komplexe und Modifikationen oder bilden sie nicht konsequent ab. Beim hier rekonstruierten Netzwerk lag jedoch ein besonderes Augenmerk auf der verlustfreien Abbildung von Prozessen, Proteinkomplexen und Modifikationen, deshalb repräsentiert der abgeleitete Graph das verfügbare Wissen detailgenau.

Da das Wissen über die in menschlichen Zellen ablaufenden molekularen Prozesse lückenhaft ist, wurden Mensch, Maus und Ratte durch eine Orthologie zu-

---

<sup>43</sup>Organellen: Teile einer Zelle, die spezialisiert sind und funktionale Aufgaben übernehmen (zum Beispiel im Zellkern, wo die Transkription der Gene stattfindet) (Smith *et al.*, 2003; Alberts *et al.*, 2001).

sammengefasst (siehe Abschnitt 3.5 und Unterabschnitt 4.2.2). Mit Hilfe dieser Abstraktion wurde ein vollständiges Bild über die stattfindenden Signaltransduktionsprozesse in Säugetieren erzeugt. Auch hier ist die Granularität, also die Qualität von repräsentiertem Wissen, unterschiedlich. Um dies adäquat abzubilden, wurden hierarchische Kanten von Ebenen mit hoher Informationsdichte zu Ebenen mit niedriger Informationsdichte eingeführt (siehe 3.5 und 4.2.3). Ein Beispiel ist die Gruppierung verschiedener Proteine zu Proteinfamilien wie in Abb. 4.7 gezeigt wurde.

Nachdem die aus Datenbanken und zusätzlicher Literatur verfügbaren Informationen in einem Signaltransduktionsgraph abgebildet wurden, repräsentiert das Netzwerk den rezenten Wissensstand. Mittels Filterung wurde das Netzwerk weiter spezifiziert, sodass die für IL-3 und Flu möglichen Signalübertragungswege in pDC untersucht werden können. Dabei können sowohl proteomische Signaltransduktion als auch genomische Genexpression und Genregulation untersucht werden.

Mit mehr als 15.000 Molekülknoten (mehr als 10.000 Komplexe), mehr als 8.000 mechanistisch abgebildeten Reaktionen<sup>44</sup> und mehr als 2.000 Genen bietet der rekonstruierte IL-3- und Flu-spezifische Netzwerkgraph in pDC eine gesamtzelluläre Sicht auf die stattfindenden Signalprozesse (siehe Tab. 4.2). An diesem Graphen wird im Folgenden algorithmisch die Signalintegration der beiden Stimuli untersucht.

---

<sup>44</sup>Eine mechanistische Darstellung von Reaktionen bedeutet, dass, soweit möglich, sämtliche bekannte Einzelheiten dieser Reaktion berücksichtigt werden. Da biochemische Reaktionen inhärent bipartit sind (Heiner, Gilbert und Donaldson, 2008), bedeutet dies, dass sich Reaktionen in Prozesskomponenten und Molekülkomponenten zerlegen lassen. Ein bipartites Netzwerk wird dieser Abbildung gerecht und bildet die zur Verfügung stehenden Reaktionsinformationen ohne Verlust ab (siehe 4.2.1 und Abb. 4.5). Im Gegensatz dazu stehen semantische Abbildungen, in denen Schritte zusammengefasst werden oder einzelne Informationen nicht berücksichtigt werden (z.B. Modifikationen von Komplexen). Während mechanistische Repräsentationen viel Wert auf die genaue Abbildung legen, liegt bei einer semantischen Repräsentation der Fokus auf der Erlangung eines generellen Verständnisses.

### 4.3 Detektion und Propagierung synergistischer Effekte im bipartiten Signaltransduktionsnetzwerk

In Netzwerken werden wichtige Knoten durch topologische Parameter bestimmt. Diese bestimmen strukturelle Eigenschaften von Knoten (Barabási und Oltvai, 2004; Mason und Verwoerd, 2007; Opsahl, Agneessens und Skvoretz, 2010). Der einfachste Parameter ist der Grad eines Knotens (degree). Dabei werden alle mit einem Knoten verbundenen Kanten gezählt (Diestel, 2000). Knoten mit einem hohen degree bilden sogenannte Hub-Knoten, welche sehr gut mit ihrer Umgebung verknüpft sind. In der Biologie sind dies Moleküle, die stark von ihrer Umgebung beeinflusst werden, oder die selbst viele andere Moleküle beeinflussen, wie beispielsweise NF- $\kappa$ B, welches unter anderem an der Immunantwort beteiligt ist und viele zelluläre Reaktionen auf verschiedenste Stressoren reguliert (Brasier, 2006; Gilmore, 2006; Perkins, 2007). Bei Synergie ist jedoch die Verbindung eines Knotens mit den verschiedenen Stimuli wichtiger als seine Bedeutung, beispielsweise als Hubknoten. Der degree untersucht diese Wege nicht und ist deshalb zur Bestimmung von Synergien ungeeignet.

Topologische Parameter, welche Wege zwischen Knoten berücksichtigen, sind Betweenness- und Closeness-Centrality. Diese betrachten die Verknüpfung eines Knotens mit anderen Knoten im Netzwerk. So wertet die Closeness-Centrality eines Knotens  $v$  diesen als wichtig, wenn im Netzwerk alle anderen Knoten mit  $v$  über möglichst kurze Wege verbunden sind. Es werden Moleküle identifiziert, die in der Lage sind, Informationen möglichst schnell im gesamten Netzwerk zu verteilen (Newman, 2005). Die Betweenness-Centrality eines Knoten  $v$  ist hoch, wenn viele kürzeste Wege zwischen anderen Knotenpaaren im Netzwerk über  $v$  verlaufen (Freeman, 1977; Wuchty und Stadler, 2003; Newman, 2005; Mason und Verwoerd, 2007). Dies identifiziert Knoten, die die Informationsverteilung kontrollieren (Newman, 2005). Unter beiden Parametern verlaufen viele verschiedene Wege über die wichtigen Knoten, was sie jedoch als Kandidaten für Synergien *de facto* ausschließt. Die vielen über diese Knoten verlaufenden Wege verhindern, dass ein experimentell beobachteter potentiell synergistischer Effekt auf eine exklusive Interaktion der beiden Stimuli zurückgeführt werden kann.

Einen anderen Ansatz verfolgt der Pairwise Disconnectivity Index (PDI). Hier werden Knoten gesucht, die möglichst viele andere Knotenpaare in einem Netzwerk trennen, wenn sie aus dem Netzwerk entfernt sind (Potapov, Goemann und Wingender, 2008). Ein Knoten  $v$  ist daher wichtig, wenn Wege zwischen Paaren anderer Knoten über  $v$  verlaufen und diese sich ohne  $v$  nicht mehr erreichen. Somit ist  $v$  für viele Verbindungen alternativlos (Potapov, Goemann und Wingender, 2008). Biologisch betrachtet werden essentielle Moleküle identifiziert, die verschiedene Funktionen einer Zelle miteinander verbinden. Moleküle mit einem hohen PDI stellen Flaschenhälse dar, beispielsweise in der intrazellulären Kommunikation.

Für synergistisch induzierte Knoten gilt, dass sie von allen Stimuli aus *gemeinsam erreichbar* sind. Um einen synergistischen Effekt jedoch sicher zu detektieren, dürfen *keine alternativen Pfade* von einem beliebigen Stimulus zu  $v$  existieren. Bei Synergien, wie auch beim PDI, sind alternativlose Wege relevant. Das Entfernen eines essentiellen Knotens verhindert einerseits das Auftreten von Synergien und separiert andererseits durch die Alternativlosigkeit Teile des Netzwerks, was wiederum den PDI beeinflusst. Vergleicht man nun das Konzept des PDI mit Synergien insgesamt, so identifiziert der PDI Knoten, die ein Netzwerk global möglichst stark separieren. Die Anzahl nicht mehr verbundener anderer Knotenpaare ist der bestimmende Faktor (Potapov, Goemann und Wingender, 2008). Synergien hingegen können auch lokal auftreten und kleine Teile separieren. Es ist entscheidend, dass alle Stimuli diesen Knoten gemeinsam erreichen und dass bei Fehlen eines beliebigen Stimulus der Knoten nicht mehr erreichbar ist. Abb. 4.12 verdeutlicht diesen Sachverhalt exemplarisch. Dort haben viele nicht-synergistische Knoten einen höheren PDI<sup>45</sup> (rote Knoten) als die Synergistischen (blaue Knoten). Da Synergien lokal auftreten, kann der PDI nicht zur Detektion von Synergien verwandt werden, wie in Abb. 4.12 verdeutlicht wird.

Jedoch ist die Grundidee des PDI, Knoten zu suchen, zu denen es keine Alternativwege gibt, auch für die Detektion von Synergien relevant. Daher dient der PDI als wichtiger Ideengeber. Aufgrund der Wichtigkeit von lokal integrierenden Knoten wird ein neuer Algorithmus entwickelt, der Synergien sicher detektiert.

---

<sup>45</sup>Für die Berechnung des PDI wurde DiVa (verfügbar unter: <http://diva.sybig.de/>) verwandt (Goemann, Wingender und Potapov, 2009).

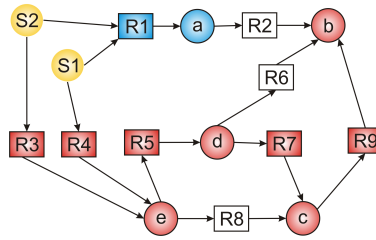


Abbildung 4.12: Vergleich von PDI und Synergie in bipartiten Graphen. Alle in rot gehaltenen Knoten haben einen höheren errechneten PDI als die blauen, synergistischen Knoten. Weiße Knoten haben einen kleineren PDI als beide. Die gelben Knoten stellen zwei Stimuli dar.

Wie *Heiner et al.* zeigen, ist biologische Signaltransduktion sowie Signalintegration inhärent bipartit (*Heiner, Gilbert und Donaldson, 2008*). Dies bedeutet, dass Signale in Prozessen miteinander interagieren und dadurch verändert werden. Man kann Signalträger (Moleküle) und Prozesse (Reaktionen) unterscheiden. An Prozessen können verschiedene Signale miteinander interagieren. Das Ergebnis der Interaktion ist an den direkt erreichbaren Molekülen<sup>46</sup> beobachtbar. Nur an den Molekülknoten können Synergien gemessen und experimentell bestätigt werden. Somit sind bipartite Graphen für die Detektion von Synergien erforderlich<sup>47</sup>. Um zu verstehen, wie Synergien in der Signaltransduktion wirken, ist es notwendig, die zellulären Gegebenheiten ganzheitlich zu betrachten. Dies bedeutet, dass Signalintegration auf proteomischer Ebene, ebenso wie die beeinflussten Gene, betrachtet werden müssen. TFs übertragen dabei die Signale von der proteomischen auf die genomische Ebene und regulieren die Genexpression. Dies hilft einer Zelle sich anzupassen. Um Synergien zu verstehen, müssen beide Ebenen gemeinsam betrachtet werden.

In Abb. 4.13 wird die Signalweiterleitung in einer Zelle ganzheitlich betrachtet. Gegeben sind zwei externe Stimuli (S1 und S2, gelbe Knoten) sowie das notwendige bipartite Netzwerk, welches aus Reaktionen (R, eckige Knoten) und Molekülen (M, runde Knoten) besteht. Durch die gepunktete Linie im unteren Teil der Abbildung wird der Transfer des Signals auf die genomische Ebene symboli-

<sup>46</sup>Dies sind all jene Moleküle  $m$ , die durch den Prozess  $r$  über eine Kante der Form  $(r, m)$  miteinander verbunden sind.

<sup>47</sup>Daher ist der Graph in Abb. 4.12 bipartit. Für die Bestimmung des PDI wird er jedoch als unipartit interpretiert, da der PDI nur auf unipartiten Graphen definiert ist (*Potapov, Goemann und Wingender, 2008*).



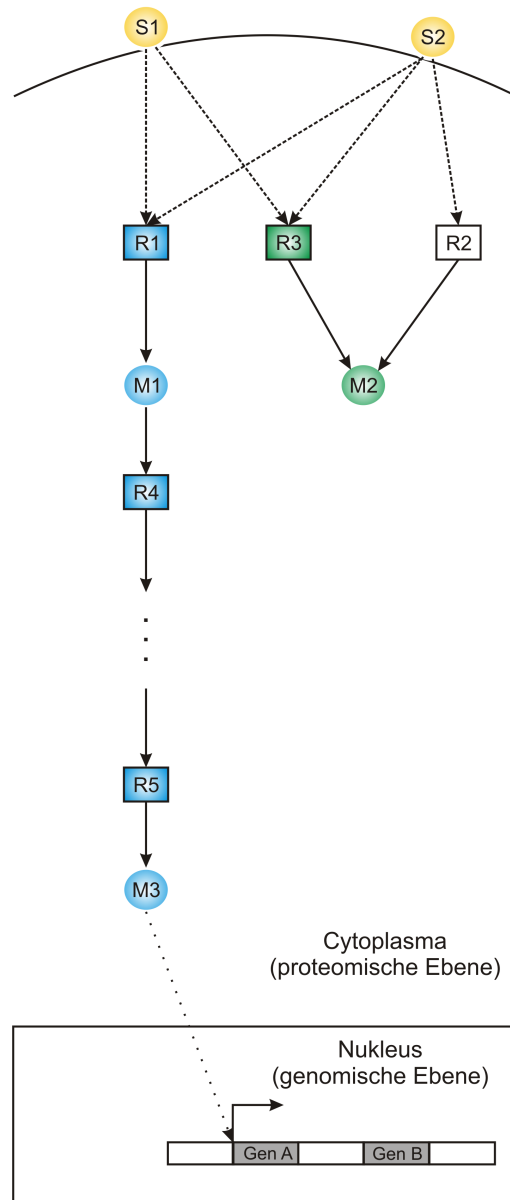


Abbildung 4.13: Schematische Darstellung von Signaltransduktion im Bezug auf Synergien. Die gelben Knoten stellen zwei gegebene Ausgangsstimuli dar. Synergistisch beeinflusste Moleküle (M) oder Reaktionen (R) werden blau hervorgehoben. Grün hervorgehobene Moleküle sind nicht exklusiv synergistisch. Zwischen R4 und R5 existiert ein Pfad beliebiger Länge, der von S1 und S2 abhängig ist, wodurch alle Komponenten entlang des Pfades ebenfalls synergistisch sind. M3 ist ein TF, welches das Signal auf die genomische Ebene überträgt (gepunktete Linie), wo es Gen A reguliert. Gestrichelte Linien zeigen eine Verbindung mit beliebig vielen Zwischenschritten auf.

siert. Zuerst werden die beiden Stimuli in die Zelle aufgenommen, beispielsweise durch Interaktion mit signalspezifischen Rezeptoren. Synergien treten nur auf, wenn die verschiedenen Stimuli interagieren. Dies geschieht in Reaktionen, die von allen Stimuli gemeinsam erreicht werden (blaue Reaktionsknoten und R3). Alle Produkte dieser Prozessknoten sind Kandidaten für synergistische Beeinflussung (blaue Molekülknoten und M2). Synergie liegt jedoch nur vor, wenn es keine alternative Möglichkeit gibt, ein Molekül ausgehend von einem beliebigen Stimulus zu erreichen. In Abb. 4.13 sieht man dies an Molekülknoten M2 (grüner Molekülknoten). Dieser wird über den parallelen Pfad ( $S2 \rightarrow R2 \rightarrow M2$ ) erreicht. Da die Synergie an M2 nicht exklusiv ist, kommt M2 nicht als Kandidat infrage (grün). Gleiches gilt für R3. Diese Reaktion integriert zwar die verschiedenen Signale, ist jedoch nicht in der Lage, sie detektierbar auf ein Molekül zu übertragen.

Während der Signalweiterleitung werden Signale oft in aufeinanderfolgenden Schritten moduliert (siehe Unterabschnitt 2.1.2 und Abb. 2.2). Auch synergistische Signale werden so moduliert. Dabei sind alle Schritte, die nicht durch einen Alternativweg mit einem Stimulus verbunden sind, ebenfalls synergistisch. In Abb. 4.13 sieht man dies an der Kette beginnend mit R1 bis einschließlich M3. Jede dieser Reaktionen ist von den beiden Stimuli aus erreichbar. Außerdem gibt es keinen Alternativweg zwischen einem beliebigen Stimulus und den Molekülknoten entlang des Weges. Es ist daher von grundsätzlichem Interesse zu wissen, wo Signalintegration erstmalig stattfindet. Die Darstellung der Abhängigkeiten zwischen synergistischen Molekülen hilft, integrierende Molekülknoten zu finden. Alle von diesen Knoten abhängigen Moleküle modulieren das Signal.

Um die zellulären Effekte synergistischer Signaltransduktion zu analysieren, werden die durch das Signal beeinflussten Gene untersucht. Gene sind der Speicher für Programme, welche beispielsweise Zellen dazu anregen, sich zu teilen, zu sterben oder auch andere Signale zu produzieren (Alberts *et al.*, 2008). Der Transfer auf die genomische Ebene gelingt mittels TFs. Diese können an der DNA binden und regulieren die betreffenden Gene. In Abb. 4.13 wird der Transfer durch die gepunktete Linie dargestellt. M3 wirkt als TF und reguliert die Expression von Gen A. Die Funktion der beeinflussten Gene determiniert die beobachtbaren zellulären Effekte. Die Charakterisierung von TFs und Genen gewährt Einblick

in die zu erwartenden Effekte synergistischer Signaltransduktion und kann somit nicht vernachlässigt werden. Aus den beschriebenen Gegebenheiten und der Interaktion zwischen verschiedenen Signalen leitet sich der folgende Graphen- bzw. Netzwerk-basierte Algorithmus ab.

### Ein Algorithmus zur Detektion von Synergien

Es sei  $\mathcal{G} := (\mathcal{V}, \mathcal{E})$  ein bipartiter Graph mit Knoten  $\mathcal{V}$  und Kanten  $\mathcal{E}$ . Die Knoten  $\mathcal{V}$  bestehen aus Molekülknoten  $\mathcal{M} := \{m_1, \dots, m_n\}$  und Reaktionsknoten  $\mathcal{R} := \{r_1, \dots, r_n\}$ , das heißt  $\mathcal{V} := \mathcal{M} \cup \mathcal{R}$ . Für jedes Molekül sei bekannt, ob es ein TF ist. Die Kanten  $\mathcal{E}$  des Graphen  $\mathcal{G}$  sind geordnete Paare  $(a, b) \in \mathcal{M} \times \mathcal{R} \cup \mathcal{R} \times \mathcal{M}$ . Es gilt also:  $G = (\mathcal{V}, \mathcal{E}) = (\mathcal{M} \cup \mathcal{R}, \mathcal{M} \times \mathcal{R} \cup \mathcal{R} \times \mathcal{M})$ . Außerdem ist die Menge von Stimuli ( $\mathcal{S} := \{s_1, \dots, s_n\} \in \mathcal{M}$ ) gegeben. Nicht im Graphen enthalten, aber für die Untersuchung von Synergien notwendig, ist die Anbindung an die genomische Ebene. Die Menge an geordneten Tupeln  $TF \rightarrow Gen := \{(a, b) \mid a \in \mathcal{M} \text{ ist ein TF, } b \text{ ist ein Gen}\}$  beschreibt diese, wobei der TF  $a$  das Gen  $b$  reguliert.

Das Ziel des Algorithmus ist die Identifikation von synergistisch beeinflussten Genen sowie die Darstellung der Abhängigkeiten zwischen synergistisch beeinflussten Molekülen, um ihre gegenseitige Regulation zu verstehen. Die jeweiligen Schritte des Algorithmus werden im Folgenden durch Kästen abgetrennt dargestellt, wobei Funktionen **fett** hervorgehoben werden. Die Eingabeparameter werden in Klammern dargestellt. Die folgende Box beschreibt die vier grundlegenden Schritte des Algorithmus.

**UntersucheSynergien**( $\mathcal{G}, \mathcal{S}, TF \rightarrow Gen$ )

- (1)  $\mathcal{K} :=$  **DetektiereSynergiekandidaten** ( $\mathcal{S}, \mathcal{G}$ )
- (2)  $\mathcal{K}' :=$  **EntferneUninformativeKandidaten** ( $\mathcal{G}, \mathcal{K}$ )
- (3)  $\mathcal{G}' :=$  **SynergistischeAbhängigkeiten** ( $s_1, \mathcal{K}', \mathcal{G}$ )
- (4)  $\mathcal{SG} :=$  **VerbindeTFsMitGenen** ( $\mathcal{K}', TF \rightarrow Gen$ )

In Rechenschritt (1) werden zuerst alle potentiellen synergistischen Kandidaten ( $\mathcal{K}$ ) anhand des Graphen  $\mathcal{G}$  bestimmt (siehe 4.3.1). Dabei werden Integrationsknoten gesucht, die von allen Stimuli gemeinsam erreichbar sind. Die bestimmte

Kandidatenmenge wird in synergistische Reaktions- und Molekülkandidaten unterteilt, wobei zu diesem Zeitpunkt noch nicht abschließend sicher ist, ob an den identifizierten Knoten Synergie vorliegt. In Rechenschritt (2) wird die Kandidatenmenge hinsichtlich eines Abbruchs der Signalweiterleitung und der Existenz von alternativen Wegen zwischen einem synergistischen Knoten und einem beliebigen Stimulus (siehe Unterabschnitt 4.3.2) kontrolliert. Dadurch werden Knoten, die nicht synergistisch sind, aus der Kandidatenmenge entfernt. Nach diesem Schritt sind alle synergistischen Knoten sicher detektiert. Um Knoten, an denen Synergie erstmalig auftritt, zu identifizieren wird in (3) ein Abhängigkeitsgraph konstruiert, welcher die Beziehungen zwischen den synergistisch beeinflussten Molekülen aufgeklärt (siehe 4.3.3). Nur Moleküle sind die experimentell nachweisbaren Träger von synergistisch induzierten Informationen. In Schritt (4) werden die Signale durch TFs auf die genomische Ebene transferiert (siehe 4.3.4). Dadurch werden die synergistischen Gene ( $\mathcal{SG}$ ) bestimmt.

Sind alle synergistisch regulierten Gene bekannt, erfolgt für die detektierten Gene eine Funktionsanalyse (siehe 4.3.5). Diese ist notwendig, um plausible Zusammenhänge zwischen den eingesetzten Stimuli und den detektierten, synergistisch beeinflussten Genen herzustellen. Beispielsweise sollten Immunstimuli auch immunrelevante Gene ansprechen. Durch Einsatz experimenteller Daten, Literatur, einschlägiger Datenbanken und anderer Methoden werden die Gene und ihre Funktionen verifiziert.

### 4.3.1 Detektion von Synergiekandidaten

Um Synergiekandidaten zu bestimmen, wird für jeden spezifischen Stimulus  $s_i$  der Graph  $\mathcal{G}$  mittels Tiefensuche<sup>48</sup> durchsucht. Dabei werden alle erreichbaren Knoten unabhängig vom Knotentyp in  $K_i$  gespeichert (2).

---

<sup>48</sup>Die Tiefensuche ist ein Standardverfahren, um in einem Graphen alle von einem spezifischen Knoten erreichbaren Knoten zu bestimmen (Goos, 2000; Cormen *et al.*, 2003).

**DetektiereSynergiekandidaten** ( $\mathcal{S}, \mathcal{G}$ )

- (1) Für alle  $s_i \in \mathcal{S}$
- (2) bestimme in  $\mathcal{G}$  alle von  $s_i$  aus erreichbaren Knoten  $K_i$
- (3)  $KR_i := K_i \cap \mathcal{R}(\mathcal{G})$
- (4)  $\mathcal{SR} := \bigcap_i KR_i$
- (5)  $\mathcal{SM} := \{m \in \mathcal{M}(\mathcal{G}) \mid \exists e \in \mathcal{E}(\mathcal{G}) \text{ mit } e = (a, m) \text{ und } a \in \mathcal{SR}\}$
- (6)  $\mathcal{K} := (\mathcal{SR}, \mathcal{SM})$
- (7) gib  $\mathcal{K}$  zurück

Anschließend wird aus der Menge  $K_i$  in (3) die Teilmenge an Knoten bestimmt, die Reaktionen ( $KR_i$ ) sind. Da für synergistische Reaktionskandidaten bekannt ist, dass eine Reaktion von allen Stimuli aus erreichbar sein muss (siehe 4.3), sind die synergistischen Reaktionen in der Schnittmenge aller durch einen Stimulus erreichbaren Reaktionen (4) enthalten. Synergistische Reaktionen projizieren ihren Effekt auf Produkte, daher werden in (5) die zu den Prozessen zugehörigen Produktmoleküle ( $\mathcal{SM}$ ) bestimmt. Die Kandidatenmenge aller synergistischen Knoten  $\mathcal{K}$  (6) setzt sich daher aus den synergistischen Reaktionen ( $\mathcal{SR}$ ) und den durch sie beeinflussten Molekülen ( $\mathcal{SM}$ ) zusammen. Die synergistischen Moleküle müssen jedoch noch auf einen alternativen Weg hin kontrolliert werden, was im nächsten Unterabschnitt (4.3.2) des Algorithmus geschieht.

Der in Abb. 4.14 illustrierte Graph veranschaulicht, wie **DetektiereSynergiekandidaten** die Menge  $\mathcal{K}$  für zwei Stimuli  $\mathcal{S} := \{S1, S2\}$  bestimmt. Für jeden Stimulus werden die jeweiligen Erreichbarkeitsmengen  $K_i$  und  $KR_i$  berechnet. Dabei wird zuerst  $S1$  und dann  $S2$  bearbeitet. In Abb. 4.14 sind die Mengen  $K_i$  und  $KR_i$  in der Tabelle dargestellt, wobei oberhalb der gestrichelten Linie  $S1$  bearbeitet wird. Die synergistischen Reaktionskandidaten ( $\mathcal{SR}$ ) bestimmen sich durch den Schnitt von  $KR_1$  und  $KR_2$  (4). Für die Ermittlung der synergistischen Molekülkandidaten ( $\mathcal{SM}$ ) werden all jene Kanten besucht, bei denen eine synergistische Reaktion den Ausgangspunkt bildet (5). Die detektierten  $\mathcal{SR}$  und  $\mathcal{SM}$  sind im Graphen (Abb. 4.14) blau markiert und tragen synergistische Informationen.

