

Untersuchung der Genexpression von *Dictyostelium discoideum* nach Infektion mit *Legionella* mittels DNA-Microarrays

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Patrick Farbrother
aus Düsseldorf

Hundt Druck GmbH, Köln

2004

Berichtersteller:

Frau Prof. Dr. A. A. Noegel

Herr Prof. Dr. D. Tautz

Tag der mündlichen Prüfung:

17. Juni 2004

Die vorliegende Arbeit wurde in der Zeit von April 2001 bis April 2004 unter der Anleitung von Frau Prof. Dr. A. A. Noegel am Institut für Biochemie I der Medizinischen Fakultät der Universität zu Köln angefertigt.

1	EINLEITUNG	1
1.1	<i>Dictyostelium discoideum</i>	1
1.2	<i>Legionella</i>	2
1.2.1	<i>Legionella pneumophila</i> Δ dotA	4
1.2.2	<i>Legionella hackeliae</i>	5
1.3	<i>Dictyostelium discoideum</i> als Wirt für <i>Legionella</i>	6
1.4	Das <i>D. discoideum</i> cDNA-Projekt	7
1.5	cDNA-Microarrays	8
1.6	Ziele der Arbeit	9
2	MATERIAL UND METHODEN	11
2.1	Material	11
2.1.1	Labormaterialien	11
2.1.2	Bausätze	11
2.1.3	Chemikalien	11
2.1.3.1	Radiochemikalien	12
2.1.4	Oligonukleotide	12
2.1.5	Puffer und Lösungen	13
2.1.6	Medium für die Kultur von <i>D. discoideum</i>	13
2.1.7	Bakterienstämme	14
2.1.8	Geräte	14
2.1.9	Computer-Programme	15
2.2	Methoden	15
2.2.1	Amplifikation der Sonden für das <i>D. discoideum</i> Microarray	15
2.2.1.1	Amplifikation von 450 Sonden für publizierte Gene von <i>D. discoideum</i>	15
2.2.1.2	Amplifikation von 5423 Sonden des japanischen <i>D. discoideum</i> cDNA-Projekts	16
2.2.1.3	Herstellung von 33 Kontrollsonden	17
2.2.2	Herstellung von Microarrays	18
2.2.2.1	Transfer der Sonden von 96-Loch- in 384-Loch-Mikrotiterplatten	18
2.2.2.2	Druck der Microarrays	18
2.2.2.3	Qualitätskontrolle der Microarrays	19
2.2.3	Zellbiologische Methoden	19
2.2.3.1	Kultivierung von <i>D. discoideum</i> in Flüssigkultur	19
2.2.3.2	Bestimmung der Zelldichte von <i>D. discoideum</i>	19
2.2.3.3	Isolation von totaler RNA aus <i>D. discoideum</i>	19
2.2.3.4	Infektion von <i>D. discoideum</i> mit <i>Legionella</i>	20
2.2.4	Molekularbiologische Methoden	20
2.2.4.1	Isolation von mRNA aus totaler RNA	20
2.2.4.2	Bestimmung von DNA-Konzentrationen durch UV-Absorption	20
2.2.4.3	Bestimmung von DNA-Konzentrationen mit der PicoGreen Methode	20
2.2.4.4	Bestimmung von RNA-Konzentrationen durch UV-Absorption	20
2.2.4.5	Elektrophoretische Auftrennung von DNA in Agarosegelen	21

2.2.4.6	Elektrophoretische Auftrennung von RNA in Agarosegelen	21
2.2.4.9	Northernblot-Analyse	22
2.2.4.7	Präzipitation von DNA und RNA	23
2.2.4.8	Reinigung von PCR-Produkten mit QIAvac	23
2.2.4.10	Quantitative PCR	23
2.2.5	Microarray-Methoden	24
2.2.5.1	Zugabe von SpotReport RNA zur RNA von <i>D. discoideum</i>	24
2.2.5.2	Reverse Transkription von RNA und Fluoreszenzmarkierung mit FairPlay	24
2.2.5.3	Prähybridisierung der Microarrays	24
2.2.5.4	Hybridisierung der Microarrays	24
2.2.5.5	Scannen von Microarrays	25
2.2.5.6	Quantifizieren von Microarrays	25
2.2.6	<i>In silico</i> Methoden	26
2.2.6.1	Identifikation differentiell exprimierter Gene mit Significance Analysis of Microarrays (SAM)	26
2.2.6.2	Clusteranalyse in GeneSpring 6.1	26
2.2.6.3	BLAST-Suchen	26
3	ERGEBNISSE	28
3.1	Herstellung des <i>D. discoideum</i> DNA-Microarrays	28
3.1.1	Vergleich von Methoden zur Amplifikation der Sonden für die Herstellung von DNA-Microarrays	28
3.1.2	Amplifikation der Sonden	30
3.1.3	Transfer der Sonden aus 96-Loch- in 384-Loch-Mikrotiterplatten	31
3.1.4	Druck der Microarrays	32
3.1.5	Das <i>D. discoideum</i> DNA-Microarray	33
3.2	Quantifizierung der gemessenen Signalintensitäten	35
3.3	Entwicklung eines automatisierten Auswerteverfahrens	36
3.3.1	Array Tools	38
3.3.2	Normalisierung der Microarray-Daten in R	41
3.3.3	Identifikation differentiell exprimierter Gene mit SAM	43
3.4	Validierung des Microarray-Systems	43
3.4.1	Das SpotReport System	43
3.4.1.1	Faktoren der Induktion und Repression	44
3.4.1.2	Nachweisgrenze des Microarrays	44
3.4.2	Einfluss der Anzahl von Experimentwiederholungen auf die Messunsicherheit	46
3.4.3	Faktoren, die Einfluss auf die detektierten Genexpressionsänderungen haben	51
3.5	Identifikation der differentiell exprimierten Gene von <i>D. discoideum</i> 24 h nach Infektion mit <i>L. pneumophila</i>	55
3.5.1	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> im Vergleich zu uninfizierten Zellen	55
3.5.1.1	Annotation der Sonden induzierter Gene	57
3.5.1.2	Annotation der Sonden reprimierter Gene	61
3.5.1.3	Kategorisierung der differentiell exprimierten Gene	63
3.5.2	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> im Vergleich zur Koinkubation mit <i>L. hackeliae</i>	66

3.5.3	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> im Vergleich zur Koinkubation mit <i>L. pneumophila</i> Δ dotA	70
3.6	Zeitreihe der Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i>	73
3.6.1	Annotation der Sonden der Kategorien mit regulierten Genen	79
3.6.2	Vergleich von SAM mit dem Lord-Test	81
3.6.2.1	Statistische Analyse	81
3.7	Validierung der Microarray-Ergebnisse mit quantitativer PCR und Northernblot-Analyse	83
4	DISKUSSION	86
4.1	Herstellung des <i>D. discoideum</i> DNA-Microarrays	86
4.1.1	Vergleich von Methoden zur Amplifikation der Sonden für die Herstellung von DNA-Microarrays	86
4.1.2	Sonden des <i>D. discoideum</i> DNA-Microarrays	87
4.2	Das automatisierte Auswerteverfahren	88
4.3	Validierung des Microarray-Systems	88
4.3.1	Das SpotReport System	88
4.3.2	Reproduzierbarkeit der Microarray-Ergebnisse	89
4.3.3	Validierung der Microarray-Ergebnisse mit quantitativer PCR und Northernblots	90
4.4	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> nach 24 Stunden	90
4.4.1	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> im Vergleich zu uninfizierten Zellen	90
4.4.1.1	Annotation der differentiell exprimierten Gene	91
4.4.1.2	Kategorisierung der differentiell exprimierten Gene	94
4.4.2	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> im Vergleich zur Koinkubation mit <i>L. hackeliae</i>	95
4.4.3	Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i> im Vergleich zur Koinkubation mit <i>L. pneumophila</i> Δ dotA	97
4.5	Zeitreihe der Infektion von <i>D. discoideum</i> mit <i>L. pneumophila</i>	99
4.6	Ausblick	100
5	ZUSAMMENFASSUNG	102
6	ABSTRACT	103
7	LITERATURVERZEICHNIS	104
8	ABKÜRZUNGSVERZEICHNIS	114

1 Einleitung

1.1 *Dictyostelium discoideum*

Die soziale Amöbe *Dictyostelium discoideum* wurde 1935 von Raper beschrieben. In seinem natürlichen Habitat, dem Waldboden, lebt der Protist von Bakterien, die er phagozytiert, und vermehrt sich durch mitotische Zellteilung. Die eukaryotischen Zellen haben ein ca. 34 Mbp großes haploides Genom mit etwa 12 000 Genen.

D. discoideum gehört zur Klasse der Mycetozoa, die evolutionär vor der Ausbildung der Metazoa und Fungi eine eigene Klasse gebildet haben (Baldauf und Doolittle, 1997; Baldauf *et al.*, 2000). Die Mycetozoa werden auch als Myxomycetes oder Schleimpilze bezeichnet und sind durch die Bildung von Zellaggregaten und die anschließende Verbreitung von Sporen über die Luft gekennzeichnet. Ist genügend Nahrung vorhanden, lebt *D. discoideum* als einzelne Amöbe. Bei Nahrungsmangel wird ein Entwicklungszyklus initiiert, der mit der Aggregation der Einzelzellen zu einem 20 000 bis 100 000 Zellen umfassenden Pseudoplasmodium beginnt (Brock und Gomer, 1999). Die Ausscheidung des Lockstoffs cAMP steuert das Zusammenströmen der Zellen (Gerisch *et al.*, 1975). Die Zellen differenzieren sich in Präspor- und Prästielzellen und es entsteht ein sogenannter *slug*, der zunächst phototaktisch migriert. In der Endphase der Entwicklung bilden die Prästielzellen dann einen Stiel, das Sorophor, aus, an dem die Präsporen empor wandern und einen Sporenkopf, das Sorokarp, bilden. Die Sporen werden verweht und keimen unter günstigen Umweltbedingungen zu Amöben aus (Abb. 1).