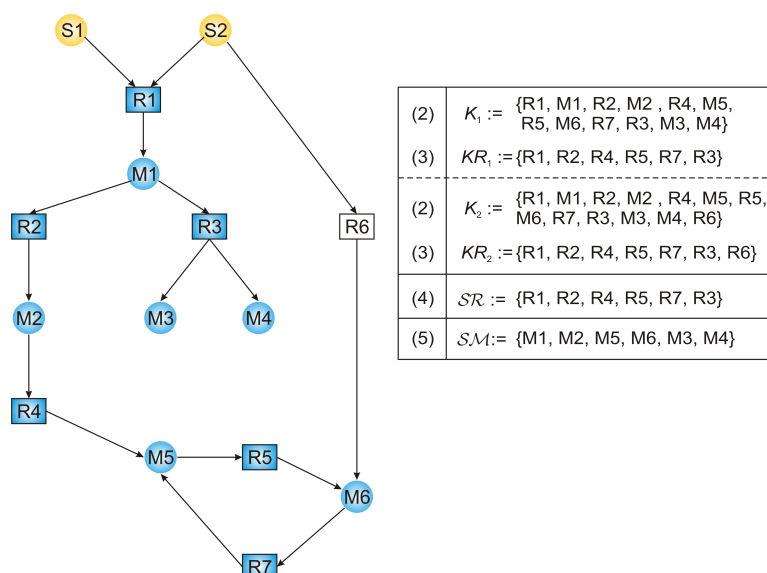


Abbildung 4.14: Beispielhafte Detektion von Synergiekandidaten durch den Algorithmus. Kandidaten, welche durch die Stimuli S1 und S2 (gelbe Knoten) synergistisch beeinflusst werden, sind blau hervorgehoben. Dabei wird unterschieden zwischen synergistischen Reaktionen (R) und Molekülen (M). Weiße Knoten bedeuten, dass keine Synergie vorliegt. Die Tabelle rechts stellt in der ersten Spalte den Rechenschritt für **DetektierereSynergiekandidaten**( $\mathcal{S}, \mathcal{G}$ ) dar und in der zweiten die jeweils berechneten Mengen. Die gestrichelte Linie unterteilt die Bearbeitung des jeweiligen Stimulus.

### 4.3.2 Entfernung uninformativer synergistischer Kandidaten

Nicht alle der bisher identifizierten synergistischen Moleküle in  $\mathcal{K}$  sind exklusiv durch synergistische Signalintegration erklärbar oder leiten das Signal in Richtung genomischer Ebene weiter. Beispielsweise bieten parallele Wege zwischen einem Stimulus und einem Molekül ( $\in \mathcal{SM}(\mathcal{K})$ ) eine alternative Erklärung an. Diese Wege werden identifiziert und Molekülkandidaten, die über einen solchen Weg verfügen, werden aus der Kandidatenmenge gelöscht. Ebenso wie synergistische Moleküle, an denen die Signalübertragung endet und die keine TFs sind. Diese Moleküle sind nicht mit weiteren Reaktionen verbunden und daher nicht in der Lage das Signal auf die genomische Ebene zu transferieren. Dieser Transfer ist jedoch unabdingbar für die Untersuchung von synergistischen Effekten, da heutzutage der experimentelle Nachweis der durch Stimuli ausgelösten Effekte zumeist auf der genomischen Ebene erfolgt. Nur TFs realisieren die Übertragung der Signale von der proteomischen Ebene auf die genomische Ebene (siehe Abb. 4.13, Knoten M3) und bleiben deshalb immer erhalten.

Synergistische Knoten, zu denen ein alternativer Weg existiert oder an denen das Signal abbricht, werden aus der Kandidatenmenge  $\mathcal{K}$  entfernt. Um diese Aufgaben zu lösen, werden zwei Hilfsfunktionen definiert. Die Erste, **EntferneKnoten** (Box unten), löscht einen gegebenen synergistischen Molekülknoten und aktualisiert die Kandidatenmenge.

**EntferneKnoten**( $\mathcal{G}, m \in \mathcal{SM}(\mathcal{K}), \mathcal{K}$ )

- (1)  $R' := \{r \in \mathcal{SR}(\mathcal{K}) \mid \exists e \in \mathcal{E}(\mathcal{G}) \text{ mit } e = (r, m)\}$
- (2)  $\forall r \in R'$
- (3)     wenn:  $\nexists e \in \mathcal{E}(\mathcal{G}) : e = (r, n)$ , mit  $n \in \mathcal{SM}(\mathcal{K})$  und  $n \neq m$
- (4)     lösche  $r$  aus  $\mathcal{SR}(\mathcal{K})$
- (5)  $R'' := \{r \in \mathcal{SR}(\mathcal{K}) \mid \exists e \in \mathcal{E}(\mathcal{G}) \text{ mit } e = (m, r)\}$
- (6)  $\forall r \in R''$
- (7)     lösche  $r$  aus  $\mathcal{SR}(\mathcal{K})$
- (8)     lösche  $m$  aus  $\mathcal{SM}(\mathcal{K})$
- (9)     gib  $\mathcal{K}$  zurück

Um einen Knoten  $m$  aus der Menge der synergistischen Molekülkandidaten  $\mathcal{SM}(\mathcal{K})$  zu entfernen, werden zuerst alle Reaktionen  $R'$  bestimmt, die das zu löschende Molekül erreichen (1). Wenn diese Reaktionen keine weiteren Kandidatenmoleküle außer  $m$  anbinden (3), propagieren sie nach dem Löschen von  $m$  keinen synergistischen Effekt mehr. Daher werden sie in (4) aus der Kandidatenmenge  $\mathcal{SR}(\mathcal{K})$  entfernt. Anschließend werden alle Reaktionen  $R''$  betrachtet, die von  $m$  direkt erreicht werden (5). Da  $m$  nach Abschluss des Entfernens nicht mehr synergistisch ist, werden alle von  $m$  abhängigen Reaktionen ebenfalls aus  $\mathcal{SR}(\mathcal{K})$  gelöscht (7). Abschließend wird in (8) das zu löschende Molekül  $m$  aus  $\mathcal{SM}(\mathcal{K})$  entfernt und das aktualisierte  $\mathcal{K}$  zurückgegeben.

Die zweite Hilfsfunktion **SucheAlternativenWeg** gibt an, ob ein alternativer Weg zwischen einem beliebigen Stimulus und einem Molekülknoten  $m$  existiert. Dabei darf der Weg jedoch nicht über eine bekannte synergistische Reaktion gehen.

**SucheAlternativenWeg**( $\mathcal{G}, m \in \mathcal{SM}(\mathcal{K}), \mathcal{S}, \mathcal{K}$ )

- (1)  $R' := \mathcal{R}(\mathcal{G}) \setminus \mathcal{SR}(\mathcal{K})$
- (2)  $E' := \mathcal{E}(\mathcal{G}) \setminus \{(a, b) \in \mathcal{E}(\mathcal{G}) \mid a \vee b \in \mathcal{SR}(\mathcal{K})\}$
- (3)  $G' := (R' \cup \mathcal{M}(\mathcal{G}), E')$
- (4)  $\forall s \in \mathcal{S}$
- (5)           wenn: in  $G'$  ein Weg von  $s$  zu  $m$  existiert
- (6)                   gib *wahr* zurück
- (7)   gib *falsch* zurück

$R'$  (1) ist die Menge an nicht-synergistischen Reaktionen und  $E'$  (2) die Menge aller möglichen Kanten, nachdem die synergistischen Reaktionen *verboten* wurden. In (3) wird ein neuer Graph ( $G'$ ) definiert, in dem alle Moleküle aus  $\mathcal{G}$  enthalten sind. Jedoch bleiben nur die nicht-synergistischen Reaktionsknoten  $R'$  (1) und die entsprechenden Kanten aus  $E'$  (2) erhalten. Somit sind nur nicht-synergistische Wege erlaubt. Um zu untersuchen, ob es einen alternativen Weg in  $G'$  gibt, wird für jeden Stimulus kontrolliert (4), ob dieser noch  $m$  erreicht (5). Ist dies der Fall, ist  $m$  nicht alleinig synergistisch beeinflussbar und es existiert ein alternativer, nicht synergistischer, Weg in  $G'$ .



Beide Hilfsfunktionen sind Teil von **EntferneUninformativeKandidaten** (Box unten), die alle synergistischen Molekülkandidaten  $\mathcal{SM}(\mathcal{K})$  hinsichtlich eines Abbruchs der Signalweiterleitung oder der Existenz eines alternativen Weges untersucht. Dadurch wird die Menge der Kandidaten  $\mathcal{K}$  verifiziert und Knoten, die nicht synergistisch sind, werden aus  $\mathcal{K}$  gelöscht.

**EntferneUninformativeKandidaten**( $\mathcal{G}, \mathcal{K}$ )

- (1)  $L := \mathcal{SM}(\mathcal{K})$
- (2) Solange  $L \neq \emptyset$
- (3)  $m$  ist das erste Element in  $L$
- (4) wenn:  $\exists e \in \mathcal{E}(\mathcal{G}) : e = (m, a)$ , mit  $a \in \mathcal{R}(\mathcal{G})$
- (5) wenn: **SucheAlternativenWeg**( $\mathcal{G}, m, \mathcal{S}, \mathcal{K}$ ) ist wahr
- (6)  $\mathcal{K} := \mathbf{EntferneKnoten}(\mathcal{G}, m, \mathcal{K})$
- (7)  $L_{neu} := \{\text{alle von } m \text{ aus erreichbaren Knoten in } \mathcal{G}\}$
- (8)  $L'_{neu} := L_{neu} \cap \mathcal{SM}(\mathcal{K})$
- (9) Hänge alle  $l \in L'_{neu} \notin L$  an  $L$  an
- (10) sonst:
- (11) wenn:  $m$  ist ein TF
- (12) wenn: **SucheAlternativenWeg**( $\mathcal{G}, m, \mathcal{S}, \mathcal{K}$ ) ist wahr
- (13)  $\mathcal{K} := \mathbf{EntferneKnoten}(\mathcal{G}, m, \mathcal{K})$
- (14) sonst:
- (15)  $\mathcal{K} := \mathbf{EntferneKnoten}(\mathcal{G}, m, \mathcal{K})$
- (16) lösche  $m$  aus  $L$
- (17) gib  $\mathcal{K}$  zurück

Für die Untersuchung wird in (1) eine Liste  $L$  als Kopie aller synergistischen Moleküle erzeugt. Aus dieser Liste wird sukzessive jedes einzelne Element  $m$  abgearbeitet und schließlich aus  $L$  gelöscht (16). Werden von  $m$  weitere Reaktionen erreicht (4), so wird die Existenz eines alternativen Weges geprüft (5). Gibt es diesen, so wird  $m$  aus  $\mathcal{K}$  gelöscht (6). Da wegen (4) noch weitere Teile des Graphen angebunden sind, werden diese bestimmt (7). Unter den bestimmten Knoten können sich synergistische Moleküle befinden (8), die möglicherweise schon bearbeitet wurden. Für diese Moleküle ändern sich die Möglichkeiten, sie zu erreichen. So können vor dem Löschen von  $m$  keine alternativen Wege über

direkt von  $m$  abhängige Reaktionen verlaufen, da diese als synergistisch angenommen werden und somit *verboten* sind (siehe **SucheAlternativenWeg** Zeile (1)). Nach dem Löschen von  $m$  aus  $\mathcal{K}$  verändern sich  $R'$  und  $E'$  in **SucheAlternativenWeg**, wobei nun direkt von  $m$  abhängige Reaktionen erreichbar sind. Diese Veränderung eröffnet gegebenenfalls neue Wege zu den von  $m$  aus erreichbaren synergistischen Molekülen. Daher müssen synergistische Moleküle ( $\in \mathcal{SM}(\mathcal{K})$ ), welche von  $m$  aus erreichbar sind (7) und schon bearbeitet wurden, neu evaluiert werden. Sie werden in (9) an  $L$  angehängt.

In den Zeilen 10-15 werden synergistische Moleküle kontrolliert, an denen die Signalweiterleitung abbricht. Diese Moleküle müssen TFs sein (11), damit das Signal auf die genomische Ebene transferiert werden kann. Nur an TF-Knoten wird die Existenz eines alternativen Weges kontrolliert (12). Ist das untersuchte Molekül kein TF (14) oder gibt es einen alternativen Weg zu diesem TF (12), wird es aus der Kandidatenmenge  $\mathcal{K}$  gelöscht.

Durch die Evaluation der synergistischen Molekülknoten nach den oben beschriebenen Schritten wird die Kandidatenmenge vollständig verifiziert. Das Löschen von  $sm$  aus  $L$  führt sukzessive dazu, dass  $L$  kleiner wird und der Algorithmus terminiert. An den verbleibenden Kandidaten ist Synergie sicher detektierbar.

Zur Veranschaulichung von **EntferneUninformativeKandidaten** wird erneut das Beispiel aus 4.3.1; Abb. 4.14 betrachtet. Neben den Ergebnissen aus der Detektion von Synergiekandidaten ( $\mathcal{K}$ ) sei bekannt, dass M4 ein TF ist. Das sukzessive Abarbeiten der synergistischen Moleküle in  $L$  (2) zeigt, dass weder M1, M2 oder M5 problematisch sind. An diesen bricht weder die Signalweiterleitung ab, noch existiert ein alternativer Pfad zu diesen Molekülen. Betrachtet man M6, so wird der alternative Pfad ( $S2 \rightarrow R6 \rightarrow M6$ ) identifiziert (5) und M6 wird gelöscht (6). Durch **EntferneKnoten** wird außerdem R5 und R7 aus  $\mathcal{SR}(\mathcal{K})$  gelöscht. Es werden die von M6 abhängigen Knoten bestimmt (7), wodurch M5 erneut in  $L$  eingefügt wird (9). Jetzt ist M5 alternativ über ( $S2 \rightarrow R6 \rightarrow M6 \rightarrow R7 \rightarrow M5$ ) erreichbar, wodurch M5, ebenso wie R4, aus  $\mathcal{K}$  entfernt wird. Nun wird M3 betrachtet, wobei hier die Signalweiterleitung abbricht, da das Molekül M3 an keine weitere Reaktion gebunden ist. M3 ist auch kein TF und wird daher gelöscht (15). Nicht betroffen ist R3, da mit M4 noch ein weiteres, ebenfalls synergistisches Molekül angebunden ist und R3 das synergistische Signal nicht nur auf M3 transferiert.

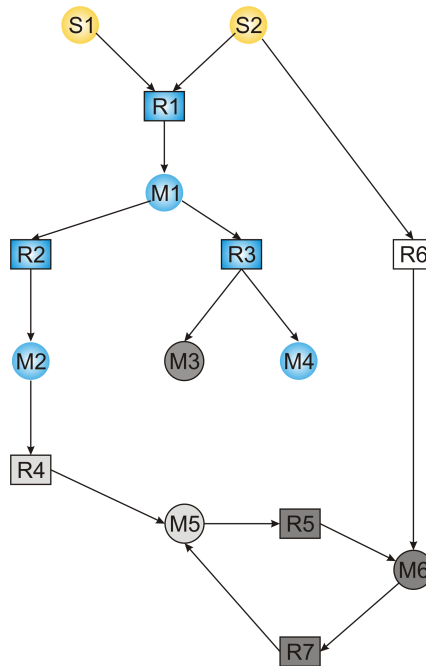


Abbildung 4.15: Algorithmische Entfernung von Synergiekandidaten. Blaue Knoten repräsentieren synergistisch beeinflusste Knoten nach Entfernung von uninformativen Kandidaten. Graue Knoten stellen ehemalige Synergiekandidaten dar (siehe 4.3.1). Die aufhellende Grau-Färbung symbolisiert Abhängigkeiten zwischen einzelnen Schritten (M5 wird erst entfernt, wenn M6 nicht mehr synergistisch ist).

Zuletzt wird M4 überprüft, wobei hier ebenfalls die Signaltransduktion abbricht. Jedoch ist M4 als TF bekannt und es existiert kein alternativer Pfad zu diesem Molekül, wodurch es nicht gelöscht werden darf (11). Abb. 4.15 zeigt die vollständige Menge an synergistisch beeinflussten Knoten im Graphen  $\mathcal{G}$ .

### 4.3.3 Detektion von Abhängigkeiten zwischen synergistischen Molekülen

Zu wissen, an welchen Molekülen Synergie erstmalig nachweisbar ist und in welchen Abhängigkeiten sie zueinander stehen, ist wichtig um zu verstehen, wie verschiedene Signale miteinander interagieren, wie sie sich gegenseitig modulieren und welche biologischen Effekte durch sie ausgelöst werden. Die Funktion **SynergistischeAbhängigkeiten** bestimmt ausgehend von einem gegebenen Stimulus  $s_1 \in \mathcal{S}$  die Abhängigkeiten zwischen den synergistischen Molekülen. Mittels einer Tiefensuche werden dabei die Beziehungen zwischen verschiedenen synergistischen Molekülen aufgeklärt, wobei die Anzahl der von einem Knoten aus erreichbaren synergistischen Moleküle bestimmt wird. Die folgende Box beschreibt den Eingang in die rekursive Tiefensuche.

**SynergistischeAbhängigkeiten**( $s_1 \in \mathcal{S}, \mathcal{K}, \mathcal{G}$ )

- (1)  $E' := \emptyset$
- (2)  $AK := \emptyset$
- (3)  $G' := (AK, E')$
- (4) **Abhängigkeiten**( $s_1, s_1, \mathcal{K}, G', \mathcal{G}$ )
- (5)  $G'' := (AK, E' \setminus \{(s_1, a) \in E'\})$
- (6) gib  $G''$  zurück

Die definierten leeren Mengen  $E'$  (1) und  $AK$  (2) werden zum Abhängigkeitsgraphen  $G'$  zusammengefasst (3). Im Rekursionsschritt (4) wird  $G'$  fortwährend aktualisiert. Da  $s_1$  selbst kein synergistisches Molekül ist, es jedoch alle anderen synergistischen Moleküle erreicht, und in  $G'$  von  $s_1$  ausgehende Kanten enthalten sind, werden diese in (5) ausgeschlossen, wodurch nur detektierte synergistische Moleküle und ihre Beziehungen zueinander dargestellt werden. Anschließend wird der so erzeugte Abhängigkeitsgraph  $G''$  zurückgegeben (6).

In der Funktion **Abhängigkeiten** findet die Rekursion statt. Die Tiefensuche besteht aus zwei Teilen. Im ersten Teil wandert man entlang des Graphen  $\mathcal{G}$  in die Tiefe, wobei die Beziehungen der synergistischen Moleküle zueinander aufgeklärt werden. Im zweiten Teil wandert man zurück zum Ausgangspunkt, wobei für jedes synergistische Molekül die Anzahl an anderen erreichbaren synergistischen Molekülen bestimmt wird.

**Abhängigkeiten**(*letzter*, *aktueller*,  $\mathcal{K}$ ,  $G'$ ,  $\mathcal{G}$ )

- (1)  $wert := 0$
- (2) wenn:  $E := \{e \in \mathcal{E}(\mathcal{G}) \mid e = (aktueller, a)\} \neq \emptyset$
- (3)  $\forall e \in E$
- (4) wenn:  $a \in \mathcal{SM}(\mathcal{K})$
- (5)  $E'(G') := E'(G') \cup (letzter, a)$
- (6)  $wert := wert + \mathbf{Abhängigkeiten}(a, a, \mathcal{K}, G', \mathcal{G})$
- (7) sonst:
- (8)  $wert := wert + \mathbf{Abhängigkeiten}(letzter, a, \mathcal{K}, G', \mathcal{G})$
- (9) wenn:  $aktueller \in \mathcal{SM}(\mathcal{K})$
- (10)  $AK(G') := AK(G') \cup (aktueller, wert)$
- (11) inkrementiere  $wert$  um 1
- (12) gib  $wert$  zurück
- (13) sonst:
- (14) wenn:  $aktueller \in \mathcal{SM}(\mathcal{K})$
- (15)  $AK(G') := AK(G') \cup (aktueller, wert)$
- (16) inkrementiere  $wert$  um 1
- (17) gib  $wert$  zurück

Die wichtigsten Parameter der Funktion **Abhängigkeiten** sind der Eingabeparameter *letzter* und der Rückgabeparameter *wert*. Ersterer benennt den zuletzt besuchten synergistischen Knoten und *wert* repräsentiert die Anzahl der vom *aktuellen* Knoten aus erreichbaren synergistischen Moleküle. Beim Wandern in die Tiefe ist noch nicht bekannt, wie viele andere synergistische Moleküle erreicht werden, daher wird *wert* auf 0 gesetzt (1). Existieren am aktuellen Knoten weitere Kanten, so werden diese in den Rechenschritten 3-12 verfolgt. Ist das Ziel einer in (3) untersuchten Kante ein synergistisches Molekül (4), wird eine Kante

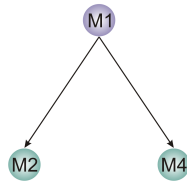


Abbildung 4.16: Abhängigkeiten zwischen den bestimmten synergistischen Molekülen im Beispiel (siehe Abb. 4.15). M1 beeinflusst die anderen synergistischen Moleküle M2 und M4.

zwischen dem letzten bekannten synergistischen Molekül und dem gefundenen in  $E'(G')$  eingefügt (5). Anschließend wird der Pfad weiter untersucht, wobei das gefundene synergistische Molekül nun das letzte Bekannte ist (6). Wird ein Knoten, der über keine weitere Kante verfügt (13-16) erreicht, so beginnt der Rückweg. Dabei werden sukzessiv alle synergistischen Molekülknoten betrachtet und gezählt. Dadurch ist an jedem Knoten bekannt, wie viele andere synergistische Knoten erreicht werden (6 oder 8). Die Tiefensuche garantiert für Knoten mit mehr als einer ausgehenden Kante, dass alle Kanten betrachtet wurden, bevor der jeweilige synergistische Knoten mit *wert* in  $AK(G')$  eingetragen (10,15) wird. Außerdem wird durch die Tiefensuche garantiert, dass keine Zyklen<sup>49</sup> permanent durchlaufen werden. Der *wert* eines Knotens ergibt sich dabei über die Summe der von diesem Knoten erreichbaren anderen synergistischen Knoten (6,8). Wird  $s_1$  erneut erreicht, ist  $\mathcal{G}$  komplett untersucht. Die Abhängigkeiten der synergistischen Knoten zueinander sind bestimmt und werden in  $G'$  abgebildet.

Aus dem Beispiel in Abb. 4.15 sind nur die Molekülknoten M1, M2 und M4 relevant. Nach Initialisierung von **Synergistische Abhängigkeiten** mit  $S1, \mathcal{K}, \mathcal{G}$  wird der Knoten R1 erreicht, wobei  $(S1, M1)$  in  $E'(G')$  eingefügt wird (5). Da M1 an weitere Reaktionen gebunden ist, wird der Pfad weiter verfolgt. Bei R2 wird die Kante  $(M1, M2)$  in  $E'(G')$  eingefügt (5). Anschließend wird M2 weiter untersucht und der Pfad wird weiterverfolgt. Jedoch befinden sich unterhalb von M2 keine weiteren synergistischen Moleküle. Daher wird beim Abschluss der Untersuchung von M2 das Tupel  $(M2,0)$  in  $AK(G')$  eingefügt (10). An den Knoten R2 und später M1 wird *wert* = 1 zurückgegeben (11,12). Es wird R3 betrachtet, wobei  $(M1, M4)$  in  $E'(G')$  eingefügt wird (5). Bei der Betrachtung von M4 ist

<sup>49</sup>Zyklen werden durch Markierung der schon besuchten Knoten ausgeschlossen. Aus Gründen der Verständlichkeit wird dies jedoch nicht dargestellt.

keine weitere Kante verfügbar, deshalb wird das Tupel (M4, 0) in  $AK(G')$  eingefügt (15). Man geht zurück zum Knoten M1, welcher nun komplett untersucht ist. Das Tupel (M1, 2) wird in  $AK(G')$  eingefügt. Wird S1 erreicht, werden alle von  $s_1$  ausgehenden Kanten in  $E'(G')$  (siehe **Synergistische Abhängigkeiten** Schritt(5)) gelöscht und man erhält Abb. 4.16. M1 ist das erste Molekül, welches synergistische Effekte vermittelt. Außerdem beeinflusst es die beiden Moleküle M2 und M4. Ein Beispiel zum erweiterten Verständnis findet sich im Appendix (siehe S. 157).

#### 4.3.4 Transkriptionsfaktoren propagieren Synergien auf die genomische Ebene

Um die detektierten Effekte an zelluläre Funktionen zu binden, ist die Rückkopplung auf die genomische Ebene notwendig. Hier werden die verschiedenen genetischen Programme, die beispielsweise die Produktion von Proteinen oder Hormonen kontrollieren, gesteuert (Alberts *et al.*, 2008). Durch Transkriptionsfaktoren sind die beiden Ebenen miteinander verbunden.

**VerbindeTFsMitGenen ( $\mathcal{K}$ ,  $TF \rightarrow Gen$ )**

- (1)  $\mathcal{SG} := \{b \mid (a, b) \in TF \rightarrow Gen, a \in \mathcal{SM}(\mathcal{K}), b \text{ ist ein Gen}\}$
- (2) gib  $\mathcal{SG}$  zurück

Um die detektierten synergistischen Moleküle auf Gene zu mappieren, werden in  $TF \rightarrow Gen$  alle Tupel  $(a, b)$  betrachtet, bei denen  $a \in \mathcal{SM}(\mathcal{K})$  gilt, also in denen  $a$  ein synergistisches Molekül und gleichzeitig ein TF ist (1). Alle durch  $a$  regulierten Gene  $b$  werden als synergistisch regulierte Gene ( $\mathcal{SG}$ ) zurückgegeben (2).

Durch Rauschen oder ungenaue Informationen in den genutzten Wissensressourcen ist mit falsch positiv identifizierten Genen zu rechnen. Zur Verbesserung der Vorhersagequalität müssen die identifizierten Gene verifiziert werden. Dies geschieht beispielsweise durch den Einsatz von experimentellen Genexpressionsdaten. Anhand dieser werden vorhergesagte Gene gefiltert, wodurch beispielsweise in einem Gewebe nicht-exprimierte Gene entfernt werden. Ergänzend dazu

müssen die TFs hinsichtlich ihrer Funktion charakterisiert werden. Es ist zu erwarten, dass bei einer Analyse der Immunantwort TFs identifiziert werden, die mit der Immunantwort korreliert sind. Insgesamt ist festzuhalten, dass die Validierung, auch von Teilergebnissen, das Vertrauen und die Sicherheit der durchgeführten Analyse erhöht.

### 4.3.5 Funktionen synergistisch beeinflusster Gene

Um Signaltransduktion zu verstehen, müssen die im vorherigen Unterabschnitt identifizierten Gene ( $\mathcal{SG}$ ) eine Rolle in den stimulierungsvermittelten Interaktionen spielen. Diese Rolle wird mittels Recherche von Literatur und Datenbanken oder Gene Set Enrichment Analysis<sup>50</sup> (GSEA), beispielsweise durch Verwendung von DAVID<sup>51</sup> und anderer Methoden, untersucht. Durch DAVID und einer GSEA werden GO-Terme<sup>52</sup> oder andere Deskriptoren gesucht, die unter den Beschreibungen der detektierten Gene überrepräsentiert sind (Beißbarth und Speed, 2004; Subramanian *et al.*, 2005; Huang *et al.*, 2007; Huang, Sherman und Lempicki, 2009). Die genaue Analyse hängt jedoch vom Fokus der gesamten Untersuchung ab und muss darauf zugeschnitten sein. Zum Verständnis synergistisch induzierter Effekte und zur Vorhersage der Effekte ist die funktionale Untersuchung unerlässlich. Aufgrund der Heterogenität der möglichen Fragestellungen wird kein generelles Ablaufschema für die Zuordnung von Funktionen vorgestellt. In dieser Arbeit wurde die Funktion synergistisch regulierter Gene durch Analyse der KEGG-Datenbanken bestimmt (siehe Unterabschnitt 3.1.3 sowie Seite 55 ff. in 4.1.1 oder Seite 111 ff. in 4.3.6).