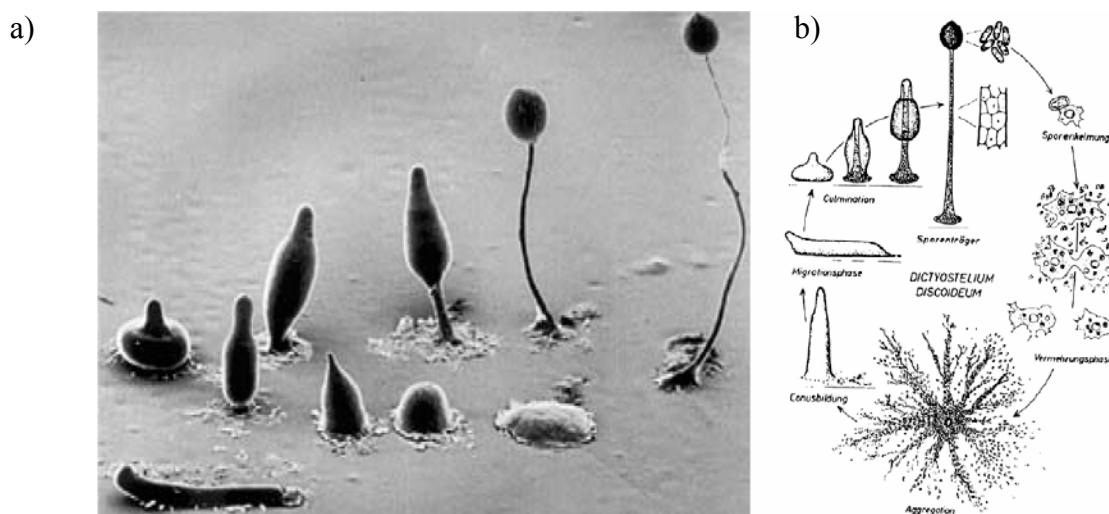


Abb. 1: Entwicklungszyklus von *D. discoideum*. a) Fotomontage rasterelektronenmikroskopischer Aufnahmen der Entwicklungsschritte nach Bildung des Aggregats von Blanton und Grimson (aus Kessin, 2001). b) Schematische Darstellung des Entwicklungszyklus von der Sporenkeimung bis zur Bildung des Sporenträgers (aus Gerisch, 1965).

Aufgrund der Lebensweise und des Entwicklungszyklus ist *D. discoideum* geeignet, um zelluläre Prozesse wie Zellmotilität, Chemotaxis, Phagozytose, Zytokinese, Signaltransduktion und Zelldifferenzierung zu untersuchen (Kessin, 2001; Parent und Devreotes, 1999; Schleicher und Noegel, 1992).

Die Kultivierung im Labor wird dadurch vereinfacht, dass durch Mutagenese des Wildtyps NC4 die Stämme Ax2 und Ax4 erhalten wurden, die sich auch in Flüssigmedium vermehren (Sussman und Sussman, 1967). Durch die Entwicklung einer Transformationsmethode (Nellen *et al.*, 1984), eines Selektionsverfahrens für Mutanten (Gerisch, 1987) sowie der REMI-Mutagenese (Kuspa und Loomis, 1992) ist *D. discoideum* genetisch gut zugänglich. Darüber hinaus gibt es für *D. discoideum* ein cDNA-Projekt (siehe 1.4) und ein Genomprojekt, das sich nach dem Abschluß der Sequenzierung jetzt in der Analysephase befindet.

1.2 *Legionella*

Legionella pneumophila wurde erstmals 1977 aus dem Lungengewebe von Opfern isoliert, die beim Besuch der American Legion Convention 1976 in Philadelphia an Lungenentzündung erkrankt waren. Es hatten sich 182 Personen infiziert, 29 starben. Diese Form der Pneumonie wurde als Legionärskrankheit bekannt (Fraser *et al.*, 1977).

Jährlich werden ca. 25 000 Fälle von *L. pneumophila*-bedingter Lungenentzündung gemeldet (Kwaik, 1998) und inzwischen sind 49 Spezies der Familie *Legionellaceae* beschrieben worden (Cordevant *et al.*, 2003). Die Subspezies von *L. pneumophila* lassen sich in 15 Serogruppen einteilen, während alle übrigen 48 *Legionella*-Spezies zusammen 17 weiteren Serogruppen angehören. Etwa 70 bis 90 % aller *Legionella*-Infektionen werden von *L. pneumophila* der Serogruppen 1 und 6 verursacht. Von 18 weiteren Arten ist bekannt, dass sie Legionärskrankheit oder das mildere Pontiac-Fieber auslösen können (Cordevant *et al.*, 2003).

Legionella ist ein Gram-negatives Bakterium, das sich fakultativ intrazellulär vermehrt (Roy *et al.*, 1998). Es lebt in natürlichen oder künstlichen Gewässern und Wassersystemen. Rowbotham beschrieb 1980 die Fähigkeit von *L. pneumophila*, sich in Protozoen zu vermehren (Rowbotham, 1980). Inzwischen sind 13 Amöbenarten und zwei Ciliaten bekannt, die als Wirte dienen können. Man geht davon aus, dass nicht der Mensch sondern Frischwasser-Amöben den natürlichen Wirt von *L. pneumophila* darstellen (Swanson und Hammer, 2000). Die Infektion alveolarer Makrophagen des Menschen erfolgt durch Inhalation von *L. pneumophila*-haltigen Aerosolen (Wintermeyer *et al.*, 1995). Trotz

Vermehrung in Makrophagen ist die Lunge für *Legionella* eine Sackgasse, denn es wurde bislang noch keine Übertragung von Mensch zu Mensch beobachtet (Swanson und Hammer, 2000).

Die Infektion von Makrophagen beginnt mit der Bindung von *L. pneumophila* an die Zelloberfläche. Dafür ist vermutlich das *major outer membrane protein* (MOMP) von *L. pneumophila* verantwortlich, das die Opsonisierung mit dem Komplementprotein C3 erlaubt. Letzteres wird von den Komplementrezeptoren CR1 und CR3 der Makrophagen gebunden. Darüber hinaus scheint es allerdings auch Komplement-unabhängige Mechanismen zu geben, an denen der Typ IV Pilus und das Hitze-Schock Protein Hsp60 beteiligt sind. Dann erfolgt die Aufnahme durch „wickelnde Phagozytose“ (*coiling phagocytosis*). Innerhalb weniger Minuten bildet sich ein Phagosom mit einer einzigen Membran, das jedoch nicht mit endosomalen Vesikeln fusioniert. Auch eine Fusion mit Lysosomen findet zunächst nicht statt und der pH-Wert des Phagosoms sinkt nicht unter 6,1 (Horwitz, 1983; Horwitz und Maxfield, 1984). Nach etwa einer Stunde interagiert das Phagosom mit Mitochondrien und nach vier Stunden ist es mit Ribosomen besetzt. *L. pneumophila* vermehrt sich dann, wird Säure-tolerant und bildet schließlich eine Flagelle aus. Nach 16 bis 20 Stunden fusioniert die Vakuole der Legionellen mit Lysosomen und der pH-Wert sinkt auf 5,5 (Abb. 2). Swanson *et al.* (2000) vermuten, dass die Legionellen dadurch Nährstoffe erhalten. Der Infektionszyklus endet mit einer von *L. pneumophila* induzierten Nekrose der Wirtszelle. Dabei werden die Legionellen freigesetzt (Gao und Abu Kwaik, 1999).

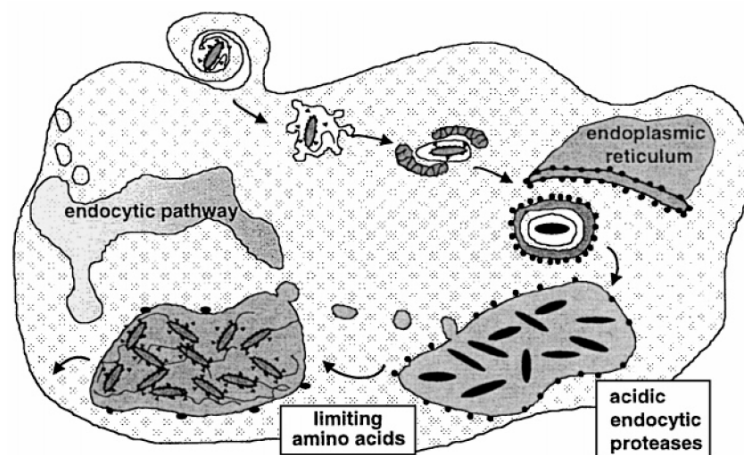


Abb. 2: Modell der Infektion von Makrophagen mit *L. pneumophila*. Die Aufnahme von *L. pneumophila* erfolgt durch *coiling phagocytosis*. Danach entwickelt sich ein Phagosom, das weder mit Endosomen noch mit Lysosomen fusioniert und auch nicht angesäuert wird. Statt dessen lagern sich zunächst Mitochondrien an, später Ribosomen. *Legionella* vermehrt sich, wird Säure-tolerant, bildet ein Flagellum und hört auf, Virulenzfaktoren zu exprimieren. Dadurch kommt es zur Fusion mit Lysosomen, wodurch Nährstoffe bereitgestellt werden. Schließlich werden die Legionellen zytotoxisch und zerstören die Wirtszelle (aus Swanson und Hammer, 2000).