---

<sup>50</sup>Bei der Gene Set Enrichment Analysis werden einer gegebenen Menge an Genen Funktionen zugewiesen. Die Zuweisung basiert dabei auf verfügbaren Datenressourcen wie beispielsweise KEGG oder anderen Datenbanken (Beißbarth und Speed, 2004; Subramanian *et al.*, 2005).

<sup>51</sup>Bei DAVID handelt es sich um ein Web-Interface-gesteuertes System zur Detektion überrepräsentierter Terme mittels GSEA, basierend auf der Auswahl diverser Datenbanken (Dennis *et al.*, 2003; Huang *et al.*, 2007; Huang, Sherman und Lempicki, 2009). Ziel ist es, in einem gegebenen Genset funktionale Gegebenheiten zu detektieren (Huang *et al.*, 2007). DAVID implementiert dabei verschiedene GSEA-Methoden.

<sup>52</sup>GO-Terme beschreiben ein Gen hinsichtlich Lokalisation, Funktion, damit verbundenen Krankheiten oder anderen Parametern (Ashburner *et al.*, 2000; Harris *et al.*, 2004).



### 4.3.6 Anwendung des Algorithmus auf das IL-3- und Flu-spezifische Signaltransduktionsnetzwerk

Der in 4.3.1-4.3.5 vorgestellte Algorithmus wird nun auf das IL-3- und Flu-Signalnetzwerk (siehe Unterabschnitt 4.2.5) angewandt. Dabei wurden weit mehr als 2000 synergistische Moleküle im gesamten Signaltransduktionsnetzwerk identifiziert, wobei 1157 synergistische TFs in der Lage sind, das Signal auf die genomische Ebene zu übertragen. Nur die durch diese TFs regulierten Gene lassen sich anhand der vorliegenden Experimentaldaten verifizieren (siehe Abschnitt 3.4), weshalb sich die weitere Analyse auf TFs und die durch sie regulierten Gene beschränkt (siehe dazu 4.3.4 und 4.3.5).

Die hohe Anzahl an identifizierten TFs war unerwartet, da TFs nur einen kleinen Teil der an der Signaltransduktion beteiligten Moleküle darstellen. Deshalb wurde untersucht, ob regulatorische Abhängigkeiten zwischen diesen TFs existieren. Hierzu eignet sich die in 4.3.3 vorgestellte Methodik. Jedoch ist die Darstellung von 1157 verschiedenen TFs und all ihren Interaktionen unübersichtlich und schlecht auswertbar, weshalb die Methode modifiziert wurde. TFs werden dabei anhand der Anzahl der durch sie regulierten TFs in Regulationsebenen gruppiert. Die wichtigsten TFs, welche die stärkste Regulationswirkung ausüben, also die meisten anderen TFs regulieren, werden in der obersten Ebene (kleinste Nummerierung) zusammengefasst und bilden die Hauptregulationsebene (HRE). Die Verbindungen zwischen den verschiedenen Ebenen zeigen, wie stark eine Ebene auf eine andere einwirkt, d.h. wie viele regulatorische Prozesse zwischen ihnen bekannt sind. Dadurch werden Abhängigkeiten zwischen den Ebenen komprimiert und verständlich dargestellt sowie wichtige Regulationsebenen identifiziert. Zur Veranschaulichung wird die Umwandlung eines Abhängigkeitsgraphen (nach 4.3.3) in den Regulationsgraphen in Abb. 4.17 beispielhaft dargestellt. Wie zu sehen ist, sind einige Ebenen mit sich selbst verbunden, was bedeutet, dass sich die TFs gegenseitig erreichen und beeinflussen, wodurch graphentheoretisch eine starke Zusammenhangskomponente vorliegt. Möglicherweise sind diese Ebenen besonders wichtig für die Regulation der propagierten Signale.

Die Analyse der 1157 synergistischen TFs ergibt, dass sich diese auf 10 verschiedene Regulationsebenen verteilen, wobei die 71 TFs der HRE sich gegenseitig sowie die anderen Ebenen regulieren (Abb. 4.18). An den TFs der HRE findet

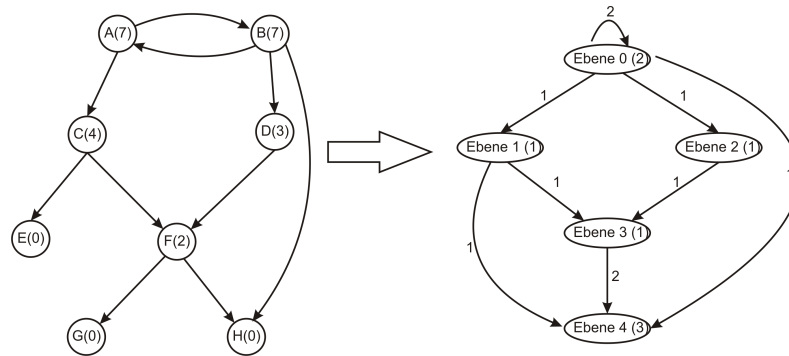


Abbildung 4.17: Beispielhafte Transformation eines Abhängigkeitsgraph (links) (nach 4.3.3) in den Regulationsgraph (rechts). Die Anzahl anderer erreichbarer synergistischer Knoten im Abhängigkeitsgraph wird in Klammern dargestellt. Im Regulationsgraph werden in Klammern die in einer Regulationsebene zusammengefassten Moleküle dargestellt. Kantenmarkierungen zeigen an, wie viele Prozesse die jeweiligen Ebenen miteinander verbinden. Die Regulationsebenen sind nach ihrer Wichtigkeit gewertet, je kleiner die Nummer, desto wichtiger ist die Ebene.

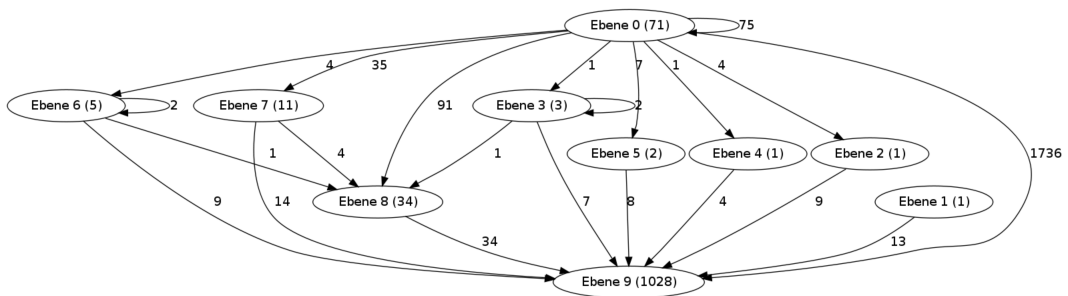


Abbildung 4.18: Beziehungen zwischen den Regulationsebenen der synergistisch durch IL-3 und Flu beeinflussten TFs in pDC. Die 71 TFs in Ebene 0 regulieren sich selbst (75 Prozesse) sowie die anderen TFs. Die Regulationsebenen 1-8 verteilen die Signale und bieten Möglichkeiten zur weitergehenden Signalmodulation. Ebene 9 wird von den anderen Ebenen reguliert, wobei sie mit Ebene 0 direkt über 1736 Prozesse verbunden ist.

	Parameter	Ebene 0	Ebene 5	Ebene 7	Ebene 8	Ebene 9
1 a	#TFs	71	2	11	34	1028
1 b	#Komplexe	24	0	4	17	912
2 a	an Gene gebundene TFs (44)	18	1	3	1	21
2 b	davon Komplexe	1	0	3	0	20
3	angebundene Gene (329)	310	5	17	1	48

Tabelle 4.3: Beziehungen zwischen TFs und den dadurch regulierten Genen. 1) Die Verteilung von TFs (a) und ihren Komplexen (b) im Signaltransduktionsnetzwerk. 2) TFs, welche an Gene gebunden werden (a), wobei Komplexe (b) separat dargestellt sind. 3) Die Anzahl der durch TFs in einer Regulationsebene angebundenen Gene. Auf Ebene 9 befinden sich 912 Komplexe, von denen 48 Gene reguliert werden. Die 18 TFs der Hauptregulationsebene (HRE) regulieren 310 Gene, wobei 17 unkomplezierte TFs 302 Gene regulieren (nicht gezeigt).

daher die synergistische Signalintegration statt. Anschließend wird das Signal auf die anderen Ebenen propagiert, wo es beispielsweise durch Phosphorylierungen und Komplexbildungen moduliert wird. Ebene 9 stellt die Zielebene der Signalmodulation dar und sollte die Verbindung zur genomischen Ebene herstellen, welche die für die Antwort relevanten Gene reguliert. In Abb. 4.18 ist eine starke Kopplung zur HRE zu erkennen, diese umfasst mehr als 1700 Prozesse. Dies unterstreicht die Bedeutung dieser beiden Ebenen. Eventuell wird das modifizierte Signal hier einer weiteren Feinjustierung oder einer Kontrolle, beispielsweise der weiteren Notwendigkeit einer Antwort, unterzogen.

Betrachtet man jedoch die Anbindung aller TFs an ihre korrespondierenden Gene, so entfalten nur 44 TFs eine genregulatorische Wirkung. Die Zielgene der restlichen identifizierten TFs erscheinen zuerst unbekannt. Auffällig ist, dass die Mehrheit der TFs in komplexierter Form vorliegen (siehe Tab. 4.3). Dies gilt sowohl für die an Gene gebundenen TFs (24 Komplexe) wie für TFs (973 Komplexe) insgesamt.

### Komplexe und ihre Rolle bei der Signaltransduktion

Von den an Gene gebundenen komplexierten TFs werden nur etwa 20% der erreichbaren Gene reguliert (siehe Tab. 4.3). Im Vergleich dazu regulieren die 17 unkomplexierten TFs der HRE 302 verschiedene Gene. Dies ist insofern erstaun-

lich, da Komplexe einerseits die Mehrheit der identifizierten TFs darstellen und andererseits nicht in der Lage sind, das Signal auf die genomische Ebene zu transferieren.

Ein möglicher Grund hierfür ist, dass die TF-enthaltenden Komplexe im Cytoplasma verbleiben und nicht in den Zellkern gelangen. Importine, welche den Transport in den Nukleus steuern, können möglicherweise nicht an diesen TFs binden, weshalb sie nicht in den Zellkern transportiert werden (Watson *et al.*, 2008; Alberts *et al.*, 2008). Eine andere Erklärung ist die durch Komplexierung veränderte Tertiär-<sup>53</sup> oder Quartärstruktur<sup>54</sup>. Durch konformationale Wechselwirkungen können Bindedomänen maskiert und inaktiviert werden, wodurch DNA-Bindungen unterbunden werden (Nelson und Cox, 2001; Watson *et al.*, 2008; Alberts *et al.*, 2008). In beiden Fällen nehmen komplexierte TFs eine wichtige Rolle bei der Signalregulation ein, wobei durch diese Mechanismen Signale gefiltert, gepuffert oder moduliert werden. Beispielsweise können sich so inaktive TFs ansammeln und durch verzögerte Aktivierung konzentrations- oder zeitabhängige genetische Prozesse auslösen (Giorgetti *et al.*, 2010). Außerdem konkurrieren verschiedene TFs auch um die Bindung an die DNA. Die Komplexierung zweier TFs, deren DNA-Bindestellen in sterischer Nähe<sup>55</sup> liegen, kann die Bindespezifität erhöhen, wodurch sich die Wahrscheinlichkeit einer erfolgreichen kooperativen DNA-Bindung im kompetitiven Umfeld der Zelle erhöht (Giorgetti *et al.*, 2010; Whittington *et al.*, 2011).

Die konträre, aber gleichwohl wahrscheinlichere, Erklärung für die Akkumulation von Komplexen und deren Unfähigkeit Signale zu transferieren ist, dass Komplexe und ihre Strukturen nicht ausreichend untersucht sind, wodurch sie in den einschlägigen Datenbanken nur unzureichend repräsentiert werden. So sind oft die durch Komplexe induzierten Effekte nicht bekannt oder sie werden unter der Wirkweise einfacher TFs in den Datenbanken subsumiert.

Für die Bestimmung der durch Komplexe ausgelösten Effekte muss bekannt sein, welche Domänen mit der DNA interagieren und welche Konformation vor-

---

<sup>53</sup>Mit Tertiärstruktur bezeichnet man den räumlichen konformationellen Aufbau von Proteinen (Nelson und Cox, 2001; Smith *et al.*, 2003).

<sup>54</sup>Unter Quartärstruktur versteht man die räumliche Anordnung von Makromolekülen wie Proteinen, wobei hier die Anordnung der Untereinheiten im Vordergrund steht, ohne dabei die interne Struktur der Untereinheiten zu berücksichtigen (Nelson und Cox, 2001; Smith *et al.*, 2003).

<sup>55</sup>Beispielsweise ist der AP-1-TF ein Homo- oder Heterodimer, bestehend aus Jun- oder Fos-TFs (Karin, Liu und Zandi, 1997).

liegt. Mittels ChIP-Seq<sup>56</sup> lassen sich Bindungsstellen an der DNA bestimmen, wobei Antikörper das zugehörige Protein, aber nicht die vorliegende Konformation detektieren (Johnson, Martazavi und Myers, 2007; Whittington *et al.*, 2011). Komplexe werden daher nicht direkt betrachtet, sondern mittels computergestützter Methoden, wie SpaMo, vorhergesagt (Whittington *et al.*, 2011). Zur experimentellen Untersuchung von Interaktion innerhalb von Komplexen dient das Mammalian-two-hybrid-System (Ravasi *et al.*, 2010), wobei keine Informationen über DNA-Bindestellen bestimmt werden (Ravasi *et al.*, 2010; Whittington *et al.*, 2011). Die Informationen der verschiedenen Quellen werden nur selten zusammengeführt, wodurch komplexierte TFs bisher nicht vollständig erfasst werden. Die vorgestellten Ergebnisse und die große Menge an komplexierten TFs (siehe Tab. 4.3) unterstreichen die Notwendigkeit einer genaueren Aufklärung zur Rolle von Komplexen in Signaltransduktion und Genregulation.

#### **Die Hauptregulationsebene induziert genomische Funktion und Immunrelevanz**

Wie in Tab. 4.3 gezeigt, wirken die 18 TFs der HRE auf 310 (> 90%) verschiedene Gene ein. Alle anderen Ebenen zusammen wirken auf weniger als ein Viertel der erreichbaren Gene ein, wobei ein Großteil der Regulation durch Ebene 9 realisiert wird. Es zeichnet sich ein schematisches Bild (Abb. 4.19), bei dem die HRE die wesentlichen synergistischen Signale auf die genomische Ebene transferiert, ab. In der Modulationsebene (Ebene 1-8) wird das Signal verändert und später auf die Zielebene (Ebene 9) propagiert. Dabei wirkt die HRE sehr stark auf die Zielebene ein (siehe Abb. 4.18), wobei sich die TFs gegenseitig beeinflussen. So können beispielsweise zusätzliche Gene aktiviert werden.

In den meisten Fällen findet jedoch eine Co-Regulation von Genen statt, wobei keine neuen Funktionen hinzugefügt werden (Daten nicht gezeigt). Dabei wirken die TFs der Zielebene auf Gene ein, die ebenfalls durch die HRE angesprochen werden, wodurch eine fein-justierte Genregulation gewährleistet wird. Dies zeigt

---

<sup>56</sup>ChIP-Seq: Beschreibt ein Verfahren, in dem eine Protein-DNA-Verbindung fixiert wird, wodurch die bindenden Sequenzen identifiziert werden. Das bindende Protein-DNA-Konglomerat wird dabei mittels Chromatin-Immunoprecipitation markiert und isoliert. Das Protein wird anschließend entfernt, wodurch das bindende DNA-Fragment sequenziert wird (Johnson, Martazavi und Myers, 2007).

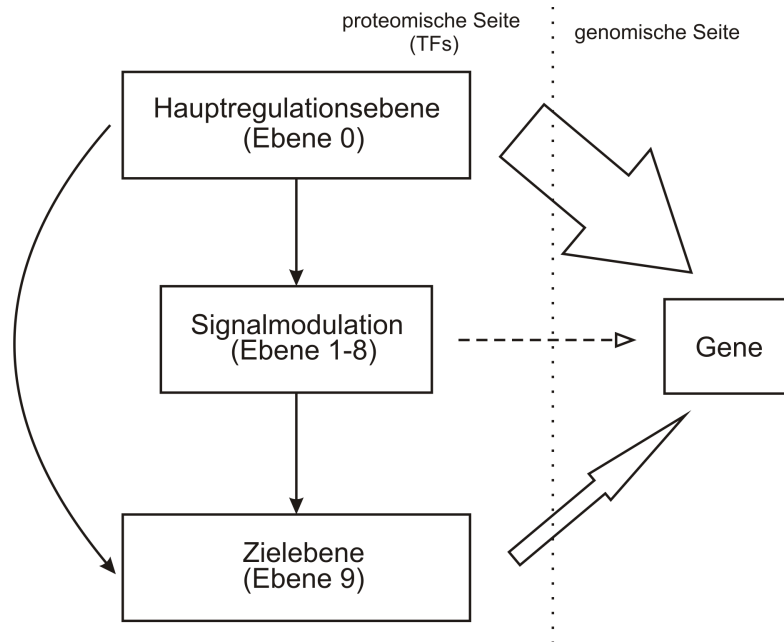


Abbildung 4.19: Schema der proteomischen Regulation zwischen den TFs (siehe auch Abb. 4.18). Rechts ist die genomische Ebene dargestellt. Durch die Größe der Pfeile wird abgebildet, welche Ebenen den Transfer von synergistischen Informationen auf Gene leisten.

sich auch in Abb. 4.19, da vor allem Hauptregulations- und Zielebene gut mit der Genomischen verbunden sind, während die Modulationsebene kaum angebunden ist. Die Zielebene setzt sich im Wesentlichen aus Komplexen zusammen, die, wie schon beschrieben, unzureichend untersucht sind. Bei der Untersuchung der Funktion der aktivierten Gene müssen daher Ziel- und Modulationsebene nicht betrachtet werden. Dennoch sind hier in Zukunft interessante Beobachtungen möglich, wenn sich die Datenlage insgesamt verbessert. So sollte die Zielebene an Bedeutung gewinnen und besser mit der genomischen Ebene verknüpft sein, wenn die Aufklärung von Komplexen und ihr jeweiliger Bezug zur genomischen Ebene besser verstanden sind. Funktionen, die momentan noch unter der Wirkung der einfachen (unkomplexierten, unmodifizierten) TFs subsumiert werden, können dann möglicherweise voneinander unterschieden werden. Für die Plausibilitätsprüfung der Funktion der TFs ist die Betrachtung der HRE hinreichend. Daher sollte für die 18 algorithmisch detektierten TFs der HRE ein Zusammenhang zwischen beschriebener Funktionalität und der Immunantwort bestehen.

Name	Signalpfad in KEGG	Literaturnachweis
IkappaB-alpha (IKK $\alpha$ )	B Cell receptor signaling	(Clément, Meloche und Servant, 2009)
HSF1	Legionellosis	(Singh und Aballay, 2006)
STAT1	TLR signaling	(Nagata und Todokoro, 1996)
STAT3	Jak-STAT signaling	(Liu <i>et al.</i> , 2011)
STAT5A	Jak-STAT signaling	(Yao <i>et al.</i> , 2006)
STAT5B	Jak-STAT signaling	(Yao <i>et al.</i> , 2006)
CREB1	Antigen processing/presentation	(Wen, Sakamoto und Miller, 2010)
CBP (CREBBP)	Jak-STAT signaling	(Levy, 2007)
<i>c-Ets-1</i>	Cancer	(Russell und Garrett-Sinha, 2010)
<i>RXR-alpha (NR2B1)</i>	Adipocytokine signaling	(Hettne <i>et al.</i> , 2007)
NF-AT2 (NFAT)	B Cell receptor signaling	(Monticelli und Rao, 2002)
HIF-1alpha	mTor signaling	(Sitkovsky und Lukashev, 2005)
<i>MEF-2D</i>	-	(Aude-Garcia <i>et al.</i> , 2010)
NF-kappaB1	TLR signalling	(Pikarsky <i>et al.</i> , 2004)
p53{p}	p53 signaling	(Ishida <i>et al.</i> , 1999)
STAT3{p}	Jak-STAT signaling	(Liu <i>et al.</i> , 2011)
CREB1{p}	Antigen processing/presentation	(Wen, Sakamoto und Miller, 2010)
RelA-p65:NF-kappaB1	TLR signalling	(Tripathi und Aggarwal, 2006)

Tabelle 4.4: Immunrelevanz der 18 synergistisch regulierten TFs der Hauptregulationsebene. Die identifizierten TFs wurden mit Hilfe der Signalpfade aus KEGG (siehe Unterabschnitt 3.1.3) und Literatur verifiziert. TFs, bei denen eine Verbindung zur Immunantwort unsicher erscheint, werden kursiv dargestellt.

Dazu wurde die Präsenz der TFs in Signalpfaden der KEGG-Datenbanken (siehe 3.1.3) sowie in Literatur, bezüglich ihrer Funktion in der Immunantwort oder Immunregulation untersucht (siehe Tab. 4.4) (Kanehisa *et al.*, 2008). Dabei zeigt sich, dass für jeden TF eine Rolle in der Immunantwort angenommen werden kann.

Die größte Gruppe der identifizierten TFs waren STATs<sup>57</sup>. Nach Aktivierung der IL-3- und Flu-Signal-Rezeptoren dimerisieren STATs an ihnen und bilden Komplexe aus zwei Untereinheiten. Anschließend lösen sie sich ab und wandern direkt in den Zellkern, wo sie ihre Wirkung entfalten (Darnell, Kerr und Stark, 1994; Frank, 1999; Levy und J. E. Darnell, 2002; Yao *et al.*, 2006). Interaktionen mit den hier identifizierten STATs spielen eine wichtige Rolle bei der durch IL-3 induzierten Signalprozessierung (Daley und Baltimore, 1988; Chaturvedi, Reddy und Reddy, 1998; Bright *et al.*, 2004; Yao *et al.*, 2006). Außerdem sind STATs Komplexe (Martinez2003, Frank1999), die jedoch nicht als solche annotiert sind. Dies unterstreicht erneut, wie wichtig Komplexe sind und wie wichtig eine genaue Untersuchung und Repräsentation von Komplexen ist (siehe auch Seite 103 ff.).

<sup>57</sup>STAT *engl.* für signal transducer and activator of transcription. Eine TF-Familie, welche die Antwortkaskade auf viele verschiedene Cytokine und Wachstumsfaktoren vermittelt (Smith *et al.*, 2003).

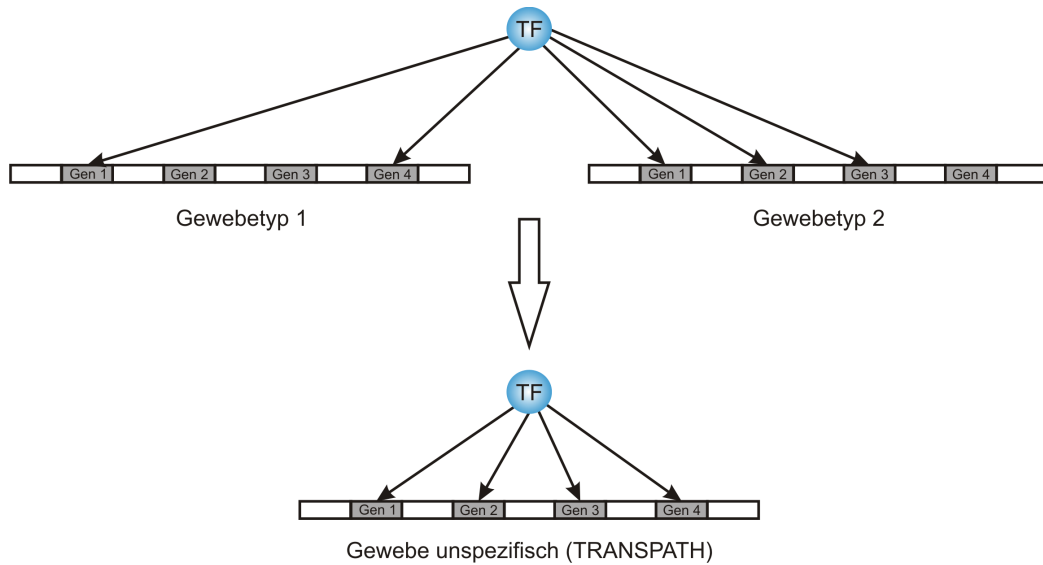


Abbildung 4.20: Schematische Darstellung von gewebespezifischer Genregulation durch einen TF (blau) und die Folgen einer gewebeunspezifischen, orthologen Darstellung (unten), wie sie bspw. von TRANSPATH<sup>®</sup> verwandt wird. Dabei werden einem TF regulierte Gene unabhängig vom Gewebetyp zugeordnet.

Mit p53 und NF- $\kappa$ B werden zwei weitere synergistisch beeinflusste TFs detektiert, die verschiedenste Gene, darunter auch immunrelevante, regulieren (Ishida *et al.*, 1999; Fujioka *et al.*, 2004; Tripathi und Aggarwal, 2006; Salminen *et al.*, 2008; Farnebo, Bykov und Wiman, 2010; Gupta *et al.*, 2011). CREB, CBP und NF-AT sind relevant in MHC-II-Signalpfaden. Sie unterstützen die Immunantwort in B- und T-Zellen, dienen als positive Regulatoren für Cytokin-signale und verbinden angeborenes und adaptives Immunsystem (Monticelli und Rao, 2002; Vivier, 2004; Nakayama *et al.*, 2011). Da diese wenigen TFs fast alle angebotenen Gene regulieren (Tab. 4.3), reichte die funktionale Kontrolle der TFs (Tab. 4.4) jedoch nicht aus. Zusätzlich mussten die Funktionen der Gene betrachtet werden. Dabei war zu berücksichtigen, dass bei der Mappierung von TFs auf Gene mit Hilfe von Datenbanken die Gewebespezifität (siehe Abb. 4.20) verloren geht.