Die Pathogenität von *L. pneumophila* wird von zahlreichen Virulenzfaktoren bestimmt. Bislang sind das Gen *mip* (*macrophage infectivity potentiator*), die Loci *dot* (*defect in organelle trafficking*), *icm* (*intracellular multiplication*), *pmi* (*protozoa and macrophage infectivity*), *mil* (*macrophage infectivity loci*) und ein Komplex für die Aufnahme von Eisen bekannt (Wintermeyer *et al.*, 1995; Gao *et al.*, 1997; Hickey und Cianciotto, 1997; Segal *et al.*, 1998; Vogel *et al.*, 1998). Die Produkte von 24 *dot/icm* Genen bilden ein sogenanntes Typ IV Sekretionssystem aus, das Effektor-Proteine ins Cytosol der Wirtszelle transportiert (Vogel *et al.*, 1998; Swanson und Hammer, 2000). Dieses System ist essentiell für die modifizierte Phagosomenreifung, die *L. pneumophila* das Überleben in Makrophagen ermöglicht. Über die sekretierten Proteine ist wenig bekannt. Im Gegensatz zu anderen Bakterien mit Typ IV System scheint *L. pneumophila* jedoch keine DNA zu transferieren (Solomon *et al.*, 2000).

Es ist außerdem bekannt, dass *L. pneumophila* zytotoxisch wirkt. Die Zellen bilden Legiolysin, ein 39 kDa Protein, das, wenn in *Escherichia coli* produziert, hämolytisch wirkt. Sie sekretieren auch eine 38 kDa große Metalloprotease, die unter anderem in der Lunge von Meerschweinchen pathologische Effekte hervorruft, die für Legionellose typisch sind (Conlan *et al.*, 1986; Wintermeyer *et al.*, 1991). Gao *et al.* (1999) haben gezeigt, dass *L. pneumophila* in Makrophagen durch die Aktivierung der Caspase-Kaskade wenige Stunden nach Infektion Apoptose induzieren kann. Darüber hinaus weisen die Legionellen auch eine kontaktabhängige Zytotoxizität auf, die durch die Bildung von Poren in der Zellmembran der Makrophagen mit weniger als 3 nm Durchmesser hervorgerufen wird (Kirby *et al.*, 1998).

1.2.1 *Legionella pneumophila* Δ dotA

DotA ist eines der 15 Proteine, die von den *dot* Genen kodiert werden. Roy *et al.* (1998) haben gezeigt, dass es erforderlich ist, um die Fusion des Phagosoms mit Lysosomen zu verhindern. Dabei ist die Expression von DotA vor der Aufnahme durch die Wirtszelle und zu Beginn der Infektion nötig. Nachdem ein replikatives Phagosom etabliert wurde, kann sich *L. pneumophila* ohne DotA vermehren. Durch den Vergleich mit einer *dotA*-Deletionsmutante von *L. pneumophila* konnte gezeigt werden, dass die Umgehung der Fusion mit Lysosomen bereits innerhalb der ersten fünf bis 15 Minuten initiiert wird. In der Mutante kolokalisiert das Lysosomen-assoziierte Protein LAMP-1 (*lysosome-associated membrane glycoprotein*) bereits nach fünf Minuten mit 37 % der Phagosomen und nach 15 Minuten mit den meisten Phagosomen. Im Wildtyp kommt es bei der Mehrheit der Phagosomen nicht zu einer Kolokalisation (Roy *et al.*, 1998).

DotA ist ein 1048 Aminosäuren langes Membranprotein der inneren Membran von *L. pneumophila*. Es hat acht Transmembrandomänen, zwei große periplasmatische Domänen mit ca. 503 und 73 Aminosäuren Länge sowie eine cytoplasmatische, C-terminale Domäne aus 122 Aminosäuren (Abb. 3). Der Aufbau von DotA zeigt Ähnlichkeit zu ABC-Transportern von *E. coli* (Roy und Isberg, 1997; Roy *et al.*, 1998). DotA ist Teil des Typ IV Sekretionssystems und die Aminosäuresequenz hat signifikante Ähnlichkeit zu der von TraY auf dem Plasmid Collb-P9 von *Shigella sonnei*. TraY ist Teil eines Typ IV Sekretionssystems, welches das Plasmid während der Konjugation transferiert (Ko *et al.*, 2003).

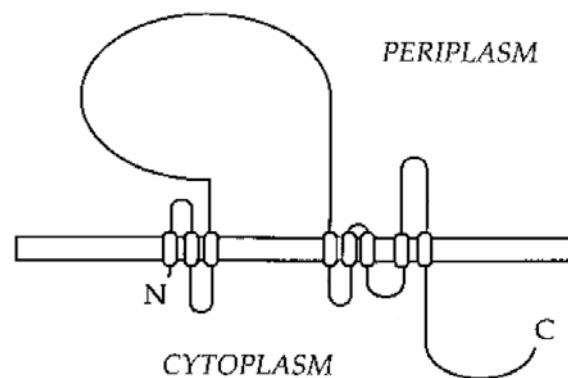


Abb. 3: Vorhergesagte Topologie von DotA. Dargestellt ist die Anordnung der acht Transmembrandomänen in der inneren Membran von *L. pneumophila*. In den periplasmatischen Raum ragen eine 503 AS und eine 73 AS lange Domäne. Der C-Terminus ist 122 AS lang und befindet sich im Cytosol (aus Roy und Isberg, 1997).

1.2.2 *Legionella hackeliae*

Die Spezies *Legionella hackeliae* besteht aus zwei Serogruppen, dem Stamm Lansing 2 (ATCC 35250), der von Hackel aus einer Bronchialbiopsie isoliert wurde und Stamm 798-PA-H (ATCC 35999), der aus einem Lungenaspirat gewonnen wurde (Brenner *et al.*, 1985; Wilkinson *et al.*, 1985). *L. hackeliae* gehört zu den 18 Arten, die gelegentlich Infektionen auslösen. Die Virulenz ist jedoch geringer als die von *L. pneumophila*.

In einem Vergleich haben Alli *et al.* (2003) die Virulenz-Charakteristika von 27 *Legionella*-Stämmen in Makrophagen untersucht. Während *L. pneumophila* die größte Virulenz aufweist, hat *L. hackeliae* eine mittlere Virulenz. Die Zytopathogenität von *L. hackeliae*, gemessen in Prozent der getöteten Wirtszellen, beträgt 45 %. Bei *L. pneumophila* sind es 100 %, bei dem nicht pathogenen Stamm *L. santicrucis* sind es 0 %. Wie die meisten *Legionella*-Stämme induziert *L. hackeliae* nicht Apoptose und bildet keine Poren in der Membran der Wirtszelle. Dazu sind nur *L. pneumophila* und *L. spiritensis* in der Lage. *L. hackeliae* vermehrt sich

intrazellulär in 24 Stunden um den Faktor 10^3 . Bei *L. pneumophila* beträgt der Faktor ca. 10^5 , während sich *L. spiritensis* nicht vermehrt. In einer Southernblot-Analyse mit *EcoRI*-geschnittener genomischer DNA wurde gezeigt, dass *L. hackeliae* wie viele andere Spezies auch zwei Fragmente besitzt, die mit *dot/icm*-Loci hybridisieren. Bei *L. pneumophila* sind es fünf bis sechs (Alli *et al.*, 2003).

In *D. discoideum* kann sich *L. hackeliae* nicht vermehren. Hägele *et al.* (2000) haben gezeigt, dass die Zahl der Kolonie-bildenden Einheiten (cfu) über 96 Stunden kontinuierlich abnimmt. Bei 81,6 % der Phagosomen wurde eine Fusion mit Lysosomen beobachtet. Bei *L. pneumophila* sind es 15,8 %, während es bei *E. coli* 100 % sind. Auch in *Acanthamoeba castellanii* findet keine Vermehrung statt, allerdings nehmen die cfu nur sehr langsam ab.

1.3 *Dictyostelium discoideum* als Wirt für *Legionella*

Die Verwendung von *D. discoideum* als Wirt für *Legionella* wurde erstmals 2000 beschrieben (Hägele *et al.*, 2000). Als Modelle der Infektion wurden zuvor Meerschweinchen, andere Protozoen, Monozyten und verschiedene menschliche Zellen verwendet. Die Verwendung von Amöben ist besonders vorteilhaft, da diese die natürlichen Wirte darstellen (Swanson und Hammer, 2000). Daher sind Amöben und Legionellen auch am besten aneinander angepasst.