### Selektion pDC-spezifischer Gene durch Experimentaldaten

Die Expression von Genen ist abhängig vom Gewebe. Nicht in jedem Gewebe wird jedes Gen exprimiert, wodurch je nach Gewebe ein unterschiedlicher Satz an TFs exprimiert wird, welche unterschiedliche Gene regulieren. Zusätzlich kann die Struktur oder die Verpackung der DNA verändert sein, sodass TF-spezifische Bindestellen nicht mehr erkannt werden und der TF dadurch nicht an der DNA binden kann (Alberts *et al.*, 2008). In der Folge können die Gene nicht weiter abgelesen und exprimiert werden. Dadurch kann ein spezifischer TF, welcher in verschiedenen Geweben exprimiert wird, gewebeabhängig einen unterschiedlichen Satz an Genen regulieren (Abb. 4.20). Da viele Datenbanken, wie auch TRANSPATH<sup>®</sup> und TRANSFAC<sup>®</sup>, keine gewebespezifischen Informationen bereithalten, sind einige der als synergistisch reguliert vorhergesagten Gene falsch Positive und müssen vor der weiteren funktionalen Analyse entfernt werden. Um diese Gene zu filtern, werden die experimentell gemessenen Expressionsraten genutzt (siehe Abschnitt 3.4). Dabei wurden alle Gene entfernt, bei denen sich die Expressionsrate der synergistischen Bedingung (IL-3+Flu) nicht von der Kontrollbedingung (Medium) unterschied.

Um die Expressionsraten zu vergleichen, wurde die Differenz zwischen den beiden Expressionswerten gebildet. Da es sich bei diesen um logarithmierte Expressionswerte handelt, entspricht die Differenz dem *fold-change*, also dem Verhältnis der Änderung zwischen den beiden Bedingungen. Man nimmt an, dass Gene, welche stärker variieren als die Standardabweichung aller 329 Gene (gestrichelte Linien), einen Unterschied in ihren Expressionwerten besitzen und dieser auf die synergistische Signalintegration zurückgeht. Außerdem sieht man in Abb. 4.21, dass der Mittelwert nahe 0 liegt und viele vorhergesagte Gene ihre Genexpression nicht ändern<sup>58</sup>. Dies zeigt, dass viele Gene falsch-positiv vorhergesagt wurden. Von den insgesamt 329 vorhergesagten Genen konnten 298 (90, 58%) auf experimentelle Werte mappiert<sup>59</sup> werden. Davon waren 40 (13, 42%) Gene differentiell exprimiert. Die differentiell exprimierten Gene werden in Abb. 4.21 benannt. Die

---

<sup>58</sup>Ein zweiseitiger-ein-Stichproben-t-test gegen  $\mu = 0$  wird mit  $p > 0.05$  abgelehnt. Die Mehrheit der Gene ändert demnach im Vergleich von IL-3+Flu und Medium ihre Genexpression nicht. Deshalb ist ein Großteil der Gene falsch positiv identifiziert.

<sup>59</sup>Der eingesetzte Microarray deckt nicht alle möglichen Gene ab, sondern nur eine Auswahl, wodurch sich die Differenz der Vorhergesagten und der Mappeden erklärt.

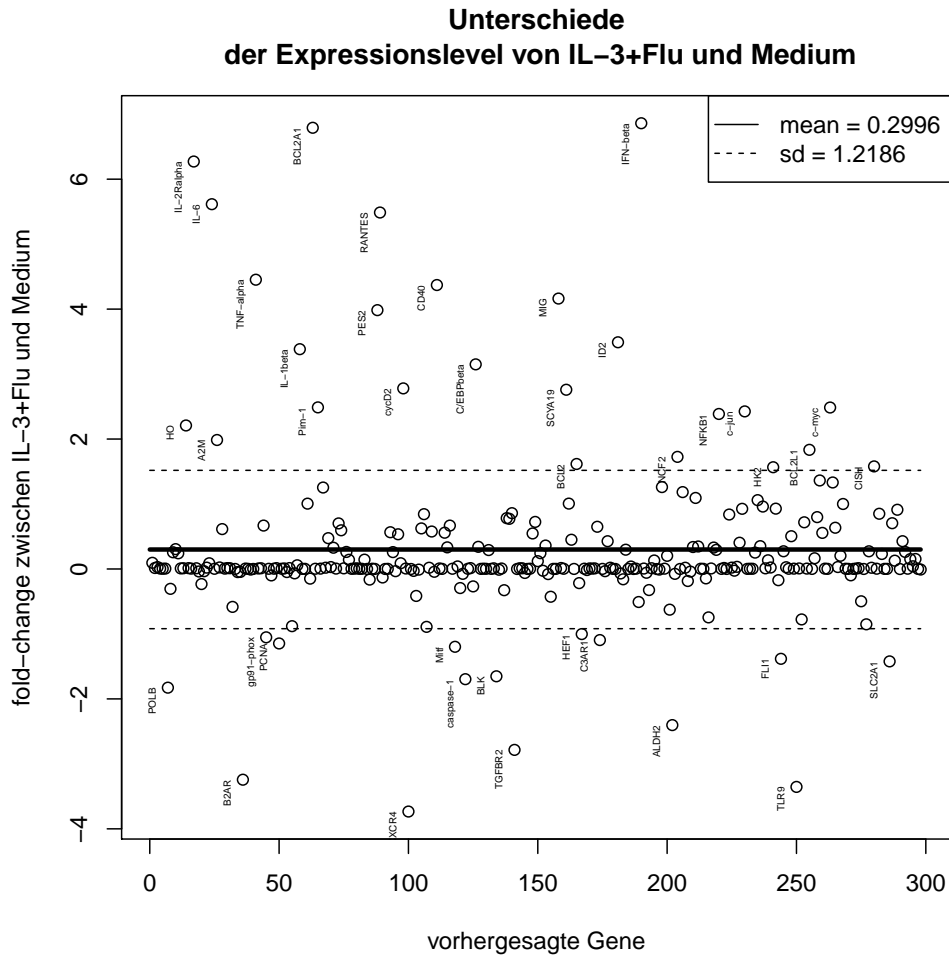


Abbildung 4.21: Änderung der Genexpression der algorithmisch bestimmten synergistischen Gene (*fold-change*) zwischen der IL-3+Flu-Bedingung und Medium (Kontrolle). Viele vorhergesagte Gene zeigen keine Änderung (eine Erklärung liefert Abb. 4.20). Gene, deren Expressionsraten sich im Vergleich der Bedingungen stärker als der Mittelwert (mean)  $\pm$  Standardabweichung (sd) verändern, sind benannt und vermitteln synergistische Effekte.

anderen Gene (258) zeigten keine ausreichende Änderung ihrer Expressionsraten und wurden nicht weiter betrachtet. Diese Gene wurden fälschlicherweise als synergistisch reguliert vorhergesagt. Der wesentliche Grund ist dabei die fehlende Gewebsspezifität bei der Bestimmung der durch einen TF regulierten Gene. Ein TF kann in verschiedenen Geweben einen unterschiedlichen Satz an Genen regulieren (siehe Abb. 4.20). Gewebe- und zelltypspezifische Informationen über die durch TFs regulierten Gene sind für eine präzise systemische Analyse von Signaltransduktion und den ausgelösten Effekten notwendig und sollten in Zukunft in den einschlägigen Datenbanken abgebildet werden. Für die 40 vorhergesagten und als differentiell exprimiert bestimmten Gene wird ein funktionaler Zusammenhang zwischen synergistischer Beeinflussung und Immunantwort angenommen.

#### **Funktionen synergistisch regulierter Gene**

Unter Verwendung der gleichen Methodik zur Funktionsbestimmung der TFs wurde auch den 40 Genen eine Funktion zugewiesen. Tabelle 4.5 zeigt dabei die Funktionen der jeweiligen Gene. Diese sind immunrelevant, wenn ihnen innerhalb der hier verwendeten KEGG-Datenbank (siehe Unterabschnitt 3.1.3) ein immun- oder stimularelevanter Term zugeordnet werden konnte, wobei typische und häufig auftretende Funktionen zahlenkodiert sind. Tabelle 4.5 ist absteigend nach dem absolutem *fold-change* sortiert, sodass sich die am stärksten ändernden Gene im oberen Teil der Tabelle wiederfinden. Unter den ersten 12 Genen befinden sich hauptsächlich jene, die direkte stimuli- und immunrelevante Funktionen vermitteln. Insgesamt wurde für 28 (70%) von 40 Genen Immunrelevanz (Spalte 3 in Tab. 4.5) festgestellt. Die in Tabelle 4.5 mit † markierten Gene kodieren für Proteine, die zu einer gemeinsamen Proteinfamilie (Bcl-2) gehören und deren Wirkung eng mit Immunantwort sowie der Kontrolle von Zellzyklus und Apoptose verbunden ist (Peso *et al.*, 1997; Khaled *et al.*, 2005; Youle und Strasser, 2008; Chipuk *et al.*, 2010). Durch Fehlregulation der letzten beiden Mechanismen kann Krebs ausgelöst werden (Khaled *et al.*, 2005; Youle und Strasser, 2008; Chipuk *et al.*, 2010). Auch 8 weitere Gene zeigten eine Verbindung zu Krebs und spiegeln damit die in der Literatur diskutierte Verbindung zwischen Immunsystem und Krebs wider (Wakefield und Roberts, 2002; Pikarsky *et al.*, 2004; Visser,

Genname	fold-change	immun-relevant	Funktion
IFN-beta	6.86157	+	1,2,3,5
BCL2A1 <sup>†</sup>	6.79177	+	hematopoietic BCL2-related protein A1
IL-2Ralpha	6.27229	+	1,2,4,5
IL-6	5.61522	+	1,2,3,4,5
RANTES	5.4872	+	1,2,3
TNF-alpha	4.45283	+	1,2,3,4
CD40	4.37116	+	1,2,3
MIG(CXCL9)	4.16285	+	1,2,3
PES2	3.98457	-	Metabolic pathways Pathways in cancer
CXCR4	-3.73327	+	1,2
ID2	3.48982	-	inhibitor of DNA binding 2B
IL-1beta	3.38358	+	1,2,3,4
TLR9	-3.35424	+	1,3
B2AR	-3.24327	-	Calcium signaling pathway Neuroactive ligand-receptor interaction Endocytosis
C/EBPbeta	3.14969	-	CCAAT/enhancer binding protein Transcriptional misregulation in cancers
TGFBR2	-2.78458	+	2, MAPK signaling pathway
cycD2	2.7798	+	5,p53 signaling pathway Transcriptional misregulation in cancer Cell Communication, Growth and Death
SCYA19	2.7594	+	1,2
Pim-1	2.48773	+	5, Chemokine signaling pathway
c-myc	2.48549	+	5, MAPK signaling pathway Transcriptional misregulation in cancer
c-jun	2.42587	+	1,3
ALDH2	-2.40282	-	Fatty acid metabolism Metabolic pathways
NFKB1	2.38638	+	1,3, B-,T-cell receptor signaling
HO(HMOX1)	2.20932	-	Metabolism of Cofactors and Vitamins Mineral absorption
A2M	1.98567	+	1, Complement and coagulation cascades
BCL2L1 <sup>†</sup>	1.83452	+	5,Transcriptional misregulation in cancer Apoptosis
POLB	-1.82654	-	Base excision repair Replication and Repair
NCF2	1.72607	+	1,Leukocyte transendothelial migration Phagosome
caspase-1	-1.69598	+	1,Cytosolic DNA-sensing pathway Influenza A
BLK	-1.65168	+	B lymphoid tyrosine kinase
BCL2 <sup>†</sup>	1.61623	+	Neurodegenerative Diseases Pathways in cancer Cell Growth and Death
CISH	1.57914	+	5
HK2	1.56635	-	Metabolic pathways Digestive Syste
SLC2A1	-1.42217	+	Adipocytokine signaling pathway Pathways in cancer
FLI1	-1.38247	-	Transcriptional misregulation in cancers
Mitf	-1.19748	-	Pathways in cancer
PCNA	-1.14626	-	Replication and Repair Base excision repair
C3AR1	-1.09384	+	1,Complement and coagulation cascades Neuroactive ligand-receptor interaction
gp91-phox	-1.05189	+	1, Leukocyte transendothelial migration Phagosome
HEF1	-1.00187	-	neural precursor cell expressed

Tabelle 4.5: Repräsentation der 40 relevanten und vorhergesagten synergistisch regulierten Gene und ihrer Funktionen in den einzelnen Signalpfaden. Durch 1=Immune System, 2=Cytokine-cytokine receptor interaction, 3=Toll-like receptor signaling pathway, 4=Hematopoietic cell lineage, 5= Jak-STAT signaling, werden wichtige Terme, die eine Rolle in der Immunantwort implizieren, kodiert. Andere Funktionen werden textuell dargestellt. Die erste Spalte benennt die Gene, die zweite den *fold-change*, die dritte zeigt die Relevanz in der Immunantwort, während die letzte Spalte die mittels Datenbanken (KEGG, siehe 3.1.3 und 4.1.1) bestimmte Funktion darstellt. <sup>†</sup> zeigt Vertreter der Bcl-2 Protein Familie an.

Eichten und Coussens, 2006; Lin und Karin, 2007; Liu *et al.*, 2011). Synergien können daher bei Prozessen, die Fehlregulationen auslösen und zur Bildung von Krebszellen beitragen, eine Rolle einnehmen.

Überraschend war das Auffinden von Funktionen, welche mit DNA-Reparatur (PCNA, POLB) oder dem Nervensystem assoziiert sind (C3AR1, BCL2, B2AR), da eine Verbindung zur Immunantwort nicht offensichtlich ist. Gasser *et al.* beschreiben die Fähigkeit, Mithilfe von DNA-Reparaturmechanismen einen Zellzyklusarrest, also die Unterbrechung der weiteren Zellteilung, und folglich Apoptose auszulösen (Gasser *et al.*, 2005). Mittels dieses Mechanismus werden schadhafte Zellen markiert, sodass sie vom Immunsystem erkannt und entfernt werden (Gasser *et al.*, 2005). Auch hier kann eine Verbindung zu Krebs gezogen werden. So ist bekannt, dass Krebszellen einer Erkennung durch das Immunsystem ausweichen (Villunger und Strasser, 1999; Garcia-Lora, Algarra und Garrido, 2003; Huang *et al.*, 2005a; Malmberg und Ljunggren, 2006). Wenn nun synergistische Effekte in einer Zelle die Markierung der Zelle als schadhaft verhindern, hilft dies krankhaften Zellen dabei, einer Detektion durch Immunzellen zu entgehen. Es zeigt sich, dass Synergien, ihr Einfluss auf das Immunsystem sowie ihr Einfluss auf Krebs, experimentell stärker untersucht werden müssen.

Die detektierten neuroassoziierten Gene sind interessant, da sie bei der Aufklärung der Beziehung zwischen Immunsystem und neurodegenerativen Erkrankungen wie Alzheimer und Schizophrenie helfen (Chen und Kendler, 2008; Cohen, 2009; Zambrano *et al.*, 2010; Swardfager *et al.*, 2010). In der Literatur wurde zum Beispiel gezeigt, dass IL-3 einen protektiven Effekt auf Neuronen ausübt, wodurch der Ausbruch von Alzheimer verzögert wird (Zambrano *et al.*, 2007; Zambrano *et al.*, 2010). Offen bleibt hierbei die Frage, welchen Einfluss Flu als zweite synergistische Komponente auf diese Krankheiten ausübt. Eventuell tritt Flu als unspezifischer Verstärker für das durch IL-3 induzierte Signal auf. In diesem Fall würde Flu, wie andere immunrelevante Stimuli auch, eine Überaktivität des Immunsystems induzieren, wodurch Immunreaktionen unspezifisch verstärkt würden. Dies ist auch konsistent mit der beobachteten Dominanz von Flu bei der Immunantwort (siehe Unterabschnitt 4.1.1 und Abb. 4.2) und zeigt, dass die IL-3-Signalfade oft downstream von Flu-Signalfaden liegen.

Insgesamt ließen sich die beobachteten und detektierten Gene in zwei Gruppen unterteilen: jene, die eine direkte Immunantwort auslösen und jene die einen mediativen Einfluss besitzen. Die direkte Immunantwort induziert eine unspezifische Reaktionen, ausgelöst durch das angeborene Immunsystem, wogegen der mediative Einfluss spezifischer Natur ist und das erworbene (adaptive) Immunsystem anspricht. Betrachtet man die in Tab. 4.5 beschriebenen Funktionen, so findet sich die Zweiteilung in den Funktionen, wie auch in den damit assoziierten Krankheiten, wieder.

Direkte Einflüsse sind mit essentiellen zellulären Prozessen wie Zellzyklus, Zellwachstums und Apoptose verbunden. Diese Prozesse greifen in grundlegende Mechanismen der Zelle ein und werden unspezifisch ausgelöst, sobald ein immunrelevantes Signal perzipiert wird (Janeway *et al.*, 2001). Ziel ist es, eine Infektion<sup>60</sup> so schnell wie möglich auszuschalten (Janeway *et al.*, 2001). Durch Auslösung des programmierten Zelltods (Apoptose) wird ein Übergreifen der Infektion auf den Organismus verhindert. Dadurch bedingen sich auch die durch Fehlregulationen assoziierten Krankheiten. So sind Prozesse sowie Gene, die dem angeborenen Immunsystem zuzuordnen sind, gleichzeitig mit der Entstehung oder der Verbreitung von Krebs verbunden (Locksley, Killeen und Lenardo, 2001; Gaur und Aggarwal, 2003; Huang *et al.*, 2005*a*; Visser, Eichten und Coussens, 2006; Youle und Strasser, 2008). Die gesteigerte Expression von Genen wie TNF-alpha, cycD2 oder A2M (siehe Tab. 4.5) ist relevant für das Funktionieren der angeborenen Immunantwort sowie für die Entstehung und Proliferation von Tumoren.

Die indirekten Einflüsse in pDC werden zwar vom angeborenen Immunsystem angestoßen, entfalten jedoch nicht sofort ihre Wirkung. Exprimierte Gene wie IL-2Ralpha, IL-1beta, TLR9<sup>61</sup> und IL-6 (siehe Tab. 4.5) transferieren das Signal auf das adaptive Immunsystem, wo eine auf den Stimulus zugeschnittene Immunantwort erzeugt wird (Janeway *et al.*, 2001). Die aktivierten Gene, Interleukine sowie Cytokine dienen als Signalgeber zur Induktion der weiteren Immunantwort. So werden über die freigesetzten Signale weitere Immunzellen, die für die spezifi-

---

<sup>60</sup>Infektion meint in diesem Zusammenhang alle Prozesse, bei denen ein externes Signal als organismische Bedrohung wahrgenommen wird.

<sup>61</sup>TLR9 aktiviert auch weitere Signalpfade der angeborenen Immunantwort und erhöht die Perzeptionsfähigkeit gegenüber Immunsignalen durch Aktivierung von pDC und B-Zellen (Akira, Takeda und Kaisho, 2001; Takeda und Akira, 2004; Viau und Zouali, 2005; Lenert, 2006; Uematsu und Akira, 2007).

sche Handhabung der Stimuli sorgen, rekrutiert (Chan *et al.*, 2005; Lenert, 2006; Lan, Selmi und Gershwin, 2008; Pirog, Stabinsky und Goldman, 2010). Außerdem entfalten Signale wie IL-6, IL-2 und RANTES inflammatorische Wirkungen, wodurch sie den Ort einer Infektion anzeigen. Diese verschiedenen Wirkweisen: die Rekrutierung von Immunzellen, die Ausschüttung inflammatorischer Signale, sowie die Möglichkeit Apoptose auszulösen, unterstreichen die Vermittlungsfunktion von pDC zwischen angeborenem und erworbenem Immunsystem (Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007; Gilliet, Cao und Liu, 2008; Kool *et al.*, 2009). Fehlregulationen bei der Übertragung auf die adaptive Immunantwort oder in der inflammatorischen Signalverarbeitung können zu Krankheiten wie Alzheimer, Schizophrenie oder allergischen Erkrankungen führen (Chen und Kendler, 2008; Cohen, 2009; Kool *et al.*, 2009; Zambrano *et al.*, 2010). Für diese Erkrankungen ist eine Beziehung zum Immunsystem in der Literatur belegt. Die detektierte Beziehung zu weiteren neurodegenerativen Erkrankungen, wie für BCL-2 angenommen (siehe Tab. 4.5), wird vermutlich ebenfalls durch die inflammatorischen Effekte, welche generell das Immunsystem in einen Alarmstatus versetzen und angeborenes sowie adaptives Immunsystem aktivieren, ausgelöst.

Weitere detektierte und nicht direkt mit dem Immunsystem verbundene Mechanismen betreffen den Fettsäurestoffwechsel und andere metabolische Signalpfade, die durch Gene wie PES2, ALDH2 oder HK2 aktiviert werden. Diese Gene sprechen generelle Funktionen der Zelle an und dienen der Ausdifferenzierung von pDC zu DC (siehe Unterabschnitt 2.2.2).

#### **pDCs realisieren den Wechsel von angeborenem zu adaptivem Immunsystem**

Anhand der bisherigen Ergebnisse zeigt sich schematisch, wie sich pDCs im Laufe von 24h in DC differenzieren (siehe Abb. 4.22). IL-3 und Flu wirken auf pDC ein, wobei sich die Funktion der Zelle über die Zeit ändert. Am Anfang wird das angeborene Immunsystem angesprochen und im Laufe der Ausdifferenzierung wird auf das adaptive Immunsystem gewechselt. Damit dienen pDC maßgeblich als Mediatoren zwischen angeborenem und adaptivem Immunsystem (Colonna, Trinchieri und Liu, 2004). Die algorithmisch detektierten Gene und ihre Funktio-

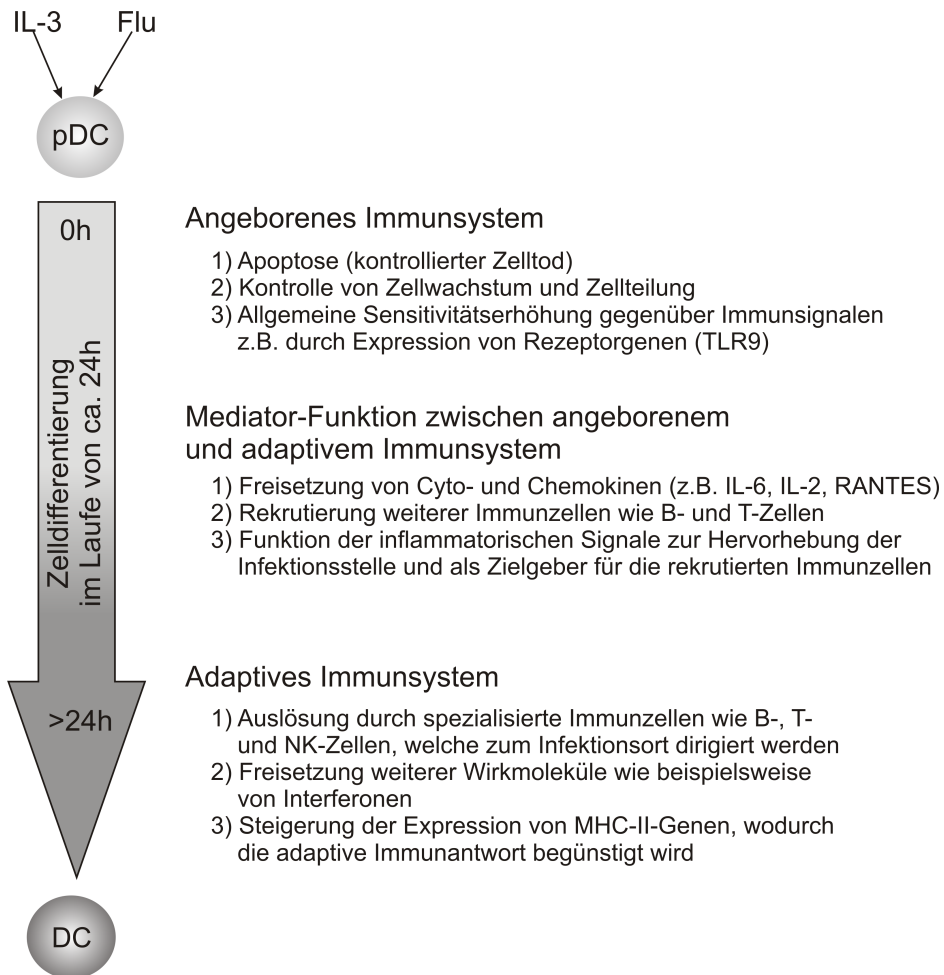


Abbildung 4.22: Schema der aktivierten Immunantwort. Während 24h (durch Pfeil symbolisiert) differenzieren sich pDC zu DC. Zuerst wird das Angeborene, später das adaptive Immunsystem angesprochen.



nen zeigen dies, da Gene aus beiden Gruppen detektiert werden (siehe Tab. 4.5). Gene, die das angeborene Immunsystem beeinflussen, sind mit regulatorischen Mechanismen des Zellzyklus verbunden und beeinflussen Zellwachstum und Zellteilung (Huang *et al.*, 2005a; Cao und Liu, 2007; Lin und Karin, 2007). Gene des adaptiven Immunsystems wirken auf das inflammatorische Verhalten der Zelle und rekrutieren weitere Immunzellen zum Infektionsort. Während sich pDCs ausdifferenzieren, verändert sich die Ansteuerung der beiden Immunsystemkomponenten kontinuierlich (Liu, 2005). Je geringer der Einfluss auf das angeborene Immunsystem wird, desto größer wird er auf das Adaptive. Daraus ergibt sich, dass kein direkter Schalter existiert, sondern dass durch die angestoßenen Prozesse ein fließender Übergang gewährleistet wird. Durch das kontinuierliche Umschalten wird eine gleichbleibend hohe Effizienz der Immunantwort gewährleistet, denn bekanntermaßen benötigt das adaptive Immunsystem längere Zeit um aktiviert zu werden (Janeway *et al.*, 2001). Eine Abschwächung der Immunreaktion würde *Druck* vom Infektionsherd nehmen und zu einer Ausbreitung oder Manifestierung der Infektion führen, was kontraproduktiv ist. Auch die in Tab. 4.5 nicht immunrelevanten Gene ordnen sich in dieses Schema ein, da sie Rollen in der Ausdifferenzierung der Zelle oder in anderen metabolischen Pfaden übernehmen.

#### 4.3.7 Zusammenfassung

Synergie zeichnet sich durch die Existenz eines integrierenden, von allen Stimuli gemeinsam erreichbaren Prozesses, sowie der Exklusivität der beeinflussten Produktknoten aus. Sie ist mit bisherigen topologischen Methoden nicht zu detektieren. Das vorgestellte algorithmische Verfahren detektiert synergistische Effekte als topologische Eigenschaft in bipartiten Netzwerken. Synergistische Prozesse und ihre Moleküle können aufeinander aufbauen. Die Darstellung der gegenseitigen Abhängigkeiten hilft dabei, wichtige Moleküle zu erkennen (siehe Unterabschnitt 4.3.3). Es wurden Regulatoren detektiert, die viele andere synergistische Moleküle beeinflussen und eine Schlüsselrolle in der Signalweiterleitung und der Signalmodulation einnehmen. Durch TFs wird das Signal von der Proteomischen auf die genomische Ebene transferiert (Unterabschnitt 4.3.4). Die beeinflussten Gene vermitteln die signalspezifischen synergistischen Effekte und beeinflussen das Verhalten der Zelle. Die algorithmisch vorhergesagten Gene sind die Signalempfänger der eingesetzten Stimuli.