Bei Makrophagen, *D. discoideum* und den Frischwasser-Amöben *A. castellanii* und *Hartmannella vermiformis* handelt es sich um motile, professionelle Phagozyten und die Infektion verläuft sehr ähnlich (Solomon und Isberg, 2000; Swanson und Hammer, 2000). In *D. discoideum* wie in Makrophagen wird die Fusion des Phagosoms mit Lysosomen zu einem frühen Zeitpunkt verhindert (Roy *et al.*, 1998; Wiater *et al.*, 1998). Das Phagosom ist in beiden Wirten nach einigen Stunden von Ribosomen umgeben (Horwitz, 1983; Solomon und Isberg, 2000) und *dot/icm*-Mutanten von *Legionella* können sich nicht vermehren (Solomon *et al.*, 2000). Gao *et al.* (1997) haben 89 *pmi*-Mutanten von *Legionella* (siehe 1.2) untersucht und festgestellt, dass die meisten Mutanten in Makrophagen und *A. polyphaga* den gleichen Infektionsverlauf zeigen. Ähnliches wurde für die *icm*-Gene in *A. castellanii* und Makrophagen gefunden (Segal und Shuman, 1999). Unterschiede zwischen *H. vermiformis* und Makrophagen wurden in der Wirkung der Phagozytose- und Pinozytose-Inhibitoren Cytochalasin D und Methylamin gefunden (King *et al.*, 1991).

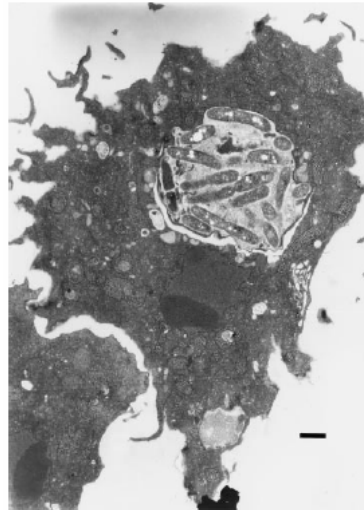


Abb. 4: Infektion von *D. discoideum* mit *L. pneumophila*. Elektronenmikroskopische Aufnahme von *L. pneumophila* in einer Vakuole von *D. discoideum* nach 48 Stunden Koinkubation. Der Balken repräsentiert 1 μm (aus Hägele *et al.*, 2000).

D. discoideum als Wirt für *Legionella* zu verwenden, bietet zusätzlich zur guten experimentellen Zugänglichkeit den Vorteil, dass die Sequenzinformationen des Genom- und des cDNA-Projekts vorhanden sind und eine Vielzahl von Mutanten zur Verfügung stehen (siehe 1.1).

1.4 Das *D. discoideum* cDNA-Projekt

Das *D. discoideum* cDNA-Projekt in Japan hat die Sequenzen von 45 386 *expressed sequence tags* (ESTs) bestimmt und daraus 5381 unabhängige Sequenzen erzeugt. Dazu wurden zunächst 18 432 cDNAs einer Bibliothek von *D. discoideum* aus der *slug*-Phase (nach 14 bis 16 Stunden des Entwicklungszyklus) sequenziert und analysiert (Morio *et al.*, 1998). Danach wurden die Sequenzen weiterer 7720 cDNAs der *slug*-Phase sowie 10 439 cDNAs der vegetativen Phase bestimmt. Diese Sequenzen wurden mit dem Programm PHRAP assembliert. Dabei wurden die ESTs identifiziert, die jeweils aus derselben mRNA und somit demselben Gen hervorgegangen sind. Dadurch wurden 5381 Sequenzen erhalten, die etwa die Hälfte aller Gene von *D. discoideum* repräsentieren und unter <http://www.csm.biol.tsukuba.ac.jp/cDNAproject.html> zugänglich sind. Anschließend wurden die Sequenzen mit allen bekannten Sequenzen verglichen, um Hinweise zur Funktion des jeweiligen Proteins zu erhalten. Etwa ein Drittel der Sequenzen konnte einem von zwölf zellulären Prozessen zugeordnet werden (siehe 3.5.1.3) (Urushihara *et al.*, 2004).

1.5 cDNA-Microarrays

Seit der Bestimmung der Nukleotidsequenz des Bakteriophagen Φ X174 durch Sanger *et al.* (1978) sind mehr als 160 pro- und eukaryotische Genome sowie unzählige cDNA-Sequenzen bestimmt worden und die Geschwindigkeit der Sequenzierung wird immer weiter gesteigert (<http://www.tigr.org>).

Mitte der neunziger Jahre des letzten Jahrhunderts wurden von Schena, Brown und Mitarbeitern sogenannte cDNA-Microarrays entwickelt. Mit ihnen kann die überwältigende Menge an Sequenzinformationen genutzt werden, um die Expression von tausenden Genen gleichzeitig zu bestimmen (Schena *et al.*, 1995; Schena *et al.*, 1996; Shalon *et al.*, 1996). Etwa zeitgleich wurden von Lipshutz *et al.* bei Affymetrix DNA-Microarrays entwickelt, die durch Photolithographie hergestellt werden und die man zur Unterscheidung von cDNA-Microarrays als Oligonukleotid-Microarrays bezeichnet (Lipshutz *et al.*, 1995; Lipshutz *et al.*, 1999).

Bei cDNA-Microarrays handelt es sich meist um gläserne Objektträger, auf die mit Hilfe eines Roboters tausende Sondenpunkte gedruckt werden. Bei den Sonden handelt es sich um DNA-Fragmente mit mehreren hundert Basenpaaren, die jeweils ein Gen repräsentieren und mittels PCR und cDNA-Klonen oder genomischer DNA amplifiziert wurden. Die Messung der Genexpression erfolgt als Vergleich eines Experiments mit einer Kontrolle. Dazu wird RNA aus den Zellen des Experiments und der Kontrolle isoliert, mittels reverser Transkription in cDNA umgeschrieben und mit verschiedenen Fluoreszenz-Farbstoffen markiert. Dann findet eine kompetitive Hybridisierung mit den Sonden des Microarrays statt. Anschließend wird das Microarray gescannt, wobei durch Anregung der Farbstoffe mit zwei unterschiedlichen Lasern die Signale des Experiments und der Kontrolle bestimmt werden. Nach Normalisierung der Signale zur Kompensation der Intensitätsunterschiede der beiden Farbstoffe werden die induzierten und reprimierten Gene durch Bildung des Quotienten aus Experiment und Kontrolle bestimmt (Duggan *et al.*, 1999) (Abb. 5). Da DNA-Microarrays meist viele oder sogar alle Gene eines Organismus repräsentieren, kann ein Profil des „Transkriptoms“ für einen experimentellen Zustand ermittelt werden. DNA-Microarrays wurden für unterschiedlichste Prozesse und Krankheitsbilder wie den Zellzyklus, diauxisches Wachstum, Hitze-Schock, Krebs, die Entzündungsreaktion und die Analyse des Entwicklungszykluses von *D. discoideum* verwendet (DeRisi *et al.*, 1996; Heller *et al.*, 1997; Eisen *et al.*, 1998; Van Driessche *et al.*, 2002). Zusätzlich zur Identifizierung differentiell exprimierter Gene, können z.B. auch Tumore anhand ihres Transkriptionsprofils klassifiziert werden

(Dyrskjot, 2003). Es können auch Entwicklungsmutanten im Vergleich zur Entwicklung des Wildtyps und anderer Mutanten charakterisiert werden (Kibler *et al.*, 2003). Die Verknüpfung von Expressionsprofilen mit der Position der Gene auf dem Chromosom hat 2002 zur Entdeckung von Chromosomenabschnitten mit ähnlich regulierten Genen in *Drosophila* geführt (Spellman und Rubin, 2002).