Unter Verwendung des vorgestellten Algorithmus 4.3.1-4.3.5 wurde die durch IL-3 und Flu induzierte Immunkaskade bezüglich Synergien und Genregulation untersucht. Dabei wurde festgestellt, dass sich die synergistisch beeinflussten TFs gegenseitig regulieren. Innerhalb der detektierten TFs wurde eine Gruppe von 71 TFs, welche als Hauptregulatoren dienen (siehe Abb. 4.18), identifiziert. Mehr als 90% der beeinflussten Gene werden von nur 18 TFs der Hauptregulationsebene kontrolliert. Diesen TFs wurde eine Rolle in der Immunantwort zugeordnet (siehe Tab. 4.4). Detektierte, komplexierte TFs transferieren das Signal im Allgemeinen nicht auf die genomische Ebene (siehe Tab. 4.3), spielen aber bei signalregulatorischen Prozessen eine wichtige Rolle und regulieren beispielsweise zeit- und konzentrationsabhängige Prozesse.

Von den 40 algorithmisch vorhergesagten synergistisch regulierten Genen besitzen 28 (70%) Gene eine immunrelevante Funktion (siehe Tab. 4.5). Es wurde festgestellt, dass diese Gene in zwei Gruppen zerfallen. Die erste ist mit dem angeborenen und die zweite mit dem adaptiven Immunsystem verbunden. Dies unterstreicht die Verbindung mit den beiden Immunsystemteilen und der besonderen Rolle von pDC in mammalischen Organismen. Anfangs aktivieren pDCs Prozesse der angeborenen Immunantwort. Diese beeinflussen grundlegende Zellmechanismen wie beispielsweise Apoptose (Gaur und Aggarwal, 2003; Cao und Liu, 2007; Youle und Strasser, 2008). Später treten adaptive Effekte, wie die Rekrutierung weiterer Immunzellen (Lin und Karin, 2007), in den Vordergrund (siehe Abb. 4.22). Während das angeborene Immunsystem auf einfache und universelle Mechanismen zurückgreift, um eine Infektion zu stoppen (z.B. Apoptose), greift das Adaptive auf langsamere aber spezifischere Prozesse zurück (Rekrutierung von Immunzellen usw.). Die durch den Algorithmus identifizierten Gene zeigen dies und realisieren einen kontinuierlichen Schalter zwischen angeborenem und adaptivem Immunsystem. Sie sollten ebenso wie die signalregulatorischen komplexierten TFs genauer untersucht werden.

## 4.4 Vergleich zwischen experimenteller und algorithmischer Methode

### 4.4.1 Methodischer Vergleich der Ansätze zur Detektion von Synergien

In Tab. 4.6 werden die Methoden und ihre Ergebnisse dargestellt und miteinander verglichen. Dabei werden Vor- und Nachteile, Ziele und Möglichkeiten der in Abschnitt 4.1 und 4.3 vorgestellten Methoden herausgearbeitet.

Der experimentelle Ansatz (Abschnitt 4.1) basiert auf der Messung von Genexpressionsdaten mittels Microarrays, wobei synergistisch regulierte Gene anhand des Vergleichs aller kombinatorisch möglichen Bedingungen bestimmt werden. Gerade bei vielen eingesetzten Stimuli ergeben sich aus der Anzahl der notwendigen Experimente hohe Kosten zur Bestimmung von Synergien, obwohl nur globale Effekte in den Genexpressionsdaten zu beobachten sind. Die proteomische Ebene, das heißt Signaltransduktion, Signalintegration sowie der Signaltransfer auf die genomische Ebene mittels TFs, werden bei dieser Art der Untersuchung nicht betrachtet (siehe Tab. 4.6). Die breitere experimentelle Basis identifiziert jedoch auch andere wechselseitige Wirkungen der Stimuli, wie beispielsweise Dominanzbeziehungen (siehe Abb. 4.23). Diese können über synergistische Komponenten verfügen, entsprechen jedoch nicht der Definition von Synergien, da echte Synergien alternativlos sind (siehe Abschnitt 4.3 und Abb. 4.23). Die Vernachlässigung der Signaltransduktionswege auf proteomischer Ebene erschwert das Auffinden von signalabhängigen Synergien und resultiert in der Detektion weniger immunrelevanter Gene (siehe Tab. 4.6). Der experimentelle Ansatz eignet sich besonders zur Erstaufklärung und Klassifikation unbekannter Sachverhalte, da er ein erstes, einfaches und globales Grundverständnis der möglichen Zusammenhänge (zum Beispiel Stimulus-Effekt-Beziehungen) vermittelt. Jedoch liefert er keine Möglichkeit, die ablaufenden Signaltransduktions- und Integrationsprozesse zu beobachten oder zu erklären, wodurch die Einordnung der Effekte in die komplexen Lebensumstände von Zellen und Organismen schwer möglich ist.

	<b>Experiment</b>	<b>Algorithmus</b>
notwendige Vorarbeiten	keine	bipartites Netzwerk, Informationen über Signaltransduktion (z.B. aus Datenbanken)
Detektion von Synergien	durch kreuzweisen Vergleich aller Bedingungen (teuer und aufwändig)	als topologisches Muster
detektierte synergistisch regulierte Gene	73	40
Beobachtung von Signaltransduktionsprozessen	nicht möglich	in jedem Schritt möglich
Sichtweise	global	global und lokal
Experimente zur Verifikation der detektierten Synergien	keine, aber Detektion ist aufwändig (siehe oben)	nur Vergleich zwischen synergistischer Bedingung und Kontrolle notwendig, daher geringere Kosten
Funktionale Verifikation	Mittels Datenbanken und funktionaler Annotation der regulierten Gene	Mittels Datenbanken und funktionaler Annotation der regulierten Gene
detektierte immunrelevante Gene	22 (30, 137%)	28 (70%)
Anwendungsmöglichkeiten	Erstauflklärung möglicher regulierter Gene; Klassifikation der Genexpression einer Zelle oder eines Gewebes; erkennen globaler Zusammenhänge auf Genexpressionsebene; einfaches Grundverständnis	Präzises Verständnis und Erklärung der ablaufenden Prozesse; Einordnung in den globalen Kontext durch Betrachtung der Anbindung zu anderen Prozessen; Interaktionen zwischen verschiedenen Komponenten;
Stärken	Dominanzbeziehungen zwischen einzelnen Stimuli werden sichtbar; unabhängig von Voraussetzungen oder Datenbankannotationen	Signalmodulation beobachtbar; Signalintegration ist in Prozessen beobachtbar; Vorhersage von Ziel-TFs, sowie Zielgenen; synergistische Signalwege werden erklärbar; jede Ebene kann einzeln verifiziert werden
Schwächen	Keine Prozesse beobachtbar; nur abschließende Effekte (Gene) werden beobachtet; identifiziert durch fehlenden Fokus auf stimulusspezifische Signaltransduktion sehr globale Effekte; kombinatorische Vielfalt führt zu schwieriger Bestimmung von Synergien; vernachlässigt die proteomische Ebene	Abhängig von der Qualität der Daten in den Datenbanken und ihren Annotationen

Tabelle 4.6: Vergleich zwischen experimenteller und algorithmischer Methode zur Bestimmung von Synergien und synergistisch regulierten Genen. In der ersten Spalte werden verschiedene Parameter, die zwischen Experiment (Spalte 2) und Algorithmus (Spalte 3) verglichen werden, dargestellt.

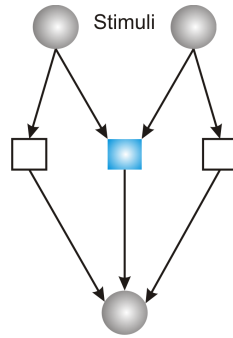


Abbildung 4.23: Dominanzeffekte und echte Synergien schließen sich gegenseitig aus. Rechtecke zeigen Signaltransduktionsprozesse an, wobei in Blau synergistische Prozesse dargestellt werden. Die ungefüllten Rechtecke zeigen Alternativwege von den Stimuli, wodurch Dominanzeffekte ausgelöst werden können. Der Zielknoten verfügt über eine synergistische Beeinflussung, welche nicht exklusiv ist (siehe Abschnitt 4.3).

Das algorithmische Verfahren (Abschnitt 4.3) hingegen nutzt ein Netzwerk, welches das aus Datenbanken extrahierte und experimentell verifizierte Wissen um Signaltransduktion, Prozessierung und Genregulation abbildet (siehe 4.2). Dadurch sind synergistische Prozesse auf proteomischer und genomischer Ebene lokal überprüfbar. Synergien werden dabei als topologisches Muster im Netzwerk detektiert (siehe 4.3.1-4.3.5), wobei die Methode abhängig von der Qualität der in den Datenbanken hinterlegten Daten ist. Fehlerhafte oder fehlende Annotationen in den Datenbanken, beispielsweise bei Funktionen oder Gewebespezifität (siehe Abb. 4.20 und Abb. 4.21), beeinflussen sowohl die Detektion der synergistischen Prozesse, wie auch die Vorhersagen der synergistisch regulierten Gene. Durch Experimentaldaten wurden putativ falsch-positiv identifizierte Synergien aus der Vorhersage entfernt (siehe Abb. 4.21 in Unterabschnitt 4.3.6). Durch die Detektion topologischer Muster, also einer strukturellen Eigenschaft des Netzwerkgraphen, konnte der experimentelle Aufwand auf den Vergleich zwischen synergistischer Bedingung und Kontrolle beschränkt werden (siehe Abschnitt 4.3). Synergien treten nämlich nur bei integrativen Prozessen (topologisch detektiert) auf und werden auf die Produkte dieser projiziert, wo der induzierte synergistische Effekt gemessen werden kann. Neben den synergistisch beeinflussten Genen werden proteomische Prozesse, in denen die aktivierten Signalfade synergistisch miteinander interagieren (lokale Sichtweise), beobachtet und nach-

verfolgt. Dadurch wurden weitergehende Fragestellungen, wie beispielsweise die gegenseitige Regulation von TFs auf der proteomischen Ebene (siehe Unterabschnitt 4.3.6), en détail analysiert. Die algorithmische Methode hilft dabei, die komplexe Biologie der Zelle besser zu verstehen und zu erklären. Die durch die Stimuli ausgelösten Signalwege sind vom Rezeptor bis zu den regulierten Genen und den ausgelösten Effekten verfolgbar.

#### 4.4.2 Vergleich der Immunrelevanz der experimentell und algorithmisch bestimmten Gene

Um die Ergebnisse der beiden Methoden zu evaluieren wurden die gemeinsam identifizierten Gene bestimmt, wobei insgesamt nur 6 Gene von beiden Methoden detektiert wurden. Diese kleine Menge war jedoch nicht überraschend, da beide Methoden auf sehr unterschiedlichen Wegen die synergistisch regulierten Gene bestimmen. Auffällig war, dass die identifizierten Gene oft in Relation zueinander stehen. Beispielsweise wurde algorithmisch IL-1 $\beta$  identifiziert (siehe Tab. 4.5), wogegen der zugehörige IL-1-Rezeptor (IL-1R1) experimentell identifiziert wurde (siehe Tab. 4.1). Durch beide Methoden wurden noch weitere Cytokin-Rezeptor- und Cytokin-kodierende Gene, wie beispielsweise IL7R oder IL-6 und TNF- $\alpha$ , bestimmt, die ebenfalls eine Rolle in der Immunantwort einnehmen (Locksley, Killeen und Lenardo, 2001; Chan *et al.*, 2005; Pirog, Stabinsky und Goldman, 2010). Um die Qualität der Methoden bezüglich der induzierten Immunantwort insgesamt zu vergleichen, wurden die detektierten Gene anhand ihrer molekularen Funktion charakterisiert und gruppiert<sup>62</sup> (siehe Tabellen 4.1 und 4.5). Dadurch konnten prinzipielle Funktionen überprüft werden, da gleichartig wirkende Gene zusammen betrachtet wurden. Dies hilft bei der Erlangung eines generellen systembiologischen Verständnisses der ausgelösten Immunantwort und evaluiert die experimentelle und algorithmische Methodik.

In Tabelle 4.6 sieht man, dass das algorithmische Verfahren im Vergleich zur experimentellen Methode zwar nur halb so viele Gene detektiert, diese jedoch spezifischer für die durch IL-3 und Flu ausgelöste Immunantwort sind. Der Anteil der algorithmisch verifizierten immunrelevanten Synergien übertrifft dabei

---

<sup>62</sup>Für die experimentell bestimmten Gene in Tab. 5.1 wurde die Funktion ebenso wie für die Gene in Tab. 4.1 bestimmt. Die Funktionen in Tab. 5.1 werden jedoch nicht abgebildet.

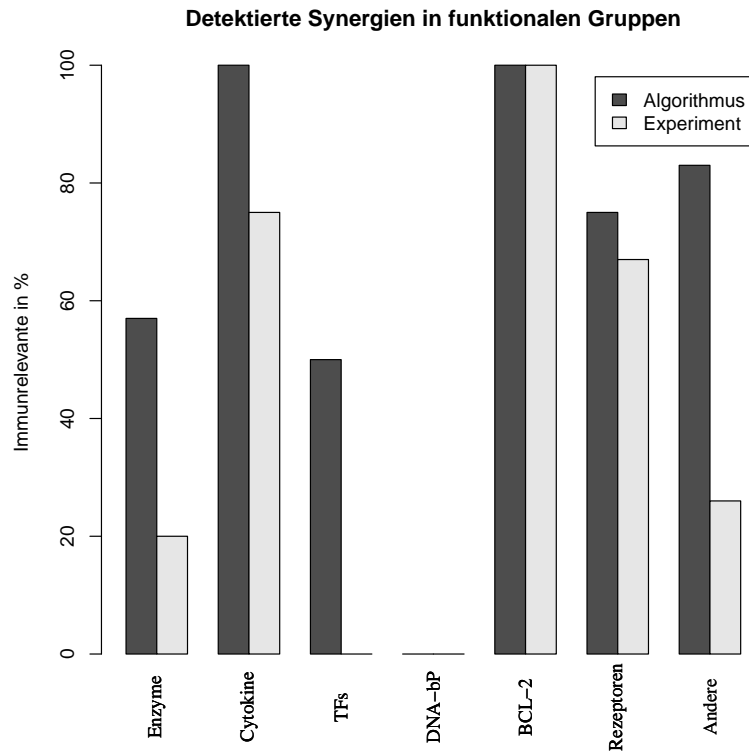


Abbildung 4.24: Prozentuale Verteilung der detektierten, verifizierten und funktional gruppierten, synergistisch induzierten Gene. Der Algorithmus detektiert Synergien in allen Gruppen besser oder genauso gut wie die experimentelle Methode.

Gruppe	detektiert		immunrelevant in %	
	Algorithmus (40)	Experiment (73)	Algorithmus	Experiment
Enzyme	7	20	57%	20%
Cytokine	11	8	100%	75%
TFs	6	13	50%	0%
DNA-bP	3	3	0%	0%
BCL-2	3	1	100%	100%
Rezeptoren	4	9	75%	67%
Andere	6	19	83%	26%

Tabelle 4.7: Ergebnisvergleich zwischen algorithmischer und experimenteller Methode. Darstellung wie viele Gene mit einer spezifischen Funktion durch die algorithmische (Spalte 2) und die experimentelle (Spalte 3) Methodik bestimmt wurden. Die Anzahl der verifizierten synergistischen und immunrelevanten Gene in % wird in den Spalten 4 (Algorithmus) und 5 (Experiment) illustriert.

immer den der experimentell Bestimmten (siehe Abb. 4.24). TFs und DNA-bindende Proteine<sup>63</sup> (DNA-bP) werden sowohl algorithmisch als auch experimentell schlecht detektiert. Während 50% der algorithmisch identifizierten TFs immunrelevant waren, konnten keine experimentell bestimmten TFs verifiziert werden. DNA-bP wurden von beiden Verfahren nie als immunrelevant gewertet (siehe Tab. 4.7). In Abgrenzung zu TFs regulieren DNA-bP hier nicht die Expression eines Gens direkt. Die in dieser Arbeit detektierten DNA-bP, wie POLB und PCNA (Tab. 4.5), sind beispielsweise in Mechanismen der DNA-Reparatur involviert (Matsumoto, Kim und Bogenhagen, 1994; Matsumoto und Kim, 1995; Sobol *et al.*, 1996; Frosina *et al.*, 1996; Sancar *et al.*, 2004). In allen anderen Gruppen detektiert das algorithmische Verfahren immunrelevante Synergien in mehr als der Hälfte der Fälle (siehe Abb. 4.24). Die grundsätzlich bessere Detektion von Synergien durch das algorithmische Verfahren geht auf die lokale Sichtweise des Algorithmus und die Detektion der integrierenden Prozesse sowie deren Projektion auf ihre Ziele zurück.

In zirka 70% der Fälle (siehe Tab. 4.6) detektiert das experimentelle Verfahren ein unspezifisches, nicht mit der Immunantwort korreliertes Gen. Gerade Enzyme und andere Moleküle können aufgrund ihrer oft mannigfaltigen Wirkung nur schlecht einer Immunrelevanz zugeordnet werden. Viele dieser Gene sind üblicherweise mit allgemeinen oder Differenzierungsaufgaben der Zelle betraut (siehe Unterabschnitt 4.1.2, Tab. 4.1). Das algorithmische Verfahren zeigt gerade in diesen eher allgemeinen Gruppen, dass es in der Lage ist, immunrelevante, synergistische Gene mit einer höheren Spezifität zu detektieren (siehe Abb. 4.24). Die Gruppen Cytokine, BCL-2 und Rezeptoren werden von beiden Verfahren gut detektiert (Abb. 4.24). Cytokine dienen als typische Immunsignale, die ins extrazelluläre Medium sekretiert werden und Signale an andere Zellen propagieren, wodurch beispielsweise zusätzliche Immunzellen aktiviert werden (Janeway *et al.*, 2001). Die Expression von BCL-2- und Rezeptor-Genen<sup>64</sup> erhöht die Sensitivität gegenüber weiteren Immunsignalen und ist ein wesentlicher Bestandteil von an-

---

<sup>63</sup>Diese Gruppe fasst in Abgrenzung zu TFs Proteine zusammen, die direkt an DNA binden, selbst aber keine TFs sind (z.B. Histone). Sie beeinflussen die Struktur und die *Verpackung* der DNA und ermöglichen dadurch beispielsweise das Binden von TFs oder epigenetische Effekte (Jones und Takai, 2001; Fuks, 2005).

<sup>64</sup>Exprimierte Rezeptorgene werden in Rezeptorproteine übersetzt, erst diese Proteine erhöhen die Spezifität gegenüber den Immunsignalen. Die vorhandene Menge an Rezeptoren ist jedoch unter anderem vom Expressionslevel der Gene abhängig.



geborener und adaptiver Immunantwort (Janeway *et al.*, 2001; Seth, Sun und Chen, 2006; Lin und Karin, 2007).

Die große Menge experimentell bestimmter, nicht immunkorrelierter Gene, wie zum Beispiel in der Gruppe der TFs und der DNA-bP, geht auf die duale Funktionalität von pDC zurück. Durch pDC wird das angeborene Immunsystem angesprochen, wobei gleichzeitig die Differenzierung zu DC ausgelöst wird, was zu einer Aktivierung des adaptiven Immunsystems führt (Cao und Liu, 2007). Die Aktivierung der für die Umwandlung zu DC notwendigen Signalpfade muss durch die ursprünglichen Stimuli induziert werden und wirkt unter anderem stark auf metabolische Pfade ein, da die Zelle einem komplettem Umbau unterzogen wird (Comeau *et al.*, 2002; Colonna, Trinchieri und Liu, 2004; Liu, 2005). Diese Pfade entfalten zwar genregulatorische Wirkungen, sind aber nicht unbedingt immunkorreliert. Das experimentelle Verfahren detektiert die Unterschiede der Genexpression, betrachtet jedoch nicht deren Ursache. Alle Gene werden gleichberechtigt behandelt, wodurch sich die detektierten und nicht immunrelevanten Gene erklären.

Im Gegensatz dazu ist die algorithmische Methode auf die stimuliinduzierte Immunantwort konzentriert, wobei neben immunrelevanten auch metabolische Funktionalitäten detektiert werden (siehe Tab. 4.5). Durch das Netzwerk (siehe Abschnitt 4.2) werden jedoch nur stimuliabhängige Auslöser metabolischer Funktionen identifiziert. Weitere nicht synergistische oder unabhängige Signalpfade werden nicht betrachtet.

Immunrelevante Langzeiteffekte wie die Verbindung zum adaptiven Immunsystem treten erst spät und bei fast vollständig abgeschlossener Differenzierung zu DC auf (Liu, 2005; Cao und Liu, 2007). Gene, die für Cytokine, Interferone oder Bestandteile von MHC-II-Komplexen kodieren, repräsentieren diese Verbindung (siehe Tab. 4.5). Während das experimentelle Verfahren immer nur punktuelle Aufnahmen der Expression von Genen zu einem bestimmten Zeitpunkt betrachtet (siehe Abschnitt 3.4), ignoriert das algorithmische Verfahren diese Abhängigkeiten. Dadurch müssen die vorhergesagten Gene zwar zeitlich eingeordnet und gegebenenfalls verifiziert werden (siehe Tab. 4.6), aber die Ergebnisse sind spezifischer und ermöglichen ein präzises Verständnis der ablaufenden Mechanismen und Prozesse (siehe Tab. 4.7).

Die eingeschränkte zeitliche Sichtweise erklärt auch, wieso Interferone experimentell nicht detektiert wurden, obwohl bekannt ist, dass sie in hohen Konzentrationen gegen Ende der Differenzierung zu DC ausgeschüttet werden (Liu, 2005; Cao und Liu, 2007). Die Expression von Interferonen war zum Betrachtungszeitpunkt nicht stark genug gesteigert, wodurch sich experimentell keine Synergie der beiden Immunsignale ableiten ließ. Gleiches gilt für Gene, welche in MHC-II-Pfaden involviert sind. Sie wurden nur algorithmisch bestimmt und repräsentieren Langzeiteffekte sowie die Verbindung zum adaptiven Immunsystem (Janeway *et al.*, 2001; Nakayama *et al.*, 2011).

Die genauen Ursachen und Mechanismen, die den Umschwung vom angeborenen hin zum adaptivem Immunsystem auslösen, sind weitgehend unbekannt, jedoch spielt die Umwandlung von pDC in DC dabei eine wichtige Rolle (Colonna, Trinchieri und Liu, 2004; Liu, 2005; Cao und Liu, 2007). Die Perzeption der Stimuli durch die Zelle löst verschiedene Signalwege aus, welche den zur Umstrukturierung notwendigen zentralen Eingriff in den Zellmetabolismus auslösen und steuern. Gene, die diesen Umschwung realisieren, sollten sowohl Einfluss sowohl auf Zellmetabolismus als auch auf die induzierte Immunantwort besitzen. Dabei ist jedoch der metabolische Einfluss und die damit verbundene Ausdifferenzierung in DC wichtiger als die direkte Immunrelevanz. Diese kann durch weitere, ebenfalls aktivierte Signalpfade reguliert werden und so den Wechsel von angeborener zu adaptiver Immunantwort realisieren. Nicht immunrelevante Gene, die sowohl auf metabolische als auch indirekt auf immunrelevante Pfade einwirken, sind PES2 und ALDH2, die durch beide Methoden detektiert wurden.

PES2 wirkt auf die durch NF- $\kappa$ B aktivierten Signalpfade sowie auf NF- $\kappa$ B selbst ein (Kosaka *et al.*, 1994). Durch diese Regulation und die zentrale Stellung von NF- $\kappa$ B im metabolischen System sowie die Rückwirkung auf immunrelevante Pfade kann PES2 eine wichtige Rolle beim Wechsel der Immunantwort wahrnehmen (Kosaka *et al.*, 1994; Fujioka *et al.*, 2004; Brasier, 2006; Clément, Meloche und Servant, 2009; Salminen *et al.*, 2008; Gupta *et al.*, 2011). Durch PES2 wird es theoretisch möglich, das angeborene Immunsystem langsam zurückzufahren und das Adaptive im Laufe der Umwandlung zu DC durch Aktivierung zusätzlicher Signalpfade sukzessive zu steigern (siehe dazu Abb. 4.22). ALDH2<sup>65</sup> erscheint

---

<sup>65</sup>ALDH2: Aldehyddehydrogenase 2, wandelt Acetaldehyd in Acetal um und ist wesentlich für den Alkoholabbau im menschlichen Metabolismus.

als Katalysator im Alkohol-Metabolismus als zu allgemein, um einen möglichen Auslöser des Umschaltens zwischen angeborenem und adaptivem Immunsystem einerseits und der Ausdifferenzierung in DC andererseits darzustellen. Dennoch wurde auch für ALDH2 gezeigt, dass es auf das Immunsystem und immunoassoziierte Krankheiten, wie Alzheimer, sowie auf generelle Zellprozesse einwirkt (Monsonogo *et al.*, 2001; Zambrano *et al.*, 2007; Wang *et al.*, 2008; Cohen, 2009; Swardfager *et al.*, 2010). Genauere Untersuchungen sind notwendig, um den Schaltmechanismus zu verstehen und aufzuklären, wobei zu beachten ist, dass dieser Mechanismus auch nicht-synergistischer Natur sein kann.

### 4.4.3 Zusammenfassung

Der Vergleich zwischen experimenteller und algorithmischer Methode zeigt, dass ein rein experimenteller Ansatz zur Untersuchung von Synergien unzureichend ist. Einerseits werden Synergien durch Dominanzbeziehungen überdeckt und andererseits ist das experimentelle Verfahren nicht in der Lage Synergien auf proteomischer Ebene zu verfolgen. Das algorithmische Verfahren detektiert Synergien anhand topologischer Muster, wobei das Signal und seine Prozessierung vom Rezeptor bis zu den Genen verfolgbar ist. Synergien und ihr Einfluss auf die Zelle werden somit präzise erklärbar. Dabei werden synergistische Interaktionen, zeitlich unabhängig in den stimulirelevanten Signalwegen detektiert wodurch sich die experimentelle Verifikation auf den Vergleich zwischen synergistischer Bedingung und Kontrolle beschränkt wird. Im experimentellen Ansatz zur Aufklärung von Synergien, ist der Vergleich zwischen allen kombinatorischen Möglichkeiten notwendig.

Der Vergleich am Beispiel von IL-3 und Flu zeigt, dass der Algorithmus weniger synergistische Gene als das experimentelle Verfahren bestimmt, diese aber spezifischer mit der Immunantwort korrelieren (siehe Tab. 4.6 und 4.7). Der algorithmische Ansatz ist dem experimentellen überlegen (siehe Abb. 4.24), da er auf die stimuliinduzierten Signalwege fokussiert ist und ausschließlich kausale Abhängigkeiten bestimmt.

Durch Perzipierung von Immunsignalen wird bekanntermaßen die Differenzierung von pDC in DC angestoßen, wobei ein Wechsel von angeborener zu adaptiver Immunantwort realisiert wird (Cao und Liu, 2007). Jedoch ist der genaue

Schaltmechanismus bisher unbekannt. Die vorliegenden Ergebnisse postulieren ein kontinuierliches Umschalten der Immunantwort, während sich die Zelle von pDC in DC wandelt. PES2 und ALDH2 sind mögliche Auslöser und werden sowohl experimentell als auch algorithmisch identifiziert. Beide regulieren generelle metabolische und auch immunrelevante Signalpfade in der Zelle. Weitere Forschungen sind notwendig, um die Verknüpfung sowie die gegenseitige Regulation zwischen angeborener und erworbener Immunantwort zu verstehen.