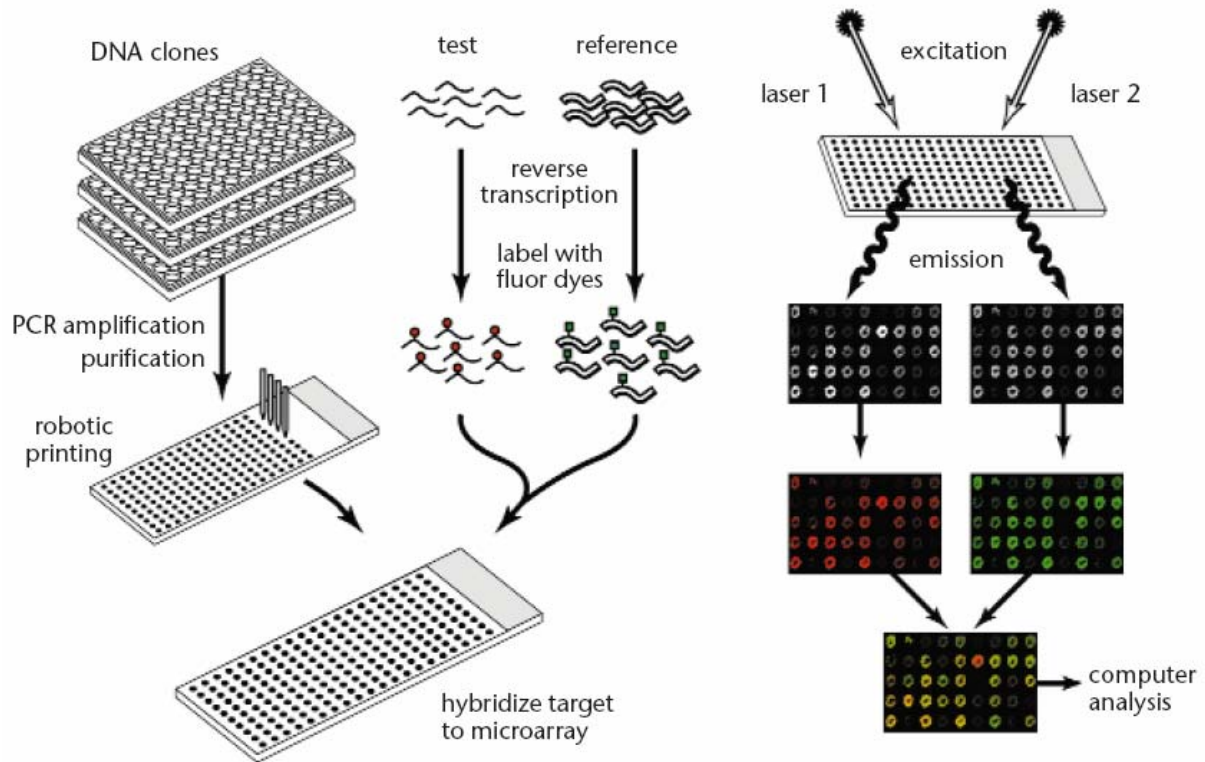


Abb. 5: Prinzip der cDNA-Microarray-Analyse. cDNA-Microarrays werden durch PCR-Amplifikation von DNA-Sonden und anschließendes Drucken auf Objektträger hergestellt. Die Messung der Genexpression wird als Vergleich eines Experiments mit einer Kontrolle durchgeführt. Dazu wird RNA aus Experiment- und Kontroll-Zellen isoliert, revers transkribiert, mit unterschiedlichen Fluoreszenz-Farbstoffen markiert und mit dem Microarray hybridisiert. Das Microarray wird nach Anregung mit zwei verschiedenen Lasern ausgelesen. Die Signale der einzelnen Sonden werden dann quantifiziert und ausgewertet (aus Duggan *et al.*, 1999).

1.6 Ziele der Arbeit

Ziel der Doktorarbeit war es, die Änderungen der Genexpression von *D. discoideum* als Reaktion auf Infektion mit *Legionella* zu untersuchen. Für die Untersuchungen sollten DNA Microarrays für *D. discoideum* entwickelt und hergestellt werden. Grundlage für die Herstellung der DNA-Microarrays war die Sequenzinformation von 450 veröffentlichten *D. discoideum* Genen sowie 5800 nicht-redundanten ESTs des *D. discoideum* cDNA-Projekts. Zusammen repräsentieren diese Sequenzen etwa die Hälfte der erwarteten 12 000 *D. discoideum* Gene.

Die Genexpressionsanalyse sollte in Zusammenarbeit mit dem Institut für Molekulare Infektionsbiologie der Universität Würzburg durchgeführt werden. Dort sollte *D. discoideum* in An- und Abwesenheit von *L. pneumophila*, *L. hackeliae* und *L. pneumophila* $\Delta dotA$ kultiviert werden. Auch die Etablierung eines automatisierten Auswerteverfahrens für DNA-Microarraydaten war geplant. Vorgesehen war neben dem Vergleich von *L. pneumophila* infizierten Zellen mit uninfizierten und mit *L. hackeliae* sowie mit *L. pneumophila* $\Delta dotA$ koinkubierten Zellen auch die Untersuchung einer Zeitreihe der Infektion. Die mit DNA-Microarrays identifizierten differentiell exprimierten Gene sollten anschließend mit alternativen Methoden wie Northernblot-Analyse und quantitativer PCR validiert und die Sonden gegebenenfalls annotiert werden.

2 Material und Methoden

2.1 Material

2.1.1 Labormaterialien

15 ml Röhrenchen, Typ 2095	Falcon
3MM Filterpapier	Whatman
50 ml Röhrenchen, Typ 2070	Falcon
Celluloseacetatfilter, 0,2 µm	Sartorius
Deckgläser	Roth
Einweg-Pipetten, 5 ml, 10 ml und 25 ml	Greiner
Einweg-Pipettenspitzen	Greiner
Hybridisierungsröhren	Hybaid
Mikrotiterplatten, 384 Loch, zylindrisch	AB Gene
Mikrotiterplatten, 96 Loch, runder Boden	AB Gene
Nitrocellulose, Typ BA85	Schleicher und Schuell
Nylonmembran, Biondine B	Gelman Science
Plastikküvette	Eppendorf
Poly-L-Lysin Objektträger	Sigma
Quarzküvette	Infrasil Hellma
Reaktionsgefäße, 0,5 ml, 1,5 ml und 2,2 ml	Sarstedt
Röntgenfilm X-omat AR-5	Kodak
Sterilfilter, 0,2 µm	Gelman Science
UltraGAPS Microarray-Objektträger	Corning

2.1.2 Bausätze

FairPlay Microarray Labeling Kit	Stratagene
Oligotex mRNA Midi Kit	Qiagen
Prime-It II Random Primer Labeling Kit	Stratagene
QIAvac PCR Purification Kit	Qiagen
QuantiTect SYBR Green PCR Kit	Qiagen
RNeasy Midi Kit	Qiagen
SpotReport – 10 Array Validation System	Stratagene

2.1.3 Chemikalien

Die verwendeten Chemikalien hatten, wenn nicht anders vermerkt, den Reinheitsgrad "pro analysi". Es wurde ausschließlich deionisiertes Wasser (Hausanlage) verwendet. Die Microarray-Waschpuffer wurden mit Reinstwasser (Millipore, Billerica, USA) hergestellt.

1kb Plus DNA Ladder	Invitrogen
Agarose	Cambrex
Borsäure	Merck
Bromphenolblau	Serva
BSA	Roth
<i>Bst</i> EIII verdaute DNA des Lambda-Phagen	NEB

Cyanin3 und Cyanin5	Amersham Biosciences
DEPC	Sigma
DMSO	Merck
EDTA	Merck
Ethanol	Riedel-de Haen
Ethidiumbromid	Sigma
Ficoll	Serva
Fisch DNA	Roche
Formaldehyd	Sigma
Formaldehydlösung 37 %ig	Calbiochem
Formamid	Merck
Hefeextrakt	Oxoid
HEPES	Biomol
Maltose	Merck
MOPS	Gerbu
Natriumchlorid	Fluka
Natriumcitrat	Fluka
Natriumdihydrogenphosphat-Dihydrat	Merck
Natriumhydrogenphosphat	Merck
Natriumhydroxid	Riedel-de Haen
Natriumlaurylsarcosinat	Sigma
Nukleotide	Sigma, Roche
Pepton	Oxoid
PicoGreen	Molecular Probes
Polyvinylpyrrolidon	Sigma
Saccharose	Fluka
Salzsäure 32%ig	Fluka
SDS	Serva
Sephadex G50	Pharmacia
Tris	Fluka, Riedel-de Haen

2.1.3.1 Radiochemikalien

α -³²P-Desoxyadenosin-5'-triphosphat (10 mCi/ml) Amersham

2.1.4 Oligonukleotide

Alle Oligonukleotide wurden von Metabion hergestellt. Angegeben sind die universellen Oligonukleotide für die Amplifikation der cDNA-Sonden sowie die Oligonukleotide der quantitativen PCR. Aus Platzgründen sind die 450 Oligonukleotid-Paare der Sonden bekannter Gene nicht aufgeführt.

Table 1: Oligonukleotid-Sequenzen

Name	Sequenz
M13F2	5' - GTA AAA CGA CGG CCA GTG -3'
Amino-M13R2	5' - ACC ATG ATT ACG CCA AGC -3'
H7-F	5' - AGT ATT AGG TGG TGC CAA TC -3'
H7-R	5' - AAA GTG GGC TCT TAA TTG AAC -3'
AnxVII-460F	5' - GGT AAT GGT ACA AAT GAA TCT GAT TTA ATT -3'
AnxVII-760R	5' - GAA GAT TTG TTT AAT GTA TTC CAT TTG TAC -3'
SLD769-F	5' - TAC ATT AAC TGA AGC TGT ACG TAG -3'
SLD769-F	5' - CGA GAT AAA CGA TTG AAA AGG ATC -3'
SSB431-F	5' - TTA TCT TGG TGA AGC AAC TTA CAC -3'
SSB431-R	5' - CGA GTG AAT GAT GGT AAT TGA TCC -3'
SSF806-F	5' - TTT CAT AAA GGT GAT CGC TTT GG -3'
SSF806-R	5' - GAT ACC AGG AGC ATC AGA AGA G -3'
SSK576-F	5' - GAA TGT GAA ACT TAT GAC CCT ACC -3'
SSK576-R	5' - ATG ACC AAT TCT ATT GTG ACC AAC -3'
SSL850-F	5' - TGA TCG TAA TAC AGG TTA CCA CTC -3'
SSL850-R	5' - GTA AAC ATT TGT ACC AGT TGC ATC -3'
VSG622-F	5' - TCA ACA ATC GAC CAA ATT CCT TG -3'
VSG622-R	5' - ACG AAT TGG GAT TGA TCT ATT ACC -3'
SSB153-F	5' - CAA TGA GGC AGC AAT GAA AGA C -3'
SSB153-R	5' - CTT TAC CAC CAA ATT TGG TAC G -3'
SSJ163-F	5' - CTG ACG TTA CAA ATG AAG AAA GTG -3'
SSJ163-R	5' - GAA ATT GTT AAA CCA GTG ACA GC -3'
SSL818-F	5' - CGC CCG CAC TTT ATG AAT TGT TC -3'
SSL818-R	5' - TCA TCA TCA GAA CCC ACA TCT TC -3'

2.1.5 Puffer und Lösungen

SSC (20 x):

3 M NaCl
300 mM Natriumcitrat

TE

10 mM TRIS-HCl (pH 8,0)
1 mM EDTA (pH 8,0)

TBE (10 x, pH 8,3)

1 M TRIS
0,83 M Borsäure
0,01 M EDTA (pH 8,0)

DEPC-Wasser

0,1 % DEPC
in Wasser
autoklavieren

Phosphatpuffer (pH 6,8)1,2 M Na₂HPO₄1,2 M NaH₂PO₄

im Verhältnis zwei zu eins mischen

2.1.6 Medium für die Kultur von *D. discoideum***AX-Medium (1000 ml):**(Claviez *et al.*, 1982)

14,3 g Pepton

7,15 g Hefeextrakt

18 g Maltose

0,616 g Na₂HPO₄ · 2 H₂O0,486 g KH₂PO₄

mit Wasser auf 1 l auffüllen

pH 6,7

autoklavieren

2.1.7 Bakterienstämme*L. pneumophila* Phil I JR32 (Marra und Shuman, 1989)*L. pneumophila* Phil I JR32 LELA 3118 (*dotA* 3118:Tn903DLL *LacZ*)*L. hackeliae* (ATCC 35250) (Brenner *et al.*, 1985)**2.1.8 Geräte**

BioPhotometer

Zentrifuge 5415 D

Entwicklermaschine für Röntgenfilme FPM-100A

Gelapparatur

Geldokumentationssystem

Heizblöcke Dri-Block DB-2D

Hybridisierungssofen

Kühlzentrifuge Megafuge 3.0 R

Lichtmikroskop

Microarray-Drucker: MicroGrid II

Microarray-Drucknadeln: MicroSpot 2500 split pins

Microarray-Hybridisierungskammern

Microarray-Scanner: Scan-Array® 4000XL

Multipette plus

PCR-Maschine PTC-200

pH-Meter 766

Pipetten

Pipettier-Roboter Proteineer dp

QIAvac 96 manifold, regulator und Luer adapter

Quantitative PCR-Maschine Opticon 2

Rüttler

Spannungsverstärker Power Pac 200 und 300

Eppendorf

Eppendorf

Fujifilm

Bio Rad

MWG-Biotech

Techne

Hybaid

Heraeus Instruments

Olympus

BioRobotics

BioRobotics

Corning

PerkinElmer Life Sciences

Eppendorf

MJ Research

Knick

Gilson

Bruker

Qiagen

MJ Research

Scientific Instruments

Bio Rad

Spektralphotometer, Typ Ultraspec 2000
 Vakuumpumpe
 Waage
 Wasserbäder

Pharmacia Biotech
 KNF
 Ohaus
 GFL

2.1.9 Computer-Programme

QuantArray 2.0, ScanArray 3.1 und ScanArray Express Versionen 2.0 und 2.1: PerkinElmer Life Sciences, Wellesley, USA

R Version 1.6.2: *The R Project for Statistical Computing*, <http://www.r-project.org/> mit den Modulen `marrayClasses`, `marrayInput`, `marrayNorm`, `marrayPlots` und `Biobase` von BioConductor (BioConductor, <http://www.bioconductor.org/>).

Significance Analysis of Microarrays (SAM) Version 1.21: SAM ist frei verfügbar beim Department of Statistics, Stanford University unter <http://www-stat.stanford.edu/~tibs/SAM/> erhältlich.

GeneSpring 6.1, Silicon Genetics, Redwood City, USA

2.2 Methoden

2.2.1 Amplifikation der Sonden für das *D. discoideum* Microarray

2.2.1.1 Amplifikation von 450 Sonden für publizierte Gene von *D. discoideum*

Anhand von 450 publizierten Gensequenzen von *D. discoideum* wurden in der Arbeitsgruppe Oligonukleotide für die Amplifikation von genspezifischen Sonden ausgewählt. Dafür wurde das Programm GeneFisher (Giegerich *et al.*, 1996) (<http://bibiserv.techfak.uni-bielefeld.de/genefisher/>) mit den folgenden Parametern verwendet.

Oligonukleotide:

Oligonukleotid-Länge:	20 bis 22 bp
GC-Gehalt:	30 bis 70 %
Schmelztemperatur:	56 bis 62 °C
Produktlänge:	400 bis 600 bp
Degeneriertheit:	0fach
Mehrfaches Vorkommen:	einmaliges Vorkommen

3' Klammer

3' Länge:	4 bp
Degeneriertheit:	0fach
3' GC-Gehalt	45 bis 75 %

Für diese Sonden sind der Name des repräsentierten Gens, die *accession number* der Sequenz in der Sequenzdatenbank GenBank, Intron- und Exongrenzen, Länge der Sonde, Schmelztemperatur der Oligonukleotid-DNA-Hybride und Referenz bekannt.

Die Sonden wurden durch PCR mit genspezifischen Oligonukleotiden und genomischer DNA als Matrize in 96-Loch PCR-Platten amplifiziert. Die Amplifikation erfolgte in einer PTC-200 PCR-Maschine und es wurden je nach Aktivität 0,5 bis 2 μl einer *Taq*-Polymerase eingesetzt, die im Institut für Biochemie I exprimiert und gereinigt wurde. Jede Sonde wurde mit zwei PCR-Reaktionen amplifiziert, um einzelne ausgefallene PCR-Reaktionen ausgleichen zu können.

Reaktionsansatz (100 μl)

83 μl	H ₂ O	
10 μl	10x <i>Taq</i> -Puffer	(100 mM TRIS-HCl pH 8,3, 0,5 M KCl, 20 mM MgCl ₂ , sterilfiltriert)
2 μl	genomische DNA	(ca. 10 ng/ μl)
2 μl	dNTP-Mix	(je 5 mM dATP, dCTP, dGTP, dTTP)
1 μl	Vorwärts-Oligonukleotid	(20 μM)
1 μl	Rückwärts-Oligonukleotid	(20 μM)
0,5 - 2 μl	<i>Taq</i> -Polymerase	

PCR-Programm

1.	5 min 94 °C	Denaturierung der DNA
2.	je 30 mal	Denaturierung der DNA
	40 s 94 °C	
	60 s 56 °C	Hybridisierung der Oligonukleotide
	40 s 72 °C	DNA-Synthese
3.	3 min 72 °C	DNA-Synthese

Als positive PCR-Kontrollen wurden eine Sonde für das Gen *lagC* sowie eine 5' und eine 3' Sonde des Gens Severin-Kinase verwendet. Bei den drei Negativkontrollen wurde die Matrizen-DNA, die Oligonukleotide oder beides weggelassen und der Reaktionsansatz entsprechend mit Wasser aufgefüllt.

Nach der Amplifikation wurde die Synthese von Produkten korrekter Länge mittels Gelelektrophorese überprüft und die Sonden mit Ethanol gefällt. Nach dem Lösen in 11 μl Microarray-Drucklösung (25 % DMSO in Wasser) wurde 1 μl zur Kontrolle der konzentrierten Sonde für eine zweite Gelelektrophorese verwendet. Anschließend wurden die Sonden mit einem Proteiner dp Pipettier-Roboter in 384-Loch Mikrotiterplatten transferiert und bis zum Druck von Microarrays bei – 20 °C gelagert.

2.2.1.2 Amplifikation von 5423 Sonden des japanischen *D. discoideum* cDNA-Projekts

Für die Amplifikation von 5423 Sonden wurde eine Plasmidbank des japanischen *D. discoideum* cDNA-Projekts (Morio *et al.*, 1998) (<http://www.csm.biol.tsukuba.ac.jp/>)

cDNAproject.html) verwendet. Die Plasmide wurden von Prof. Dr. Yoshimasa Tanaka und Prof. Dr. Hideko Urushihara, Universität Tsukuba, Japan, bereitgestellt und stellen einen nicht redundanten Satz von cDNAs dar. Die cDNAs wurden mit RNA aus der vegetativen und der *slug*-Phase des Entwicklungszyklus von *D. discoideum* gewonnen und ihre Sequenz bestimmt. Die Amplifikation der Sonden erfolgte mit den Plasmiden als Matrize und universellen Oligonukleotiden.

Reaktionsansatz (100 μ l)

73 μ l	H ₂ O	
10 μ l	10x <i>Taq</i> -Puffer	(100 mM TRIS-HCl pH 8,3, 0,5 M KCl, 20 mM MgCl ₂ , sterilfiltriert)
10 μ l	Plasmid	1:10 verdünnt
2 μ l	dNTP-Mix	(je 5 mM dATP, dCTP, dGTP, dTTP)
1 μ l	Oligonukleotid M13F2	(20 μ M)
1 μ l	Oligonukleotid Amino-M13R2	(20 μ M)
0,5 - 2 μ l	<i>Taq</i> -Polymerase	

PCR-Programm

1.	5 min 94 °C	Denaturierung der DNA
2.	je 30 mal	Denaturierung der DNA
	40 s 94 °C	Hybridisierung der Oligonukleotide
	60 s 58 °C	DNA-Synthese
	120 s 68 °C	DNA-Synthese
3.	4 min 68 °C	DNA-Synthese

Das Oligonukleotid Amino-M13R2 trägt am 5' Ende eine Aminoallyl-Modifikation, die zur Bindung der Sonden an Objektträger verwendet werden könnte, falls diese eine Beschichtung mit Aldehyd- oder Epoxygruppen tragen.

Die amplifizierten cDNA-Sonden wurden wie die Sonden der publizierten Gene überprüft, konzentriert, in 384-Loch Mikrotiterplatten transferiert und gelagert. Für alle Sonden wurden die vorhandenen Informationen zur Identität und Position in den 96-Loch Mikrotiterplatten in elektronischer Form gespeichert und zur Berechnung der Positionen in den 384-Loch Mikrotiterplatten und auf den Microarrays verwendet.