## 5 Fazit und Ausblick

Im dieser Arbeit und dem GlobCell-Projekt, in dessen Rahmen sie angefertigt wurde, wurde der synergistische Einfluss der Stimuli IL-3 und Flu auf pDC und die Immunantwort untersucht. Die Stimuli beeinflussen angeborenes und erworbenes Immunsystem und verbinden beide Komponenten miteinander, außerdem wird durch IL-3 und Flu die Differenzierung von pDC in DC induziert. Die beiden Stimuli aktivieren in pDC verschiedene Signalpfade, die auch synergistisch miteinander interagieren. Während der Signalweiterleitung wird das durch die Signalpfadinteraktionen ausgelöste Signal moduliert. Schließlich wird der Effekt auf die genomische Ebene projiziert, indem verschiedene Zielgene synergistisch reguliert werden. Diese regulierten Gene induzieren die Produktion von Proteinen, die entweder intrazelluläre Prozesse anpassen oder als Signalmoleküle in den extrazellulären Raum sekretiert werden. Für die Untersuchung von Synergien und ihrer induzierten Immunantwort wurden die aktivierten Signalpfade, ihre regulierten Gene sowie ausgelöste Effekte betrachtet.

Mittels der Auswertung der Ergebnisse eines klassischen experimentellen Verfahrens, welches durch den Projektpartner am Institut Curie in Paris durchgeführt wurde, wurden die durch synergistische Interaktionen regulierten Gene bestimmt. Dazu wurde ein statistischer Vergleich von IL-3+Flu mit den anderen Bedingungen IL-3, Flu und Medium durchgeführt. Es wurde festgestellt, dass insgesamt nur 73 Gene synergistisch beeinflusst werden, wobei nur ca. 30% dieser Gene eine Funktion in der Immunantwort erfüllen. Die geringe Anzahl erklärt sich dadurch, dass die klassische Methodik nicht die synergistischen Interaktionen der Signalpfade untersucht, sondern nur ihre globalen Effekte auf der genomischen Ebene betrachtet.

Um die Vorhersage von synergistisch regulierten Genen zu verbessern und um Synergien auch in den interagierenden Signaltransduktionswegen zu detektieren,

wurde in dieser Arbeit ein neues Netzwerk-basiertes Verfahren für die systematische Detektion von Synergien entwickelt. Die für die Rekonstruktion des Netzwerks notwendigen Informationen wurden aus den Datenbanken TRANSPATH<sup>®</sup> und TRANSFAC<sup>®</sup> sowie zusätzlicher Literatur entnommen. Das Netzwerk enthält die experimentell bestätigten Signaltransduktionsprozesse einer Säugerzelle. Durch die bipartite Repräsentation des Netzwerkgraphen wurde Signaltransduktion in eine Abfolge von Prozessen und beteiligte Signalträger zerlegt. Dies repräsentiert zum einen das grundsätzliche biochemische Verständnis von Interaktionen (Heiner, Gilbert und Donaldson, 2008) und ermöglicht zum anderen den neuartigen algorithmischen Ansatz zur Aufklärung von Synergien anhand ihrer topologischen Eigenschaften in einem bipartiten Netzwerkgraphen.

Neben einem allgemeinen Referenznetzwerk für Säugerzellen wurde in dieser Arbeit auch ein spezifisches Netzwerk, welches die Gegebenheiten von pDC und den eingesetzten Signalen abbildet, rekonstruiert. Insgesamt umfasst das Netzwerk fast 9000 verschiedene Prozesse, in denen mehr als 15.000 verschiedene Proteine, zumeist in Proteinkomplexen (>10.000), miteinander interagieren. Durch 1259 abgebildete Transkriptionsfaktoren werden potentiell 2382 verschiedene Gene reguliert.

Auf diesem Netzwerk basiert der entwickelte Algorithmus zur Detektion von Synergien. Synergien treten dabei nur in Prozessen auf, die durch die Stimuli gemeinsam erreicht werden. Die Produkte dieser Prozesse sind die synergistischen Signalträger, solange sie nicht über alternative, nicht-synergistische, Wege im Signaltransduktionsnetzwerk angebunden waren. Diese von einem Prozess abhängigen, „alternativlosen“ Produkte wurden entlang ihrer Signaltransduktionswege bis zu den synergistisch beeinflussten TFs verfolgt. Die Transkriptionsfaktoren transferieren die synergistischen Signale auf die genomische Ebene, wo durch Adaption der Genexpression zelluläre und regulatorische Effekte ausgelöst werden. Es fiel auf, dass die Anpassung der Genexpression im wesentlichen von nur 18 Transkriptionsfaktoren induziert wurde. Diese TFs sind in einer Hauptregulationsebene organisiert. Zum einen werden dadurch die synergistischen Gene direkt reguliert und zum anderen beeinflussen die TFs der Hauptregulationsebene weitere TFs, beispielsweise durch Komplexbildung, wodurch die Expression der Zielgene fein justiert wird.

---

Durch Anwendung des entwickelten Algorithmus auf das rekonstruierte Netzwerk wurden 40 synergistisch regulierte Gene bestimmt. Wie in der experimentellen Methode wurden diese Zielgene bezüglich ihrer Funktion in der Immunantwort untersucht. Dabei wurden 70% der vorhergesagten Gene eine Rolle in der Immunantwort zugeordnet. Somit korrelieren die algorithmisch detektierten, synergistischen Gene häufiger mit dem Immunsystem als die experimentell bestimmten Äquivalente.

Erstmalig wurden so Synergien in biologischen Signaltransduktionsnetzwerken durch ihre topologischen Eigenschaften detektiert, wobei die Signalweiterleitung bis zu den regulierten Genen verfolgt werden konnte. Der algorithmische Ansatz hilft dabei Synergien besser zu erklären, Interaktionsprinzipien aufzuklären und ihre Prozessierung zu verstehen. Die regulierten Gene werden als Folge von Weiterleitungsereignissen vorhergesagt. Synergien bilden deshalb ein nachvollziehbares und vorhersagbares Strukturelement in einem bipartiten zellulären Netzwerk. Die durch den Algorithmus vorhergesagten Gene sowie ihre Funktionen im Immunsystem bestätigen dies. Zudem zeigte sich, dass das algorithmische Verfahren Synergien im Kontext der Immunantwort zielgerichteter bestimmt als das vergleichbare experimentelle Verfahren. Der wesentliche Grund ist dabei der starke Fokus des Algorithmus auf signalspezifische Effekte, die das experimentelle Verfahren nicht bereitstellt. Dadurch liefert der entwickelte Algorithmus einen Beitrag zum besseren Verständnis von Signaltransduktion allgemein sowie von Synergien im Speziellen.

Bei der Aufklärung der Rolle von pDC und der Regulation von angeborenem und erworbenem Immunsystem spielen Synergien ebenfalls eine Rolle. Die identifizierten synergistischen Gene und ihre Funktionen zeigen, dass verschiedene Gene gemeinsam und kontinuierlich das Umschalten zwischen den beiden Teilen der Immunantwort realisieren.

Die Darstellung der prozessualen Gegebenheiten durch ein bipartites Netzwerk ermöglicht genauere Vorhersagen regulierter Gene und resultierender Effekte durch die Verfolgung der verschiedenen Stimuli-induzierten Signalwege. Daraus ergeben sich zukünftig vielfältige Möglichkeiten, spezifische Medikamente oder Therapien für Krankheiten zu entwickeln und diese an den detektierten Genen oder der Signalverarbeitung auszurichten, wodurch beispielsweise Wech-

selwirkungen mit anderen Medikamenten reduziert werden können. Jedoch ist die Qualität der Ergebnisse mit der Genauigkeit der im Netzwerk repräsentierten Informationen und der in den Datenbanken verfügbaren Annotationen verbunden. Gerade Daten über die Abundanz von Proteinen in einem bestimmten Zelltyp sind häufig noch nicht verfügbar. Heutzutage wird oft durch Genexpressionsdaten auf die proteomische Zusammensetzung geschlossen, was jedoch ungenau ist (siehe Cox und Mann, 2007; Cravatt, Simon und Yates, 2007; Nilsson *et al.*, 2010). Die existierenden experimentellen Methoden zur gesamtzellulären Bestimmung der Protein-Abundanz sind noch der limitierende Faktor. Die Massenspektrometrie liefert zwar einen groben Überblick über die Zusammensetzung des Proteoms, bestimmt es aber nicht in seinen notwendigen Einzelheiten, wie beispielsweise den vorliegenden Komplexen und Modifikationen. Außerdem sind bei einer gesamtzellulären Aufklärungen des Proteoms durch massenspektrometrische Ansätze die Ergebnisse verschiedener Labore oft nicht miteinander vergleichbar (Cox und Mann, 2007; Wang *et al.*, 2012). Darüber hinaus ist die coverage<sup>66</sup> bei der Bestimmung des kompletten Proteoms für einen Zelltyp häufig zu gering, wie beispielsweise PaxDb zeigt (Cox und Mann, 2007; Wang *et al.*, 2012, und <http://pax-db.org/>). Durch die Untersuchung weiterer Zelltypen und die fortschreitende technologische Entwicklung wird die Aufklärung des Proteoms an Qualität gewinnen. Anhand dieser Daten kann das rekonstruierte Netzwerk weiter verbessert werden. Zellspezifische Interaktionen werden dadurch genauer bestimmbar und die durch verschiedene Stimuli ausgelösten Effekte können besser vorhergesagt werden.

---

<sup>66</sup>Mit coverage wird die Überdeckung bezeichnet. Hier: zu welchem Anteil das Proteom eines Zelltyps aufgeklärt ist.



# Literatur

- T. Adachi und R. Alam. The mechanism of IL-5 signal-transduction. *Am. J. Physiol. Cell Physiol.*, 275:623–633, 1998.
- T. A. Addona, S. E. Abbatiello, B. Schilling, S. J. Skates, D. R. Mani, D. M. Bunk, C. H. Spiegelman, L. J. Zimmerman, A.-J. L. Ham, H. K. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.*, 7:633–641, 2009.
- R. Aebersold und M. Mann. Mass spectrometry-based proteomics. *Nature*, 422:198–207, 2003.
- A. Aguirre, M. E. Rubio und V. Gallo. Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal. *Nature*, 16:323–327, 2010.
- S. Akira. Toll-like Receptor Signaling. *J. Biol. Chem.*, 278:38105–38108, 2003.
- S. Akira, K. Takeda und T. Kaisho. Toll-Like receptors: critical proteins linking innate and acquired immunity. *Nat. Immunol.*, 2:675–68, 2001.
- B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter. *Lehrbuch der molekularen Zellbiologie*, 3. Aufl., bei: Wiley-VCH, Weinheim, 2001.
- B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter. *Molecular biology of the cell*, 4. Aufl., bei: Garland Science Taylor & Francis Group, New York, 2008.
- D. Anastassiou. Computational analysis of the synergy among multiple interacting genes. *Mol. Syst. Biol.*, 3:83, 2007.
- K.-i. Arai, F. Lee, A. Miyajima, S. Miyatake, N. Arai, T. Yokota. Cytokines: Coordinators of Immune and Inflammatory Responses. *Ann. Rev. Biochem.*, 59:783–836, 1990.

- M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.*, 1:25–29, 2000.
- C. Aude-Garcia, V. Collin-Faure, H. Bausinger, D. Hanau, T. Rabilloud, C. Lemerrier. Dual roles for MEF2A and MEF2D during human macrophage terminal differentiation and c-Jun expression. *Biochem. J.*, 430:237–244, 2010.
- M. M. Babu und S. A. Teichmann. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.*, 31:1234–1244, 2003.
- G. D. Bader, M. P. Cary und C. Sander. Pathguide: a pathway resource list. *Nucleic Acids Res.*, 34:504–506, 2006.
- A. Bairoch und B. Boeckmann. The SWISS-PROT protein sequence data bank. *Nucleic Acids Res.*, 20:2019–2022, 1992.
- A.-L. Barabási. *LINKED The New Science of Networks*, bei: Perseus Publishing, Cambridge, 2002.
- A.-L. Barabási und Z. N. Oltvai. Network Biology: Understanding the Cell's Functional Organization. *Nat. Rev. Genet.*, 5:101–113, 2004.
- T. Barrett, T. O. Suzek, D. B. Troup, S. E. Wilhite, W.-C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi und R. Edgar. NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res.*, 33:562–566, 2005.
- E. v. Behring; Kap. Serum Therapy in Therapeutics and Medical Science, S. 1–18, in: *Physiology or Medicine 1901-1921*, Bd. 1, Hrsg. J. Lindsten, bei: World Scientific Publishing Co. Pte. Ltd., Singapore, 1967.
- E. v. Behring und E. Wernicke. Ueber Immunisirung und Heilung von Versuchstieren bei der Diphtherie. *Zeitschrift für Hygiene und Infektionskrankheiten*, 12:10–24, 1892.
- T. Beißbarth und T. P. Speed. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, 20:1464–1465, 2004.
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, 2000.

- R. H. Bortfeldt, S. Schuster und I. Koch. Exhaustive analysis of the modular structure of the spliceosomal assembly network – a Petri net approach. *In Silico Biol.*, 10:89–123, 2010.
- K. d. Bosscher, W. V. Berghe und G. Haegeman. Mechanisms of anti-inflammatory action and of immunosuppression by glucocorticoids: negative interference of activated glucocorticoid receptor with transcription factors. *J. Neuroimmunol.*, 109:16–22, 2000.
- H. R. Bourne, D. A. Sanders und F. McCormick. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature*, 348:125–132, 1990.
- A. R. Brasier. The NF-kappaB regulatory network. *Cardiovasc. Toxicol.*, 2:111–130, 2006.
- J. J. Bright, C. Natarajan, S. Sriram, G. Muthian. Signaling through Jak2-Stat5 Pathway is essential for IL-3-Induced Activation of Microglia. *Glia*, 45:188–196, 2004.
- P. O. Brown und B. D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.*, 21:33–37, 1999.
- X. Bustelo. Vav proteins, adaptors and cell signaling. *Oncogene*, 20:6372–6381, 2001.
- N. A. Campbell, J. B. Reece, L. A. Urry, M. L. Cain, S. A. Wasserman, R. B. Jackson, P. V. Minorsky. *Biologie*, 8. Aufl., bei: Pearson Education, München, 2009.
- R. D. Canales, Y. Luo, J. C. Willey, B. Austermiller, C. C. Barbacioru, C. Boyesen, K. Hunkapiller, R. V. Jensen, C. R. Knight, K. Y. Lee *et al.* Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat. Biotechnol.*, 9:1115–1122, 2006.
- W. Cao und Y.-J. Liu. Innate immune functions of plasmacytoid dendritic cells. *Curr. Opin. Immunol.*, 19:24–30, 2007.
- G. N. von Caron, P. J. Stephens, C. J. Hewitt, J. R. Powell, R. A. Badley. Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. *J. Microbiol. Methods*, 42:97–114, 2000.
- M. C. W. Chan, C. Y. Cheung, W. H. Chui, S. W. Tsao, J. M. Nicholls, Y. O. Chan, R. W. Y. Chan, H. T. Long, L. L. M. Poon, Y. Guan *et al.* Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells. *Respir. Res.*, 6:135, 2005.

- P. Chaturvedi, M. V. R. Reddy und E. P. Reddy. Src kinases and not JAKs activate STATs during IL-3 induced myeloid cell proliferation. *Oncogene*, 16:1749–1758, 1998.
- X. Chen und K. S. Kendler. Interleukin 3 and Schizophrenia. *Am. J. Psychiatry*, 165:13–14, 2008.
- Y. Chen, E. R. Dougherty und M. L. Bittner. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Opt.*, 2:364–374, 1997.
- J. E. Chipuk, T. Moldoveanu, F. Llambi, M. J. Parsons, D. R. Green. The BCL-2 Family Reunion. *Mol. Cell*, 37:299–310, 2010.
- C. Choi, M. Krull, A. Kel, O. Kel-Margoulis, S. Pistor, A. Potapov, N. Voss, E. Wingender. TRANSPATH® - a high quality database focused on signal transduction. *Comp. Funct. Genomics*, 5:163–168, 2004.
- C. Choi, T. Crass, A. Kel, O. Kel-Margoulis, M. Krull, S. Pistor, A. Potapov, N. Voss, E. Wingender. Consistent Re-Modeling of Signaling Pathways and Its Implementation in the TRANSPATH Database. *Genome Inform.*, 15:244–254, 2004.
- J.-F. Clément, S. Meloche und M. J. Servant. The IKK-related kinases: from innate immunity to oncogenesis. *Cell Res.*, 18:889–899, 2009.
- R. M. Cohen. The Role of the Immune System in Alzheimer’s disease. *Focus*, 7:28–35, 2009.
- M. Colonna, G. Trinchieri und Y.-J. Liu. Plasmacytoid dendritic cells in immunity. *Nat. Immunol.*, 5:1219–1226, 2004.
- M. R. Comeau, A.-R. V. d. V. d. Vries, C. R. Maliszewski, L. Galibert. CD123bright Plasmacytoid Predendritic Cells: Progenitors Undergoing Cell Fate Conversion?. *J. Immunol.*, 169:75–83, 2002.
- T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein; Kap. Graph Algorithms, S. 525–700, in: *Introductions to Algorithms*, 2. Aufl., Hrsg. T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, bei: The MIT Press, Cambridge, 2003.
- J. Cox und M. Mann. Is Proteomics the New Genomics?. *Cell*, 130:395–398, 2007.
- N. J. Cox und K. Subbarao. Influenza. *Lancet*, 354:1277–1282, 1999.
- B. F. Cravatt, G. M. Simon und J. R. Yates. The biological impact of mass-spectrometry-based proteomics. *Nature*, 450:991–1000, 2007.

- G. Q. Daley und D. Baltimore. Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. *Proc. Natl. Acad. Sci. USA*, 85:9312–9316, 1988.
- J. E. Darnell, I. M. Kerr und G. R. Stark. Jak-STAT Pathways and Transcriptional Activation in Response to IFNs and Other Extracellular Signaling Proteins. *Science*, 264:1415–1421, 1994.
- T. C. Dawson, M. A. Beck, W. A. Kuziel, F. Henderson, N. Maeda. Contrasting Effects of CCR5 and CCR2 Deficiency in the Pulmonary Inflammatory Response to Influenza A Virus. *Am. J. Pathol.*, 156:1951–1959, 2000.
- G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, R. A. Lempicki. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.*, 4:R60, 2003.
- R. Dey, K. Ji, Z. Liu, L. Chen. A Cytokine-Cytokine Interaction in the Assembly of Higher-Order Structure and Activation of the Interleukine-3:Receptor Complex. *PLoS ONE*, 4:e5188, 2009.
- R. Diestel; Kap. Grundbegriffe, S. 1–31, in: *Graphentheorie*, 2. Aufl., Hrsg. R. Diestel, bei: Springer, Berlin, 2000.
- A. Doreau, A. Belot, J. Bastid, B. R. B. Ranchin, N. Fabien, P. Cochat, M.-C. Trescol-Biemont, C. Pouteil-Noble, P. Trolliet, I. Durieu *et al.* Interleukin 17 acts in synergy with B cell-activating factor to influence B cell biology and the pathophysiology of systemic lupus erythematosus. *Nat. Immunol.*, 10:778–785, 2009.
- H. Dudek, S. R. Datta, T. F. Franke, M. J. Birnbaum, R. Yao, G. M. Cooper, R. A. Segal, D. R. Kaplan, M. E. Greenberg. Regulation of Neuronal Survival by the Serine-Threonine Protein Kinase Akt. *Science*, 275:661–665, 1997.
- R. Edgar, M. Domrachev und A. E. Lash. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, 30:207–210, 2002.
- J. Ellson, E. R. Gansner, E. Koutsofios, S. C. North, G. Woodhull; Kap. Graphviz and Dynagraph – Static and Dynamic Graph Drawing Tools, S. 127–148, in: *Graph Drawing Software*, Hrsg. M. Jünger und P. Mutzel, bei: Springer, Berlin, 2004.
- J. Ellson, E. R. Gansner, L. Koutsofios, S. C. North, G. Woodhull. Graphviz – Open Source Graph Drawing Tools. *LNCS 2265*:483–484, 2002.

- D. H. Erwin und E. H. Davidson. The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.*, 10:141–148, 2009.
- F. Facchetti, C. de Wolf-Peeters, D. Y. Mason, K. Pulford, J. J. van den Oord, V. J. Desmet. Plasmacytoid T cells. Immuno-histochemical evidence for their mono-cyte/macrophage origin. *Am. J. Physiol. Pathol.*, 133:15–21, 1988.
- M. Farnebo, V. J. N. Bykov und K. G. Wiman. The p53 tumor suppressor: A masterregulator of diverse cellular processes and therapeutic target in cancer. *Biochem. Biophys. Res. Commun.*, 396:58–89, 2010.
- A. C. Feller, K. Lennert, H. Stein, H. D. Bruhn, H. H. Wuthe. Immunohistology and aetiology of histiocytic necrotizing lymphadenitis. Report of three instructive cases. *Histopathology*, 7:825–839, 1983.
- R. W. Finberg, J. P. Wang und E. A. Kurt-Jones. Toll like receptors and viruses. *Rev. Med. Virol.*, 17:35–43, 2007.
- W. M. Fitch. Homology a personal view on some of the problems. *Trends Genet.*, 16:227–231, 2000.
- W. Fitch. Distinguishing homologous from analogous proteins. *Syst. Zool.*, 19:99–113, 1970.
- D. A. Frank. STAT Signaling in the Pathogenesis and Treatment of Cancer. *Mol. Med.*, 5:432–456, 1999.
- L. C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40:35–41, 1977.
- G. Frosina, P. Fortini, O. Rossi, F. Carrozzino, G. Raspaglio, L. S. Cox, D. P. Lanei, A. Abbondandolo, E. Dogliotti. Two Pathways for Base Excision Repair in Mammalian Cells. *J. Biol. Chem.*, 271:9573–9578, 1996.
- S. Fujioka, J. Niu, C. Schmidt, G. M. Sclabas, B. Peng, T. Uwagawa, Z. Li, D. B. Evans, J. L. Abbruzzese und P. J. Chiao. NF- $\kappa$ B and AP-1 Connection: Mechanism of NF- $\kappa$ B-Dependent Regulation of AP-1 Activity. *Mol. Cell. Biol.*, 24:7806–7819, 2004.
- F. Fuks. DNA methylation and histone modifications: teaming up to silence genes. *Curr. Opin. Genet. Dev.*, 15:490–495, 2005.
- T. Furukawa, M. Sunamura, F. Motoi, S. Matsuno, A. Horii. Potential Tumor Suppressive Pathway Involving DUSP6/MKP-3 in Pancreatic Cancer. *Am. J. Pathol.*, 162:1807–1815, 2003.

- E. R. Ganser und S. C. North. An open graph visualization system and its applications to software engineering. *Softw. Pract. Exper.*, 00 (S1):1–29, 1999.
- A. Garcia-Lora, I. Algarra und F. Garrido. MHC Class I Antigens, Immune Surveillance, and Tumor Immune Escape. *J. Cell. Physiol.*, 195:346–355, 2003.
- R. J. Garten, C. T. Davis, C. A. Russell, B. Shu, S. Lindstrom, A. Balish, W. M. Sessions, X. Xu, E. Skepner, V. Deyde *et al.* Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans. *Science*, 325:197–200, 2009.
- S. Gasser, S. Orsulic, E. J. Brown, D. H. Raulet. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature*, 436:1186–1190, 2005.
- U. Gaur und B. B. Aggarwal. Regulation of proliferation, survival and apoptosis by members of the TNF superfamily. *Biochem. Pharmacol.*, 66:1403–1408, 2003.
- L. Gautier, L. Cope, B. M. Bolstad, R. A. Irizarry. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20:307–315, 2004.  
*GeneChip® Human Genome U133 Arrays*, Affymetrix Inc., Santa Clara.
- R. Gentleman. *R Programming for Bioinformatics*, bei: CRC Press Taylor & Francis Group, Boca Raton, 2008.
- R. Gentleman, V. J. Carey, S. Dudoit, R. A. Irizarry, W. Huber. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, bei: Springer, Berlin, 2005.
- R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry *et al.* Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.*, 5:R80, 2004.
- C. Ghirelli, R. Zollinger und V. Soumelis. Systematic cytokine receptor profiling reveals GM-CSF as a novel TLR-independent activator of human plasmacytoid dendritic cells. *Blood*, 115:5037–5040, 2010.
- M. Gilliet, W. Cao und Y.-J. Liu. Plasmacytoid dendritic cells: sensing nucleic acids in viral infection and autoimmune diseases. *Nat. Rev. Immunol.*, 8:594–606, 2008.
- T. D. Gilmore. Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene*, 25:6680–6684, 2006.

- L. Giorgetti, T. Siggers, G. Tiana, G. Caprara, S. Notarbartolo, T. Corona, M. Pasparakis, P. Milani, M. L. Bulyk und G. Natoli. Noncooperative Interactions between Transcription Factors and Clustered DNA GBinding Sites Enable Graded Transcriptional Responses to Environmental Inputs. *Mol. Cell*, 37:418–428, 2010.
- B. Goemann, E. Wingender und A. P. Potapov. An approach to evaluate the topological significance of motifs and other patterns in regulatory networks. *BMC Syst. Biol.*, 3:53, 2009.
- B. D. Gomperts, I. M. Kramer und P. E. R. Tatham. *Signal Transduction*, 2. Aufl., bei: Academic Press Elsevier, Burlington, 2009.
- G. Goos; Kap. Abstrakte Datentypen, S. 287–310, in: *Vorlesungen über Informatik*, 3. Aufl., Bd. 1, Hrsg. G. Goos, bei: Springer, Berlin, 2000.
- M. Gouwy, S. Struyf, J. Cattuse, P. Proost, J. V. Damme. Synergy between pro-inflammatory ligands of G protein-coupled receptors in neutrophil activation and migration. *J. Leukoc. Biol.*, 76:185–194, 2004.
- R. P. de Groot, P. J. Coffey und L. Koenderman. Regulation of Proliferation, Differentiation and Survival by the IL-3/IL-5/GM-CSF Receptor Family. *Cell. Signal.*, 9:619–628, 1998.
- G. Grouard, M.-C. Rissoan, L. Filgueira, I. Durand, J. Banchereau, Y.-J. Liu. The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand. *J. Exp. Med.*, 185:1101–1111, 1997.
- C. Guiducci, C. Ghirelli, M.-A. Marloie-Provost, T. Matray, R. L. Coffman, Y.-J. Liu, F. J. Barrat, V. Soumelis. PI3K is critical for the nuclear translocation of IRF-7 and type I IFN production by human plasmacytoid predendritic cells in response to TLR activation. *J. Exp. Med.*, 205:315–322, 2008.
- C. Guiducci, G. Ott, J. H. Chan, E. Damon, C. Calacsan, T. Matray, K.-D. Lee, R. L. Coffman, F. J. Barrat. Properties regulating the nature of plasmacytoid dendritic cell response to Toll-like receptor 9 activation. *J. Exp. Med.* 1999–2008, 2006.
- S. C. Gupta, J. H. Kim, R. Kannappan, S. Reuter, P. M. Dougherty, B. B. Aggarwal. Role of nuclear factor-kB-mediated inflammatory pathways in cancer-related symptoms and their regulation by nutritional agents. *Exp. Biol. Med. (Maywood)*, 236:658–671, 2011.