2.2.1.3 Herstellung von 33 Kontrollsonden

Bei der Herstellung der Microarrays wurden positive und negative Kontrollen, sowie das SpotReport – 10 Array Validation System verwendet. Die positiven Kontrollen waren Sonden für zwölf Gene von *D. discoideum* mit bekanntem Expressionsmuster für verschiedene Phasen des Wachstums sowie genomische DNA. Als negative Kontrollen wurden *Bst* EIII verdaute DNA des Lambda-Phagen und Microarray-Drucklösung eingesetzt.

Die Sonden für die zwölf *D. discoideum* Gene wurden wie in 2.2.1.1 beschrieben amplifiziert, allerdings je 32 PCRs pro Sonde. Die PCR-Reaktionen wurden anschließend vereinigt, um eine ausreichende Menge für zwölf Doppelpunkte pro Microarray zu haben. Für die Produktion der Microarrays wurden die Kontrollen in einer 384-Loch Mikrotiterplatte zusammengestellt.

2.2.2 Herstellung von Microarrays

2.2.2.1 Transfer der Sonden von 96-Loch- in 384-Loch-Mikrotiterplatten

Für den Transfer der amplifizierten Sonden aus 96-Loch-Mikrotiterplatten, in denen sie konzentriert wurden, in 384-Loch-Mikrotiterplatten für den Druck auf Microarrays, wurde der Proteiner dp Pipettier-Roboter verwendet. Für das Steuerungsprogramm DP Control wurde der Pipettiervorgang im *Add-In* 96to384.xla programmiert und eine Arbeitsliste (96to384.xls) erstellt. Damit kann der Proteiner jeweils den Inhalt von vier 96-Loch-Mikrotiterplatten in eine 384-Loch-Platte überführen.

2.2.2.2 Druck der Microarrays

Die Microarrays wurde auf einem MicroGrid II mit 16 MicroSpot 2500 split pins nach den Vorgaben des Herstellers gedruckt. Die MicroSpot 2500 Stifte besitzen eine Kapillare und ein Reservoir, wodurch nach einmaligem Befüllen mehrere hundert Punkte gedruckt werden können. Die Anzahl der möglichen Punkte ist von der Verdunstungsrate der Drucklösung abhängig. Die verwendete Lösung aus 25 % DMSO in Wasser verdunstet langsamer als reines Wasser und denaturiert die DNA, was für die spätere Hybridisierung vorteilhaft ist. DMSO-haltige Lösungen erzeugen jedoch auch größere Punkte als salzhaltige Puffer. Dies kann nachteilig sein, wenn viele tausend Sonden auf ein Microarray gedruckt werden sollen. Die Temperatur wurde bei ca. 20 °C und die relative Luftfeuchtigkeit bei 45 % konstant gehalten, um eine geringe Verdunstung der Sonden und eine homogene DNA-Verteilung der Spots zu gewährleisten. Trocknen die Sonden auf dem Objektträger zu schnell ein, entstehen häufig ringförmige Punkte, die im Zentrum wenig DNA enthalten.

Gedruckt wurde auf UltraGAPS Objektträger. Diese Objektträger besitzen eine homogene γ -Amino-Propyl-Silan-Beschichtung, welche die DNA kovalent bindet. Dazu wurden die Microarrays nach dem Trocknen der gedruckten Sonden zwei Stunden bei 80 °C inkubiert, um die DNA mit dem Objektträger zu verknüpfen. Vor Verwendung wurden die Microarrays gemäß den Herstellerangaben prähybridisiert.

Für den Vergleich von Methoden zur Amplifikation der Sonden wurde auf Poly-L-Lysin Objektträger gedruckt. Nach dem Druck wurden die Microarrays 2 Stunden bei 80 °C inkubiert, in 0,1 % SDS gewaschen, die DNA für 2 min bei 95 °C denaturiert und die Microarrays durch Zentrifugation bei 235 x g getrocknet.

2.2.2.3 Qualitätskontrolle der Microarrays

Um die Qualität der gedruckten Microarrays zu überprüfen, wurde jeweils das letzte Microarray eines Produktionsprozesses 15 Minuten mit degenerierten, Cy3-markierten Nonamer-Oligonukleotiden hybridisiert.

Hybridisierungsansatz (60 µl)

38,8 µl H₂O
12 µl 20x SSC
3 µl 1 M HEPES, pH 7,0
1,2 µl 10 % SDS
5 µl Nonamer [100 µM]

Anschließend wurde das Microarray mit Deckglas in 2 x SSC, 0,1 % SDS so lange leicht geschüttelt, bis das Deckglas abfiel. Dann wurde es für 10 s mit 0,1 x SSC gewaschen, 10 min bei 235x g zentrifugiert und gescannt. Die degenerierten Oligonukleotide können mit allen Sonden hybridisieren. Dadurch ist es möglich, die Qualität des Microarrays zu überprüfen.

2.2.3 Zellbiologische Methoden

2.2.3.1 Kultivierung von *D. discoideum* in Flüssigkultur

Der Stamm AX2 von *D. discoideum* kann in flüssigem Medium kultiviert werden. Es wurden 500 µl einer Suspension der Sporen dieses Stamms in 29,5 ml AX2-Medium gegeben und drei bis vier Tage bei 21 °C und 160 UpM angezogen. Ausgehend von dieser Vorkultur wurden Hauptkulturen mit einer Zelldichte von mindestens 2×10^4 Zellen/ml angeimpft. Die Kultivierung wurde bis zu einer Zelldichte von 5×10^6 Zellen/ml durchgeführt, da *D. discoideum* danach in die stationäre Phase übergeht.

2.2.3.2 Bestimmung der Zelldichte von *D. discoideum*

Die Zelldichte von Kulturen wurde mit einer Neubauer-Zählkammer nach Anleitung des Herstellers bestimmt.

2.2.3.3 Isolation von totaler RNA aus *D. discoideum*

Es wurden 1×10^8 bis 5×10^8 Zellen für die Isolation mit dem RNeasy Midi Kit verwendet. Abweichend vom Protokoll „Isolation of Cytoplasmic RNA from Animal Cells“ des

Herstellers wurden die Zellen mit 500 x g geerntet und zunächst zweimal mit Reinstwasser gewaschen.

2.2.3.4 Infektion von *D. discoideum* mit *Legionella*

Alle Infektionen wurden von Carina Wagner am Institut für Molekulare Infektionsbiologie, Universität Würzburg, durchgeführt. In 75 cm²-Zellkulturflaschen wurden zu 15 ml Kulturen von *D. discoideum* mit 1×10^6 Zellen/ml in einer 1:1 Mischung aus HL5-Medium und Soerensen-Puffer 150 µl einer Legionellen-Suspension von 1×10^9 Zellen/ml in Soerensen-Puffer gegeben. Die Zellen wurden für unterschiedliche Zeiten bei 24,5 °C koinkubiert und anschließend ihre RNA isoliert (Hägele *et al.*, 2000; Skriwan *et al.*, 2002).

HL5-Medium

14,3 g	Proteose Pepton Nr. 2 (Difco)
7,15 g	Hefeextrakt
15,4 g	Glukose
1,28 g	Na ₂ HPO ₄ · 2 H ₂ O
0,49 g	KH ₂ PO ₄

mit Wasser auf 1 l auffüllen

pH 7,5

autoklavieren

Soerensen-Puffer

2 g	KH ₂ PO ₄
0,36 g	Na ₂ HPO ₄ · 2 H ₂ O
mit Wasser auf 1 l auffüllen	
pH 6,0	
autoklavieren	

2.2.4 Molekularbiologische Methoden

2.2.4.1 Isolation von mRNA aus totaler RNA

Boten-RNA (mRNA) wurde aus totaler RNA mit dem Oligotex mRNA Midi Kit isoliert.

2.2.4.2 Bestimmung von DNA-Konzentrationen durch UV-Absorption

Um die Konzentration von DNA-Lösungen zu bestimmen und die Reinheit abzuschätzen, wurde die Absorption bei 260 nm und 280 nm gemessen. Eine Absorption von $A_{260} = 1$ entspricht bei doppelsträngiger DNA einer Konzentration von 50 µg/ml. Die Reinheit ergibt sich aus dem Quotienten A_{260} / A_{280} und sollte für reine DNA zwischen 1,8 und 2,0 liegen.

2.2.4.3 Bestimmung von DNA-Konzentrationen mit der PicoGreen Methode

DNA-Bestimmungen mit dem Fluoreszenz-Farbstoff PicoGreen wurden nach Anleitung des Herstellers durchgeführt.

2.2.4.4 Bestimmung von RNA-Konzentrationen durch UV-Absorption

Zur Bestimmung der Konzentration und Abschätzung der Reinheit von RNA-Lösungen wurde die Absorption bei 260 nm und 280 nm gemessen. Eine Absorption von $A_{260} = 1$ entspricht einer Konzentration von 40 µg/ml RNA. Die Reinheit ergibt sich aus dem Quotienten A_{260} / A_{280} und sollte für reine RNA zwischen 1,8 und 2,0 liegen.