- M. A. Guthridge, F. C. Stomski, D. Thomas, J. M. Woodcock, C. J. Bagley, M. C. Berndt, A. F. Lopez. Mechanism of Activation of the GM-CSF, IL-3, and IL-5 Family of Receptors. *Stem Cells*, 16:301–313, 1998.
- J. E. Hall; Kap. Unit VI and VII, in: *Guyton and Hall Textbook of Medical Physiology*, 12. Aufl., Hrsg. J. E. Hall, bei: Saunders Elsevier, Philadelphia, 2011.
- X. Han, A. Aslanian und J. R. Yates. Mass Spectrometry for Proteomics. *Curr. Opin. Chem. Biol.*, 12:483–490, 2008.
- M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, 32:D258–D261, 2004.
- T. Heinemeyer, X. Chen, H. Karas, A. E. Kel, O. V. Kel, I. Liebich, T. Meinhardt, I. Reuter, F. Schacherer und E. Wingender. Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res.*, 27:318–322, 1999.
- M. Heiner, D. Gilbert und R. Donaldson; Kap. Petri Nets for Systems and Synthetic Biology, S. 215–264, in: *Formal Methods for Systems Biology*, Hrsg. M. Bernardo, P. Degano und G. Zavattaro, bei: Springer, Berlin, 2008.
- T. Herdegen, R. Böhm, N. Cimin-Bredée, J. Culman, P. Gohlke, L. Ley, G. Luippold, M. Ufer, V. Wätzig. *Kurzlehrbuch Pharmakologie und Toxikologie*, 2. Aufl., bei: Thieme, Stuttgart, 2010.
- K. Hettne, M. Cases, S. Boyer, J. Mestres. Connecting Small Molecules to Nuclear Receptor Pathways. *Curr. Top. Med. Chem.*, 7:1530–1536, 2007.
- S. Hitoshi, T. Alexson, V. Tropepe, D. Donoviel, A. J. Elia, J. S. Nye, R. A. Conlon, T. W. Mak, A. Bernstein und D. v. d. Kooy. Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells. *Genes Dev.*, 16:846–858, 2002.
- W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, B. Goldstein. The Complexity of Complexes in Signal Transduction. *Biotechnol. Bioeng.*, 84:783–794, 2003.
- M. A. Hoeve, A. A. Nash, D. Jackson, R. E. Randall, I. Dransfield. Influenza Virus A Infection of Human Monocyte and Macrophage Subpopulations Reveals Increased Susceptibility Associated with Cell Differentiation. *PLoS ONE*, 7:e29443, 2012.

- G. M. del Hoyo, P. Martín, H. H. Vargas, S. Ruiz, C. F. Arias, C. Ardavín. Characterization of a common precursor population for dendritic cells. *Nature*, 415:1043–1047, 2002.
- B. Huang, J. Zhao, H. Li, K.-L. He, Y. Chen, S.-H. Chen, L. Mayer, J. C. Unkeless, H. Xiong. Toll-Like Receptors on Tumor Cells Facilitate Evasion of Immune Surveillance. *Cancer Res.*, 65:5009–5014, 2005.
- D. W. Huang, B. T. Sherman und R. A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, 4:44–57, 2009.
- D. W. Huang, B. T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M. W. Baseler, H. C. Lane *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.*, 35:169–175, 2007.
- H.-M. Huang, Y.-L. Lee und T.-W. Chang. Jak1 N-Terminus Binds to Conserved Box 1 and Box 2 Motifs of Cytokine Receptor Common beta Subunit but Signal Activation Required Jak1 C-Terminus. *J. Cell Biochem.*, 99:1078–1084, 2006.
- H.-M. Huang, Y.-L. Lin, C.-H. Chen, T.-W. Chang. Simultaneous Activation of JAK1 and JAK2 Confers IL-3 Independent Growth on BA/F3 pro-B Cells. *J. Cell Biochem.*, 96:361–375, 2005.
- T. Ideker, J. Dutkowski und L. Hood. Boosting Signal-to-Noise in Complex Biology: Prior Knowledge Is Power. *Cell*, 144:860–863, 2011.
- R. Ihaka und R. Gentleman. R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.*, 5:299–314, 1996.
- J. N. Ihle. Cytokine receptor signalling. *Nature*, 377:591–594, 1995.
- J. N. Ihle, Y. Weinstein, J. Keller, L. Henderson, E. Palaszynski. Interleukin 3. *Methods Enzymol.*, 116:540–552, 1985.
- N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, 324:218–223, 2009.
- R. A. Irizarry, Z. Wu und H. A. Jaffee. Comparison of Affymetrix GeneChip expression measures. *Bioinformatics*, 22:789–794, 2006.

- R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, T. P. Speed. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, 31:e15, 2003.
- T. Ishida, S. Chada, M. Stipanov, S. Nadaf, F. I. Ciernik, D. I. Gabrilovich, D. P. Carbone. Dendritic cells transduced with wild-type p53 gene elicit potent anti-tumour immune responses. *Clin. Exp. Immunol.*, 117:244–251, 1999.
- A. Iwasaki und R. Medzhitov. Regulation of adaptive immunity by the innate immune system. *Science*, 327:291–295, 2010.
- E. Izaurralde und S. Adam. Transport of macromolecules between the nucleus and the cytoplasm. *RNA*, 4:351–364, 1998.
- C. A. Janeway, P. Travers, M. Walport, M. Shlomchik. *immuno biology*, 5. Aufl., bei: Garland Publishing Taylor & Francis, New York, 2001.
- S. Janowski, B. Kormeier, T. Töpel, K. Hippe, R. Hofestädt, N. Willassen, R. Friesen, S. Rubert, D. Borck, P. Haugen *et al.* Modeling of cell-to-cell communication processes with Petri nets using the example of quorum sensing. *In Silico Biol.*, 10:27–48, 2010.
- D. Jarrossay, G. Napolitani, M. Colonna, F. Sallusto, L. A. Specialization and complementarity in microbial molecule recognition by human myeloid and plasmacytoid dendritic cells. *Eur. J. Immunol.*, 31:3388–3393, 2001.
- D. S. Johnson, A. Martazavi und R. M. Myers. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science*, 316:1497–1502, 2007.
- P. A. Jones und D. Takai. The Role of DNA Methylation in Mammalian Epigenetics. *Science*, 293:1068–1070, 2001.
- D. Kalderon, B. L. Roberts, W. D. Richardson, A. E. Smith. A Short Amino Acid Sequence Able to Specify Nuclear Location. *Cell*, 39:499–509, 1984.
- K. Kandasamy, S. S. Mohan, R. Raju, S. Keerthikumar, G. S. S. Kumar, A. K. Venugopal, D. Telikicherla, J. D. Navarro, S. Mathivanan, C. Pecquet *et al.* NetPath: a public resource of curated signal transduction pathways. *Genome Biol.*, 11:R3, 2010.
- M. Kanehisa. Linking databases and organisms: GenomeNet resources in Japan. *Trends in Biochem. Sci.*, 22:442–444, 1997.
- M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, 34:354–357, 2006.

- M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, 36:480–484, 2008.
- M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya. The KEGG databases at GenomeNet. *Nucleic Acids Res.*, 30:42–46, 2002.
- M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, M. Hattori. The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, 32:D277–D280, 2004.
- M. Karin, Z. g. Liu und E. Zandi. AP-1 function and regulation. *Curr. Opin. in Cell Biol.*, 9:240–246, 1997.
- T. Kawai und S. Akira. Innate immune recognition of viral infection. *Nat. Immunol.*, 7:131–137, 2006.
- M. K. Kerr und G. A. Churchill. Statistical Design and the Analysis of Gene Expression Microarray Data. *Genet. Res. (Camb.)*, 77:123–139, 2001.
- A. R. Khaled, D. V. Bulavin, C. Kittipatarin, W. Q. Li, M. Alvarez, K. Kim, H. A. Young, A. J. Fornce, S. K. Durum. Cytokine-driven cell cycling is mediated through Cdc25A. *J. Cell Biol.*, 169:755–763, 2005.
- M. Kool, M. v. Nimwegen, M. A. Willart, F. Muskens, L. Boon, J. J. Smit, A. Coyle, B. E. Clausen, H. C. Hoogsteden, B. N. Lambrecht *et al.* An Antiinflammatory Role for Plasmacytoid Dendritic Cells in Allergic Airway Inflammation. *J. Immunol.*, 183:1074–1082, 2009.
- E. I. Korpelainen, J. R. Gamble, M. A. Vadas, A. F. Lopez. IL-3 receptor expression, regulation and function in cells of the vasculature. *Immunol. Cell Biol.*, 74:1–7, 1996.
- T. Kosaka, A. Miyata, H. Ihara, S. Hara, T. Sugimoto, O. Takeda, E. i. Takahashi, T. Tanabe. Characterization of the human gene (PTGS2) encoding prostaglandin-endoperoxide synthase 2. *Eur. J. Biochem.*, 221:889–897, 1994.
- A. Kriete, R. Eils, C. Lawrenz, B. Mishra, F. J. Bruggeman, B. M. Bakker, J. Hornberg, H. V. Westerhoff, U. Klingmüller, H.-W. Ma *et al.*; Kap. II. Foundations of Biochemical Network Analysis and Modeling, S. 127–204, in: *Computational Systems Biology*, Hrsg. A. Kriete und R. Eils, bei: Academic Press Elsevier, Burlington, 2006.
- W. P. Krijnen. *Applied Statistics for Bioinformatics using R*, Hanze University, Institute for Life Science und Technology, Groningen, 2009.

- A. Krug, A. R. French, W. Barchet, J. A. Fischer, A. Dzionek, J. T. Pingel, M. M. Orihuela, S. Akira, W. M. Yokoyama und M. Colonna. TLR9-Dependent Recognition of MCMV by IPC and DC Generates Coordinated Cytokine Responses that Activate Antiviral NK Cell Function. *Immunity*, 21:107–119, 2004.
- M. Krull, S. Pistor, N. Voss, A. E. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. P. Potapov, C. Choi *et al.* TRANSPATH®: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.*, 34:546–551, 2006.
- M. Krull, N. Voss, C. Choi, S. Pistor, A. Potapov, E. Wingender. TRANSPATH®: an integrated database on signal transduction and a tool for array analysis. *Nucleic Acids Res.*, 31:97–100, 2003.
- R. Y. Lan, C. Selmi und M. E. Gershwin. The regulatory, inflammatory, and T cell programming roles of interleukin-2 (IL-2). *Autoimmunity*, 31:7–12, 2008.
- P. S. Lenert. Targeting Toll-like receptor signaling in plasmacytoid dendritic cells and autoreactive B cells as a therapy for lupus. *Arthritis Res. Ther.*, 8:203, 2006.
- K. Lennert und W. Remmele. Karyometric research on lymph node cells in man I. Germinoblasts, lymphoblasts and lymphocytes. *Acta Haematol.*, 19:99–113, 1958.
- D. E. Levy und J. J. E. Darnell. Stats: transcriptional control and biological impact. *Nat. Rev. Mol. Cell. Biol.*, 3:651–662, 2002.
- O. Levy. Innate immunity of the newborn: basic mechanisms and clinical correlates. *Nat. Rev. Immunol.*, 7:379–390, 2007.
- S. Li, S. M. Assmann und R. Albert. Predicting Essential Components of Signal Transduction Networks: A Dynamic Model of Guard Cell Abscisic Acid Signaling. *PLoS Biol.*, 4:1732–1748, 2006.
- W. K. Lim, K. Wang, C. Lefebvre, A. Califano. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics*, 23:282–288, 2007.
- W.-W. Lin und M. Karin. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J. Clin. Invest.*, 117:1175–1183, 2007.
- Y.-J. Liu. IPC: Professional Type 1 Interferon-Producing Cells and Plasmacytoid Dendritic Cell Precursors. *Annu. Rev. Immunol.*, 23:275–306, 2005.

- Z. Liu, I. Hazan-Halevy, D. M. Harris, P. Li, A. Ferrajoli, S. Faderl, M. J. Keating, Z. Estrov. STAT-3 activates NF-kappaB in chronic lymphocytic leukemia cells. *Mol. Cancer Res.*, 4:507–515, 2011.
- D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Norton *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, 14:1675–1680, 1996.
- R. M. Locksley, N. Killeen und M. J. Lenardo. The TNF and TNF Receptor Superfamilies: Integrating Mammalian Biology. *Cell*, 104:487–501, 2001.
- A. F. Lopez, T. R. Hercus, P. Ekert, D. R. Littler, M. Guthridge, D. Thomas, H. S. Ramshaw, F. Stomski, M. Perugini, R. D’Andrea *et al.* Molecular basis of cytokine receptor activation. *IUBMB Life*, 7:509–518, 2010.
- J. Lund, A. Sato, S. Akira, R. Medzhitov, A. Iwasaki. Toll-like Receptor 9-mediated Recognition of Herpes Simplex Virus-2 by Plasmacytoid Dendritic Cells. *J. Exp. Med.*, 198:513–520, 2003.
- D. J. Lynn, C. Chan, M. Naseer, M. Yau, R. Lo, A. Sribnaia, G. Ring, J. Que, K. Wee, G. L. Winsor *et al.* Curating the innate immunity interactome. *BMC Syst. Biol.*, 4:117, 2010.
- D. J. Lynn, G. L. Winsor, C. Chan, N. Richard, M. R. Laird, A. Barsky, J. L. Gardy, F. M. Roche, T. H. Chan, N. Shah *et al.* InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.*, 4:218, 2008.
- H. Lüllmann, K. Mohr und L. Hein. *Pharmakologie und Toxikologie*, 17. Aufl., bei: Thieme, Stuttgart, 2010.
- D. Maglott, J. Ostell, K. D. Pruitt, T. Tatusova. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, 35:D26–D31, 2007.
- K.-J. Malmberg und H.-G. Ljunggren. Escape from immune- and nonimmune-mediated tumor surveillance. *Semin. Cancer Biol.*, 16:16–31, 2006.
- P. M. V. Martin und E. Martin-Granel. 2,500-year Evolution of the Term Epidemic. *Emerg. Infect. Dis.*, 12:976–979, 2006.
- W. Martin, T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa und D. Penny. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thou-

- sands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA*, 99:12246–12251, 2002.
- M. Martinez-Moczygemba und D. P. Huston. Biology of common beta receptor-signaling cytokines: IL-3, IL-5, and GM-CSF. *J. Allergy. Clin. Immunol.*, 112:653–665, 2003.
- O. Mason und M. Verwoerd. Graph theory and networks in Biology. *IET Syst. Biol.*, 1:89–119, 2007.
- Y. Matsumoto und K. Kim. Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science*, 269:699–702, 1995.
- Y. Matsumoto, K. Kim und D. F. Bogenhagen. Proliferating cell nuclear antigen-dependent abasic site repair in *Xenopus laevis* oocytes: an alternative pathway of base excision DNA. *Mol. Cell. Biol.*, 14:6187–6197, 1994.
- V. Matys, O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, 34:D108–D110, 2006.
- V. Matys, E. Fricke, R. Geffers, E. Gößling, M. Haubrock, R. Hehl, H. K., H. Karas, A. E. Kel, O. V. Kel-Margoulis *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, 31:374–378, 2003.
- K. McKenna, A.-S. Beignon und N. Bhardwaj. Plasmacytoid Dendritic Cells: Linking Innate and Adaptive Immunity. *J. Virol.*, 79:17–27, 2005.
- K. L. Medina, J. M. Pongubala, K. L. Reddy, D. W. Lancki, R. DeKoter, M. Kieslinger, R. Grosschedl, H. Singh. Assembling a Gene Regulatory Network for Specification of the B Cell Fate. *Dev. Cell.*, 7:607–617, 2004.
- E. Metchnikoff; Kap. Elie Metchnikoff: Immunity in infective diseases, S. 280–326, in: *Physiology or Medicine 1901-1921*, Bd. 1, Hrsg. J. Lindsten, bei: World Scientific Publishing Co. Pte. Ltd., Singapore, 1967.
- A. Monsonego, R. Maron, V. Zota, D. J. Selkoe, H. L. Weiner. Immune hyporesponsiveness to amyloid b-peptide in amyloid precursor protein transgenic mice: Implications for the pathogenesis and treatment of Alzheimer’s disease. *Proc. Natl. Acad. Sci. USA*, 98:10273–10278, 2001.
- S. Monticelli und A. Rao. NFAT1 and NFAT2 are positive regulators of IL-4 gene transcription. *Eur. J. Immunol.*, 10:2971–2978, 2002.

- D. W. Mount. *Bioinformatics Sequence and Genome Analysis*, 2. Aufl., bei: Cold Spring Harbor Laboratory Press, New York, 2004.
- B. Mroczko und M. Szmitkowski. Hematopoietic cytokines as tumor markers. *Clin. Chem. Lab. Med.*, 42:1347–1354, 2004.
- Y. Nagata und K. Todokoro. Interleukin 3 Activates Not Only JAK2 and STAT5, but also Tyk2, STAT1 and STAT3. *Biochem. Biophys. Res. Commun.*, 221:785–789, 1996.
- M. Nakayama, K. Takeda, M. Kawano, T. Takai, N. Ishii, K. Ogasawara. Natural killer (NK)-dendritic cell interactions generate MHC class II-dressed NK cells that regulate CD4+ T cells. *Proc. Natl. Acad. Sci. USA*, 108:18360–18365, 2011.
- D. Nelson und M. Cox. *Lehninger Biochemie*, 3. Aufl., bei: Springer, Berlin, 2001.
- M. E. J. Newman. A measure of betweenness centrality based on random walks. *Soc. Networks*, 29:39–54, 2005.
- C. Nielsen. *Animal Evolution Interrelationships of the living phyla*, 3. Aufl., bei: Oxford University Press, New York, 2012.
- T. Nilsson, M. Mann, R. Aebersold, J. R. Yates III, A. Bairoch, J. J. M. Bergeron. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods*, 7:681–685, 2010.
- B. M. Nitsche, A. F. J. Ram und V. Meyer. The use of open source bioinformatics tools to dissect transcriptomic data. *Methods Mol. Biol.*, 835:311–331, 2012.
- K. Oda, Y. Matsuoka, A. Funahashi, H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.*, 1:2005.0010, 2005.
- U. O’Doherty, M. Peng, S. Gezelter, W. J. Swiggard, M. Betjes, N. Bhardwaj, R. M. Steinman. Human blood contains two subsets of dendritic cells, one immunologically mature and the other immature. *Immunology*, 82:487–493, 1994.
- K. Okudela, T. Yazawa, T. Woo, M. Sakaeda, J. Ishii, H. Mitsui, H. Shimoyamada, H. Sato, M. Tajiri, N. Ogawa *et al.* Down-Regulation of DUSP6 Expression in Lung Cancer. *Am. J. Pathol.*, 175:867–881, 2009.
- T. Opsahl, F. Agneessens und J. Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Networks*, 32:245–251, 2010.



- F. R. S. Owen. *Lectures on the comparative anatomy and physiology of the invertebrate animals*, bei: Longman, Brown, Green & Longmans, London, 1843.
- B. O. Palsson; Kap. 0 - II, S. 1–178, in: *System Biology - Properties of Reconstructed Networks*, Hrsg. B. O. Palsson, bei: Cambridge University Press, Cambridge, 2006.
- N. D. Perkins. Integrating cell-signalling pathways with NF-kappaB and IKK function. *Nat. Rev. Mol. Cell. Biol.*, 1:49–62, 2007.
- L. d. Peso, M. González-García, C. Page, R. Herrera, G. Nuñez. Interleukine-3 - Induced Phosphorylation of BAD Through the Protein Kinase AKT. *Science*, 278:687–689, 1997.
- C. A. Petri. Kommunikation mit Automaten. Dissertation. Fakultät für Mathematik und Physik der Technischen Hochschule Darmstadt, 1962.
- E. Pikarsky, R. M. Porat, I. Stein, R. Abramovitch, S. Amit, S. Kasem, E. Gutkovich-Pyest, S. Urieli-Shoval, E. Galun und Y. Ben-Neriah. NF-kB functions as a tumour promoter in inflammation-associated cancer. *Nature*, 431:461–466, 2004.
- K. A. Pirog, Y. Stabinsky und R. Goldman; Kap. Cytokines, S. 8–294, in: *Cytokine Index*, 3. Aufl., Hrsg. Y. Stabinsky und R. Goldman, bei: ReproTech Inc., Rocky Hill, 2010.
- L. C. Platanius und E. N. Fish. Signaling pathways activated by interferons. *Exp. Hematol.*, 27:1583–1592, 1999.
- A. Potapov, B. Goemann und E. Wingender. The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks. *Bioinformatics*, 9:227, 2008.
- J. M. Pratt, J. Petty, I. Riba-Garcia, D. H. L. Robertson, S. J. Gaskell, S. G. Oliver, R. J. Beynon. Dynamics of Protein Turnover, a Missing Dimension in Proteomics. *Mol. Cell. Proteomics*, 1:579–591, 2002.
- X. Qian, L. Esteban, W. C. Vass, C. Upadhyaya, A. G. Papageorge, K. Yienger, J. M. Ward, D. R. Lowy, E. Santos. The Sos1 and Sos2 Ras-specific exchange factors: differences in placental expression and signaling properties. *EMBO J.*, 19:642–654, 2000.
- J. Quackenbush. Computational Analysis of Microarray Data. *Nat. Rev. Genet.*, 2:418–427, 2001.

- T. Ravasi, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, 140:744–752, 2010.
- M. H. V. v. Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle *et al.* *Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses*, bei: Academic Press, San Diego, 2000.
- R. S. Riley und M. Idowu. *Principles and Applications of Flow Cytometry*, Virginia Commonwealth University, Medical College of Virginia, Department of Pathology, Richmond, 2003.
- M.-C. Rissoan, V. Soumelis, N. Kadowaki, G. Grouard, F. Briere, R. d. Waal Malefyt, Y.-J. Liu. Reciprocal Control of T Helper Cell and Dendritic Cell Differentiation. *Science*, 283:1183–1186, 1999.
- M. E. Ritchie, J. Silver, A. Oshlack, M. Holmes, D. Diyagama, A. Holloway, G. K. Smyth. A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, 23:2700–2707, 2007.
- H. Robbins. An Empirical Bayes Approach to Statistics. *Proc. Third Berkeley Symp. on Math. Statist. and Prob.*, 1:157–163, 1956.
- P. Rodriguez-Viciana, P. H. Warne, A. Khwaja, B. M. Marte, D. Pappin, P. Das, M. D. Waterfield, A. Ridley, J. Downward. Role of Phosphoinositide 3-OH Kinase in Cell Transformation and Control of the Actin Cytoskeleton by Ras. *Cell*, 89:457–467, 1997.
- A. Rojas, G. Fuentes, A. Rausell, A. Valencia. The Ras protein superfamily: Evolutionary tree and role of conserved amino acids. *J. Cell Biol.*, 196:189–201, 2012.
- L. Russell und L. A. Garrett-Sinha. Transcription factor Ets-1 in cytokine and chemokine gene regulation. *Cytokine*, 51:217–226, 2010.
- L. Sachs und J. Hedderich. *Angewandte Statistik Methodensammlung mit R*, 12. Aufl., bei: Springer, Berlin, 2006.
- J. Saldeen und N. Welsh. p38 MAPK inhibits JNK2 and mediates cytokine-activated iNOS induction and apoptosis independently of NF- $\kappa$ B translocation in insulin-producing cells. *Eur. Cytokine Netw.*, 15:47–52, 2004.

- A. Salminen, J. Huuskonen, J. Ojala, A. Kauppinen, K. Kaarniranta, T. Suuronen. Activation of innate immunity system during aging: NF-kB signaling is the molecular culprit of inflamm-aging. *Ageing Res. Rev.*, 7:83–105, 2008.
- A. Sancar, L. A. Lindsey-Boltz, K. Ünsal Kacmaz, S. Linn. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.*, 73:39–85, 2004.
- F. Schacherer, C. Choi, U. Götze, M. Krull, S. Pistor, E. Wingender. The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*, 17:1053–1057, 2001.
- M. Schena, D. Shalon, R. W. Davis, P. O. Brown. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, 270:467–470, 1995.
- M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, N. Rajewsky. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455:58–63, 2008.
- R. B. Seth, L. Sun und Z. J. Chen. Antiviral innate immunity pathways. *Cell Res.*, 16:141–147, 2006.
- H. M. Shapiro. *Practical flow cytometry*, 3. Aufl., bei: Wiley Subscription Services, Inc., A Wiley Company, Hoboken, 1995.
- J. G. Shelton, L. S. Steelman, E. R. White, J. A. McCubrey. Synergy between PI3K/Akt and Raf/MEK/ERK Pathways in IGF-1R Mediated Cell Cycle Progression and Prevention of Apoptosis in Hematopoietic Cells. *Cell Cycle*, 3:372–379, 2004.
- L. Shi, L. H. Reid, W. D. Jones, R. Shippy, J. A. Warrington, S. C. Baker, P. J. Collins, F. d. Longueville, E. S. Kawasaki, K. Y. Lee *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.*, 24:1151–1161, 2006.
- A.-B. Shyu, M. F. Wilkinson und A. v. Hoof. Messenger RNA regulation: to translate or to degrade. *EMBO J.*, 27:471–481, 2008.
- F. P. Siegal, N. Kadowaki, M. Shodell, P. A. Fitzgerald-Bocarsly, K. Shah, S. Ho, S. Antonenko, Y.-J. Liu. The nature of the principal type 1 interferon-producing cells in human blood. *Science*, 284:1835–1837, 1999.

- V. Singh und A. Aballay. Heat Shock and Genetic Activation of HSF-1 Enhance Immunity to Bacteria. *Cell Cycle*, 5:2443–2446, 2006.
- M. Sitkovsky und D. Lukashev. Regulation of immune cells by local-tissue oxygen tension: HIF1- $\alpha$  and adenosine receptors. *Nat. Rev. Immunol.*, 5:712–721, 2005.
- A. D. Smith, S. P. Datta, G. H. Smith, P. N. Campbell, R. Bentley, H. A. McKenzie. *Oxford Dictionary of Biochemistry and Molecular Biology*, 2. Aufl., bei: Oxford University Press, New York, 2003.
- W. Smith, C. H. Andrewes und P. P. Laidlaw. A virus obtained from influenza patients. *Lancet*, 222:66–68, 1933.
- G. K. Smyth; Kap. Limma: linear models for microarray data, S. 397–420, in: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Hrsg. R. Gentleman, V. Carey, S. Dudoit, W. H. R. Irizarry, bei: Springer, New York, 2005.
- G. K. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, 3:3, 2004.
- G. K. Smyth, M. Ritchie, N. Thorne, J. Wettenhall. *limma: Linear Models for Microarray Data User's Guide*, The Walter und Eliza Hall Institute of Medical Research, Melbourne, 2007.
- R. W. Sobol, J. H. Horton, R. Kühn, H. Gu, R. K. Singhal, R. Prasad, K. Rajewsky, S. H. Wilson. Requirement of mammalian DNA polymerase- $\beta$  in base-excision repair. *Nature*, 379:183–186, 1996.
- Z. Songyang, D. Baltimore, L. C. Cantley, D. R. Kaplan, T. F. Franke. Interleukin 3-dependent survival by the Akt protein kinase. *Proc. Natl. Acad. Sci. USA*, 94:11345–11350, 1997.
- A. Spagnoli, U. Lucca, G. Menasce, L. Bandera, G. Cizza, G. Forloni, M. Tetamanti, L. Frattura, P. Tiraboschi, M. Comelli *et al.* Long-term acetyl-L-carnitine treatment in Alzheimer's disease. *Neurology*, 41:1726–1732, 1991.
- A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, 102:15545–15550, 2005.

- W. Swardfager, K. Lanctôt, L. Rothenburg, A. Wong, J. Cappell, N. Herrmann. A Meta-Analysis of Cytokines in Alzheimer's Disease. *Biol. Psychiatry*, 68:930–941, 2010.
- K. Takeda und S. Akira. TLR signaling pathways. *Semin. Immunol.*, 16:3–9, 2004.
- K. Takeda und S. Akira. Toll-like receptors in innate immunity. *Int. Immunol.*, 17:1–14, 2005.
- R. L. Tatusov, E. V. Koonin und D. J. Lipman. A Genomic Perspective on Protein Families. *Science*, 278:631–637, 1997.
- S. A. Teichmann und M. M. Babu. Gene regulatory network growth by duplication. *Nat. Genet.*, 36:492–496, 2004.
- S. Tong, Y. Li, P. Rivailler, C. Conrardy, D. A. A. Castillo, L.-M. Chen, S. Recuenco, J. A. Ellison, C. T. Davis, I. A. York *et al.* A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. USA*, 109:4269–4274, 2012.
- P. Tripathi und A. Aggarwal. NF- $\kappa$ B transcription factor: a key player in the generation of immune response. *Curr. Sci.*, 90:519–531, 2006.
- S. Uematsu und S. Akira. Toll-like Receptors and Type I Interferons. *J. Biol. Chem.*, 282:15319–15324, 2007.
- N. Vezzio, M. Sarfati, L.-P. Yang, C. E. Demeure, G. Delespesse. Human Th2-like cell clones induce IL-12 production by dendritic cells and may express several cytokine profiles. *Int. Immunol.*, 8:1963–1970, 1996.
- M. Viau und M. Zouali. B-lymphocytes, innate immunity, and autoimmunity. *Clin. Immunol.*, 114:17–26, 2005.
- A. Villunger und A. Strasser. The great escape: Is immune evasion required for tumor progression?. *Nat. Med.*, 5:874–875, 1999.
- K. E. d. Visser, A. Eichten und L. M. Coussens. Paradoxical roles of the immune system during cancer development. *Nat. Rev. Cancer*, 6:24–37, 2006.
- E. Vivier. Natural Killer Cell Signaling Pathways. *Science*, 306:1517–1519, 2004.
- L. M. Wakefield und A. B. Roberts. TGF- $\beta$  signaling: positive and negative effects on tumorigenesis. *Curr. Opin. Genet. Dev.*, 12:22–29, 2002.
- B. Wang, J. Wang, S. Zhou, S. Tan, X. He, Z. Yang, Y.-C. Xie, S. Li, C. Zheng und X. Ma. The association of mitochondrial aldehyde dehydrogenase gene

- (ALDH2) polymorphism with susceptibility to late-onset Alzheimer's disease in Chinese. *J. Neurol. Sci.*, 268:172–175, 2008.
- M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner, C. v. Mering. PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life. *Mol. Cell. Proteomics*, 11:492–500, 2012.
- S. Watanabe, T. Itoh und K.-I. Arai. Jak2 Is Essential for Activation of *c-fos* and *c-myc* Promoters and Cell Proliferation through the Human Granulocyte-Macrophage Colony-stimulating Factor Receptor in BA/F3 Cells. *J. Biol. Chem.*, 271:12681–12686, 1996.
- J. D. Watson und F. H. C. Crick. Molecular Structure of Nucleic Acids. *Nature*, 171:737–378, 1953.
- J. D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, R. Losick. *Molecular Biology of the Gene*, 6. Aufl., bei: Pearson/Benjamin Cummings, San Francisco, 2008.
- R. Wehner und W. Gehring; Kap. Kapitel 4, 6 und 7, S. 250–321; 344–421, in: *Zoologie*, 23. Aufl., bei: Thieme, Stuttgart, 1995.
- A. Y. Wen, K. M. Sakamoto und L. S. Miller. The Role of the Transcription Factor CREB in Immune Function. *J. Immunol.*, 185:6413–6419, 2010.
- T. Whittington, M. C. Frith, J. Johnson, T. L. Bailey. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res.*, 39:e98, 2011.
- E. Wingender. *Gene Regulation in Eukaryotes*, bei: VCH Weinheim, Weinheim, 1993.
- E. Wingender. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.*, 9:326–332, 2008.
- E. Wingender, X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhäuser *et al.* The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, 29:281–283, 2001.
- E. Wingender, P. Dietze, H. Karas, R. Knüppel. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.*, 24:238–241, 1996.
- E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Prüß, I. Reuter und F. Schacherer. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.*, 28:316–319, 2000.

- J. M. Woodcock, C. J. Bagley und A. F. Lopez. The functional basis of granulocyte-macrophage colony stimulating factor, interleukin-3 and interleukin-5 receptor receptor activation, basic and clinical implications. *Int. J. Biochem. Cell Biol.*, 31:1017–1025, 1999.
- Z. Wu, R. A. Irizarry, F. M. Murillo, R. Gentleman, F. Spencer. A Model Based Background Adjustment for Oligonucleotide Expression Arrays. *J. Am. Stat. Assoc.*, 99:909–917, 2004.
- S. Wuchty und P. F. Stadler. Centers of complex networks. *J. Theor. Biol.*, 223:45–53, 2003.
- Z. Yao, Y. Cui, W. T. Watford, J. H. Bream, K. Yamaoka, B. D. Hissong, D. Li, S. K. Durum, Q. Jiang, A. Bhandoola *et al.* Stat5a/b are essential for normal lymphoid development and differentiation.. *Proc. Natl. Acad. Sci. USA*, 103:1000–1005, 2006.
- A. Yoshimura, M. Ichihara, I. Kinjyo, M. Moriyama, N. G. Copeland, D. J. Gilbert, N. A. Jenkins, T. Hara, A. Miyajima. Mouse oncostatin M: an immediate early gene induced by multiple cytokines through the JAK-STAT5 pathway. *EMBO J.*, 15:1055–1063, 1996.
- R. J. Youle und A. Strasser. The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell. Biol.*, 9:47–59, 2008.
- A. Zambrano, C. Otth, R. B. Maccioni, I. I. Concha. IL-3 Control TAU Modifications and Protects Cortical Neurons from Neurodegeneration.. *Curr. Alzheimer Res.*, 7:615–624, 2010.
- A. Zambrano, C. Otth, L. Mijica, I. I. Concha, R. B. Maccioni. Interleukine-3 prevents neuronal death induced by amyloid peptide. *BMC Neurosci.*, 8:82, 2007.
- J. D. Zhang und S. Wiemann. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics*, 25:1470–1471, 2009.
- S.-Y. Zhang, E. Jouanguy, V. Sancho-Shimizu, K. Yang, L. Abel, C. Picard, A. Puel, J.-L. Casanova. Human Toll-like receptor-dependent induction of interferons in protective immunity to viruses. *Immunol. Rev.*, 220:225–236, 2007.
- J. Zhao, H. Yu, J.-H. Luo, Z.-W. Cao, Y.-X. Li. Hierarchical modularity of nested bow-ties in metabolic networks. *BMC Bioinformatics*, 7:386, 2006.

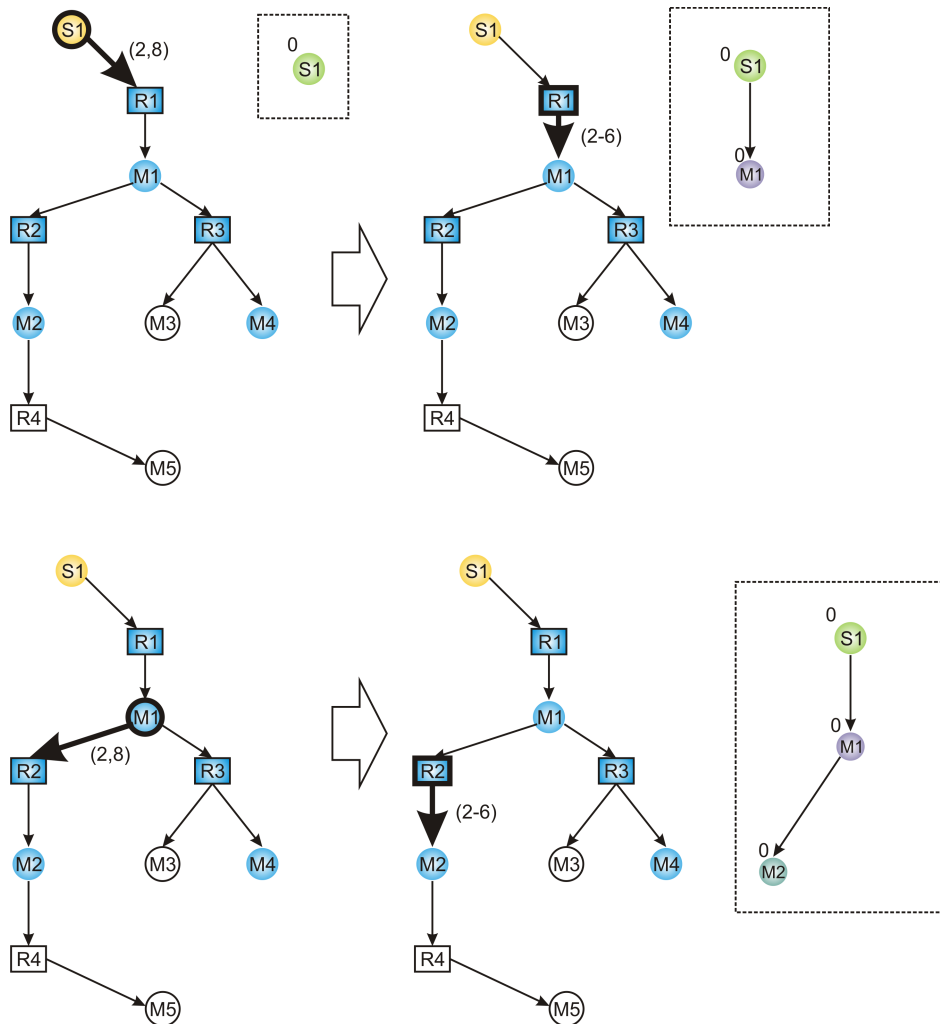
- Y. Zhu und P. J. Bertics. Chemoattractant-induced signaling via the Ras-ERK and PI3K-Akt networks, along with leukotriene C4 release, is dependent on the tyrosine kinase Lyn in IL-5- and IL-3-primed human blood eosinophils. *J. Immunol.*, 186:516–526, 2011.
- E. I. Zuniga, D. B. McGavern, J. L. Pruneda-Paz, C. Teng, M. B. A. Oldstone. Bone marrow plasmacytoid dendritic cells can differentiate into myeloid dendritic cells upon virus infection. *Nat. Immunol.*, 5:1227–1234, 2004.

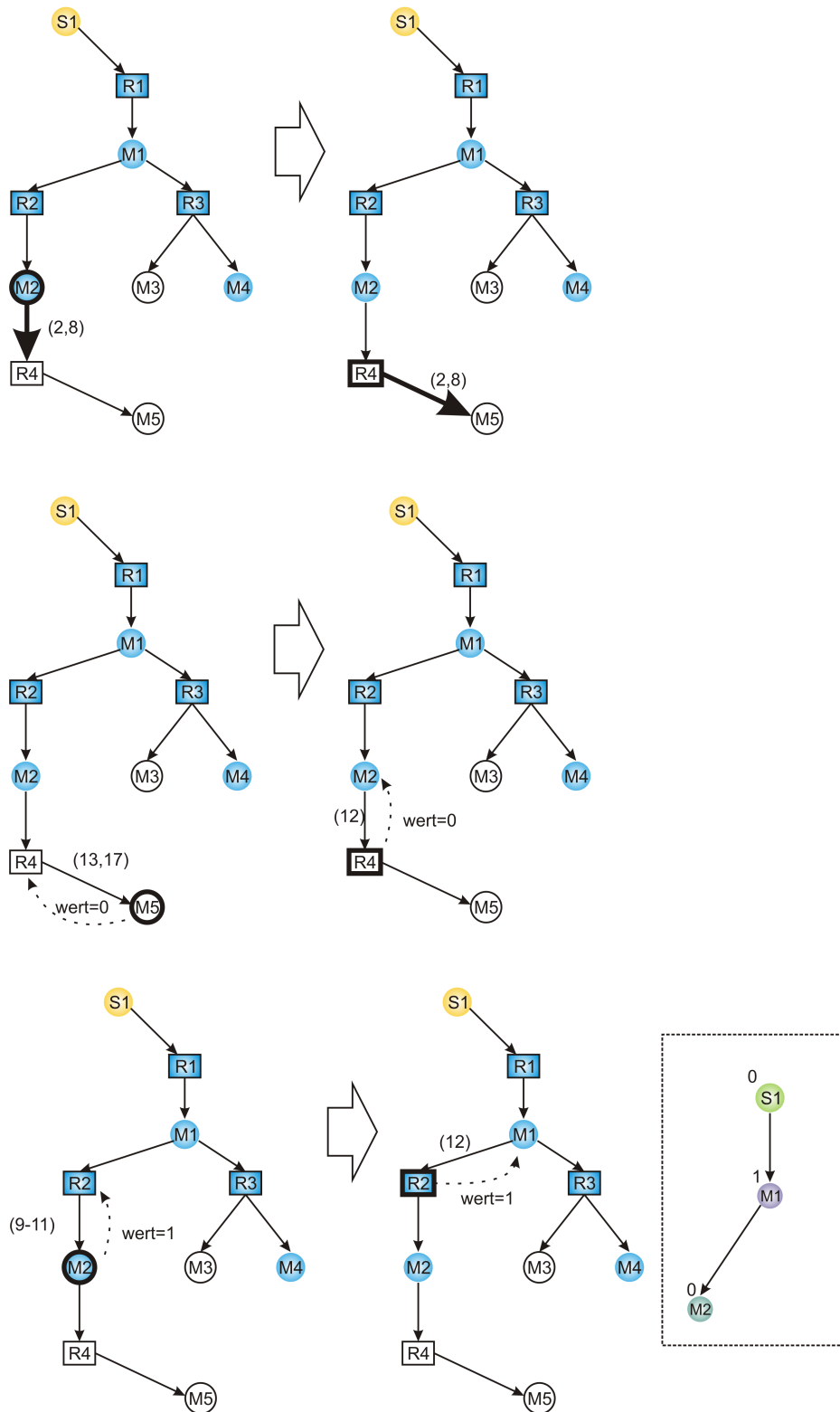


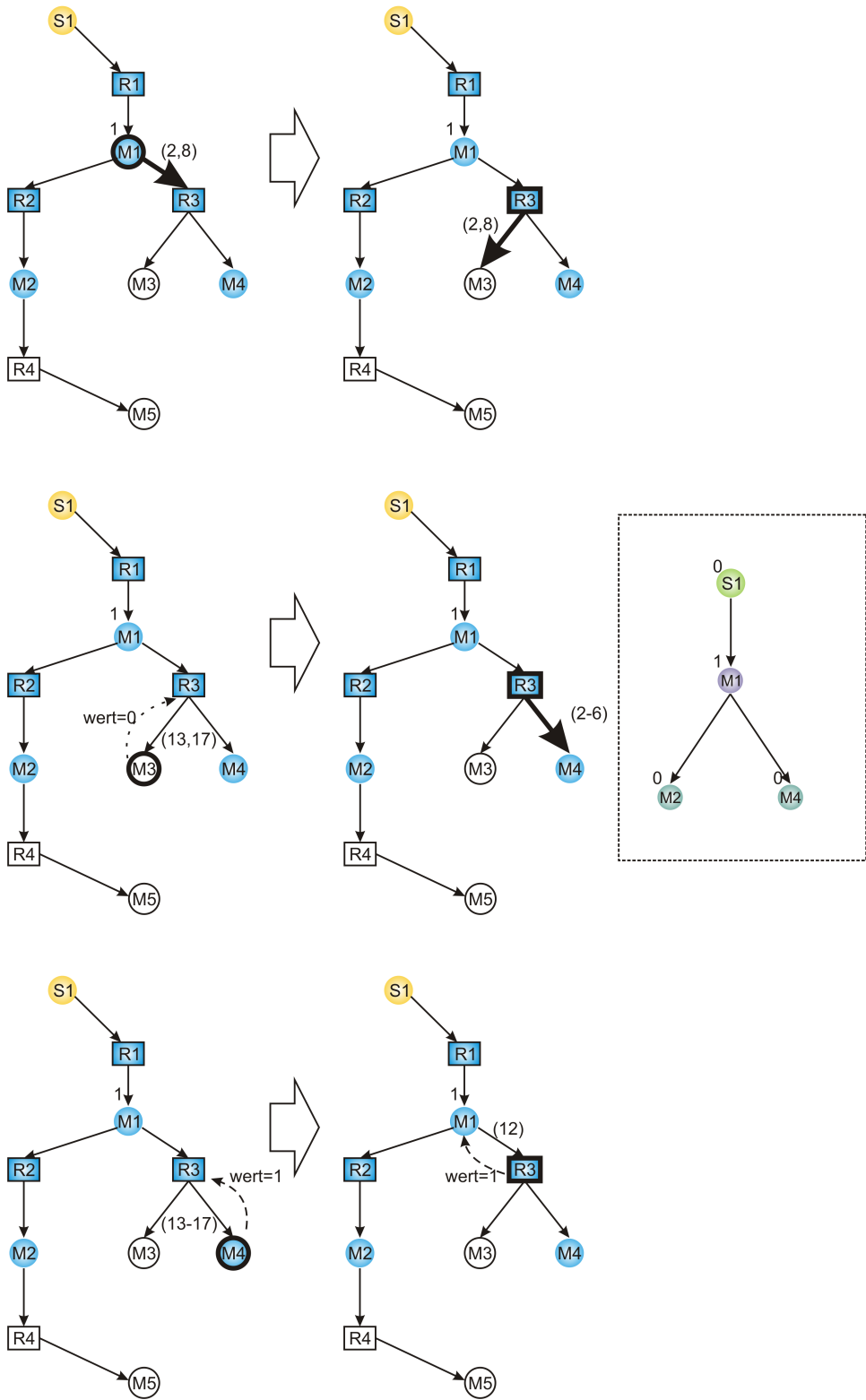
# Appendix

## Bestimmung von Abhängigkeiten nach Unterabschnitt 4.3.3

Auf den folgenden Seiten werden anhand des resultierenden synergistischen Graphen (Abb. 4.15) die einzelnen Schritte zur Bestimmung der Abhängigkeiten zwischen den synergistischen Molekülen abgebildet.







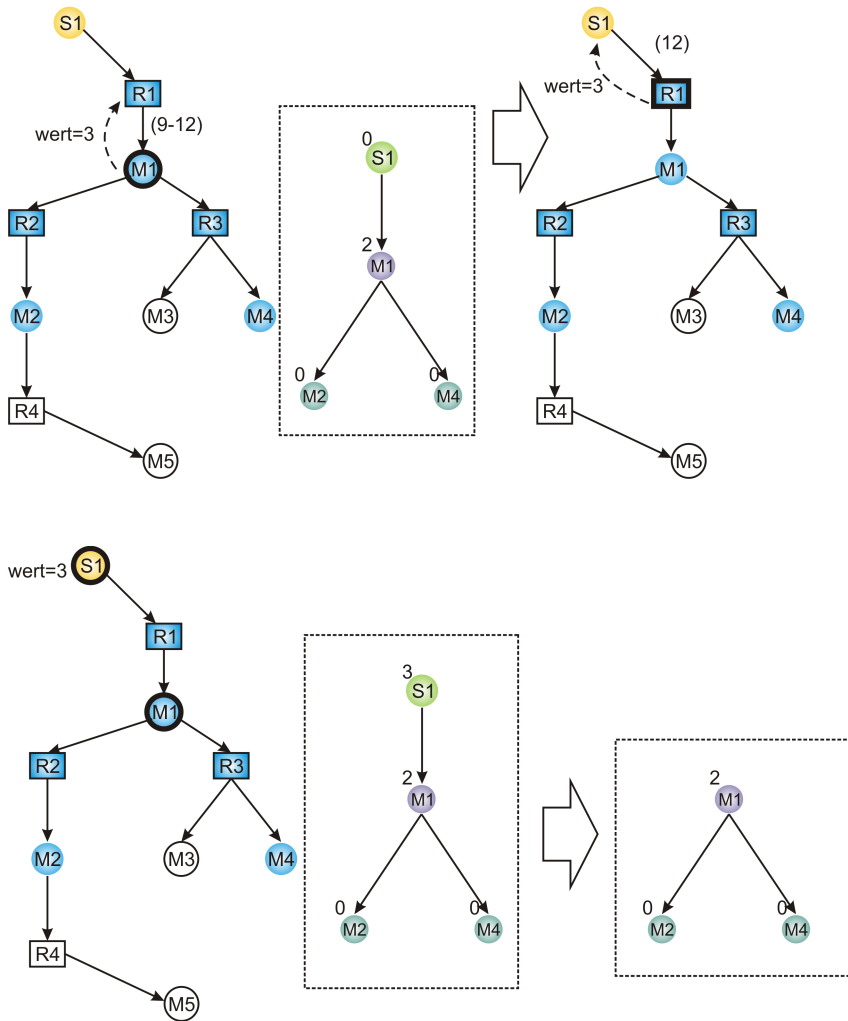


Abbildung 5.1: Beispielhafter Ablauf des Algorithmus zur Bestimmung der Abhängigkeiten (siehe Unterabschnitt 4.3.3). Dabei wird immer von links nach rechts gelesen. Hervorgehoben sind dabei immer der aktuell betrachtete Knoten und die jeweilige Kante. In Klammern an den Kanten sind die jeweiligen algorithmischen Schritte dargestellt. Gestrichelte Kanten zeigen jeweils an, welcher *wert* zurückgegeben wird, also wie viele andere synergistische Knoten erreicht werden. In jeweils abgetrennten Bild wird der Abhängigkeitsbaum sukzessive aufgebaut. Dabei repräsentieren die Zahlen oben links an einem Knoten jeweils den aktuellen *wert* des Knoten.

Gen	IL-3+Flu	Medium	IL-3+Flu - Medium	immun- relevant
NEDD4L	4.9402	2.2008	2.7394	-
ASAH1	4.88955	7.56055	-2.671	-
PDE4DIP	5.3458	2.6882	2.6576	-
CDYL	6.1604	8.8102	-2.6498	-
TZFP	4.2121	1.5655	2.6466	-
TNFRSF11A	4.8076	2.2946	2.513	+
CDC42EP3	4.3708	1.8586	2.5122	-
uPA (PLAU)	4.3524	1.8468	2.5056	+
met (HGFR)	4.4769	2.0458	2.4311	+
ALDH2	5.8429	8.2457	-2.4028	-
DDOST	3.1991	5.5389	-2.3398	-
PINK1	3.1991	5.5389	-2.3398	-
CD36	3.8192	6.02753	-2.20833	+
FLNB	4.93105	7.1341	-2.20305	-
RUNX1	5.7248	7.8922	-2.1674	-
CAT	6.51525	8.63535	-2.1201	-
MAPKAP1	5.5526	3.454	2.0986	-
SLC16A1	3.6171	1.5923	2.0248	-
ANK3	4.4773	2.5908	1.8865	-
PAI-2	3.546	1.678	1.868	-
ITGAL	3.1391	4.9969	-1.8578	+
CLASP2	3.2213	5.0385	-1.8172	-
CTSC	6.0218	7.8358	-1.814	-
PIK3CD	6.668	8.4446	-1.7766	+
EPRS	5.2498	3.5385	1.7113	-
SWAP70	7.9529	6.2725	1.6804	-
S100A6	6.7274	8.2855	-1.5581	-
SMC2L1	2.4443	3.9986	-1.5543	-
RAI17	5.265	6.8052	-1.5402	-
CTSB	6.6036	8.1338	-1.5302	+
USF2	4.3878	5.8569	-1.4691	-
SPRY2	3.0835	1.7165	1.367	+
PLAUR	3.8577	2.5188	1.3389	+
MAP4K5	6.0738	4.8319	1.2419	-
PTGER3	3.1691	1.96155	1.20755	-
SRC-3	5.8068	4.6797	1.1271	-
GPR9 (CXCR3)	5.0252	6.1416	-1.1164	+
KLF5	3.1346	2.0265	1.1081	-
GLI3	2.6711	1.5862	1.0849	-
IL27RA	5.1622	6.2442	-1.082	+
TTN	3.0653	1.9892	1.0761	-
IDE	3.1053	4.15125	-1.04595	-
RAP2B	5.31633	6.33187	-1.01554	-
THBD	3.8186	2.8148	1.0038	+
GNAS	5.6724	4.6698	1.0026	-
LRP8	5.9013	4.9093	0.992	-
CYP1A1	2.5118	1.5779	0.9339	-
Sox8	2.5724	1.7308	0.8416	-
MAD (MXD1)	5.4006	4.5642	0.8364	-
LMNA	6.3329	5.5716	0.7613	-
GNB4	4.7503	4.3376	0.4127	+

Tabelle 5.1: Experimentell identifizierte Gene, welche nicht in Tab. 4.1 aufgeföhrt sind, geordnet nach absoluter Differenz zwischen IL-3+Flu und Medium (Spalte 4). Spalte 1 zeigt den Namen des identifizierten Gens, die Spalten 2 und 3 jeweils die gemessenen Expressionswerte unter der synergistischen Bedingung (IL-3+Flu) und unter Medium. Ein „+“ in der Spalte immun-relevant zeigt die Zuordnung eines immunrelevanten Terms durch KEGG-PATHWAYS oder KEGG-GENES. Sobald ein Term wie „immune“, „cytokine“, „Toll-like“, „JAK-STAT“, „B-cell“ oder „T-cell“ in der Funktionsannotation detektiert wurde, wurde das Gen als immunrelevant gewertet. 22 von 73 ( $\approx 30.14\%$ ) Genen zeigen zusammen mit Tab. 4.1 eine Immunrelevanz. Hier sind es 13 von 51 Genen.