2.2.4.5 Elektrophoretische Auftrennung von DNA in Agarosegelen

(Maniatis *et al.*, 1982)

Zur Auftrennung und Größenbestimmung von DNA-Fragmenten wurden 1,0 %ige Agarosegele in einer horizontalen Gelapparatur verwendet. Als Gel- und Elektrophorese-Puffer diente TBE-Puffer. Für die Elektrophorese wurden in der Regel 10 ng bis 20 µg DNA (in 10 bis 20 µl TE-Puffer) mit fünffach konzentriertem Probenauftragspuffer gemischt und in die Gelauftragstaschen pipettiert. Die Elektrophorese wurde bei 1 bis 5 V/cm durchgeführt. Die Gele wurden in einer 1 %igen Ethidiumbromid-Lösung 5 min gefärbt und anschließend dreimal für 10 min mit Wasser entfärbt. Die Gele wurden mit einem Geldokumentationssystem photographiert. Zur Größenbestimmung diente ein DNA-Längenstandard.

DNA-Probenpuffer (5 x):

40 % Saccharose
0,5 % SDS
0,25 % Bromphenolblau
in TE Puffer

2.2.4.6 Elektrophoretische Auftrennung von RNA in Agarosegelen

(Lehrach *et al.*, 1977)

Für die Elektrophorese wurden 1,2 %ige Agarosegele hergestellt. Dazu wurden 0,6 g Agarose und 5 ml 10 x RNA-Gelpuffer zu 37 ml Wasser gegeben, aufgeköcht und auf 60 °C abgekühlt. Anschließend wurden 8 ml einer 37 %igen Formaldehydlösung zugegeben und das Gel in eine Gelwanne gegossen. Die Gelapparatur wurde mit RNA-Elektrophoresepuffer gefüllt und die RNA-Proben mit 10 µl –RNA-Probenpuffer versetzt und für 10 min auf 65 °C erhitzt. Dann wurden die Proben auf Eis abgekühlt und mit 1 µl RNA-Farbmarker versetzt und auf das Gel geladen. Die Elektrophorese wurde bei 1 bis 5 V/cm durchgeführt und die Gele mit einem Geldokumentationssystem photographiert.

RNA-Gelpuffer (10 x):

200 mM MOPS/NaOH
50 mM Natriumacetat
10 mM EDTA
pH 8,0
autoklavieren

RNA-Elektrophoresepuffer (10 x):

200 mM MOPS/NaOH
50 mM Natriumacetat
10 mM EDTA
pH 7,0
autoklavieren

RNA-Probenpuffer

50 % Formamid
 6 % Formaldehyd
 in 1 x RNA-Gelpuffer
 10 µg/ml Ethidiumbromid

RNA-Farbmarker

50 % Saccharose
 0,25 % Bromphenolblau
 in DEPC-Wasser

2.2.4.9 Northernblot-Analyse

Elektrophoretisch getrennte RNA wurde, wie von Maniatis *et al.* (1982) beschrieben, auf eine Nylonmembran transferiert. Die Membran wurde 5 min mit 2 x SSC gewaschen und zwischen zwei Whatmanpapieren für 2 h bei 80 °C gelagert, um die RNA mit der Membran zu vernetzen. Anschließend wurde die Membran mit 15 ml Hybridisierungspuffer und 150 µl Fisch DNA [10 µg/ml] für 3 h bei 37 °C prähybridisiert.

Radioaktiv markierte Sonden wurden ausgehend von 25 ng Sonden-DNA mit dem Prime-It II Random Primer Labeling Kit hergestellt. Nicht eingebaute, radioaktive Nukleotide wurden durch Zentrifugation für 1 min bei 800 x g über eine 1 ml Sephadex G50 Säule mit TE entfernt. Anschließend wurde die Sonde für 5 min auf 95 °C erhitzt, abgekühlt und in 10 ml Hybridisierungspuffer über Nacht bei 37 °C im Hybridisierungsöfen mit der Membran hybridisiert (Mehdy *et al.*, 1983). Gewaschen wurde die Membran im Hybridisierungsöfen bei 37 °C in folgenden Schritten:

1. 1 h mit 30 ml Waschpuffer
2. 20 min mit 40 ml 2 x SSC mit 0,1 % SDS
3. 20 min mit 40 ml 0,2 x SSC mit 0,1 % SDS
4. 20 min mit 40 ml 2 x SSC

Die Nylonmembran wurde auf Whatmanpapier getrocknet, in Cellophanfolie eingewickelt und zusammen mit einem Röntgenfilm in einer Expositions-kammer eine Stunde bis mehrere Tage bei – 70 °C gelagert. Anschließend wurde der Röntgenfilm entwickelt.

Waschpuffer

1,2 M Phosphatpuffer pH 6,8
 2 mM EDTA (pH 7,2)
 50 % Formamid
 1 % Na-Laurylsarcosinat
 0,2 % SDS
 in 2 x SSC

Hybridisierungspuffer

Waschpuffer mit
 4 x Denhardts Reagenz

Denhardts Reagenz (100 x)

2 % Ficoll
 2 % Polyvinylpyrrolidon
 2 % BSA

2.2.4.7 Präzipitation von DNA und RNA

Lösungen von DNA und RNA wurden mit einem Zehntel ihres Volumens einer 3 M Natriumacetat-Lösung (pH 4,8) und zwei Volumen Ethanol versetzt und für 1 Stunde oder länger bei -20 °C gelagert. Anschließend wurde 20 min bei 16 100 x g zentrifugiert, der Überstand verworfen und der Niederschlag mit 0,5 bis 1 ml 70 %iger Ethanollösung gewaschen. Nach einer weiteren Zentrifugation für 10 min wurde der Überstand abgesaugt und der Niederschlag getrocknet.

2.2.4.8 Reinigung von PCR-Produkten mit QIAvac

Die Reinigung von PCR-Produkten im 96-Loch-Format wurde mit dem QIAvac PCR Purification Kit und dem erforderlichen Zubehör nach Anleitung des Herstellers durchgeführt.

2.2.4.10 Quantitative PCR

Quantitative PCR Reaktionen wurden mit dem QuantiTect SYBR Green PCR Kit und der Quantitative PCR-Maschine Opticon 2 durchgeführt. Dazu wurden für die Sonden des Microarray Oligonukleotid-Paare ausgewählt, die mit der cDNA-Matrize von *D. discoideum* Produkte mit einer Länge von 200 bis 300 bp erzeugten.

Als positive Kontrolle wurde das Gen H7 verwendet, das während des Wachstums und des Entwicklungszykluses von *D. discoideum* stets gleich stark exprimiert wird (Zinda und Singleton, 1998). Um eine Eichkurve zu erstellen wurde das murine Gen Annexin 7 verwendet, das von einem pGEMTeasy Plasmid (Promega) mit der gesamten cDNA-Sequenz von Annexin 7 amplifiziert wurde. Das Plasmid wurde in den Mengen 10 pg, 100 pg, 1 ng und 10 ng eingesetzt. Die Auswertung erfolgte mit dem Programm Opticon Monitor in Version 2.02.

Reaktionsansatz (20 µl)

10 µl	2 x QuantiTect SYBR Green PCR Master Mix
0,5 µl	Vorwärts-Oligonukleotid (20 µM)
0,5 µl	Rückwärts-Oligonukleotid (20 µM)
1 µl	cDNA-Matrize
8 µl	Wasser

PCR-Programm

1. 5 min 95 °C Denaturierung der DNA
2. je 60 mal 30 s 94 °C Denaturierung der DNA
30 s 58 °C Hybridisierung der Oligonukleotide
60 s 68 °C DNA-Synthese
Fluoreszenz-Messung
3. 1 s 50 °C
4. Schmelzkurve von 60 bis 90 °C mit 5 s Pause und Messung alle 1 °C

2.2.5 Microarray-Methoden

2.2.5.1 Zugabe von SpotReport RNA zur RNA von *D. discoideum*

Das SpotReport – 10 Array Validation System dient der Überprüfung des Microarray-Systems durch Zugabe von bekannten Mengen *Arabidopsis* RNA zur RNA von *D. discoideum*. Um exakte Messwerte zu erhalten, wurden aus den zehn verschiedenen RNA-Fragmenten zwei Stocklösungen erstellt, von der jeweils 10 µl zu 20 µg RNA des Experiments bzw. der Kontrolle von *D. discoideum* gegeben wurden. Anschließend wurde die RNA präzipitiert und in 12 µl RNase-freiem Wasser gelöst.

Gen	Mischung A [pg/10 µl]	Mischung B [pg/10 µl]	Quotient A/B	Gesamtmenge [pg]
Cab	300	150	2	450
RCA	250	250	1	500
rbcL	250	250	1	500
LTP4	5	5	1	10
LTP6	2	2	1	4
XCP2	1	1	1	2
RCP1	300	150	2	450
NAC1	400	40	10	440
TIM	375	75	5	450
PRKase	375	75	5	450

2.2.5.2 Reverse Transkription von RNA und Fluoreszenzmarkierung mit FairPlay

Die reverse Transkription und die Markierung der cDNA mit den Fluoreszenz-Farbstoffen Cy3 und Cy5 wurde nach Vorschrift des Herstellers durchgeführt. Die Reinigung der markierten cDNA wurde nach Schritt 21 modifiziert. Die Säule wurde noch zweimal mit 10 mM TrisCl (pH 8,5) eluiert und die DNA-Konzentration spektroskopisch bestimmt. Anschließend wurden Cy3- und Cy5-markierte cDNA vereinigt und präzipitiert.

2.2.5.3 Prähybridisierung der Microarrays

Microarrays wurden nach dem Protokoll für UltraGAPS-Objektträger prähybridisiert.

2.2.5.4 Hybridisierung der Microarrays

Die markierte cDNA wurde in 65 µl Hybridisierungspuffer gelöst und für 10 min auf 80 °C erhitzt. Nach Abkühlung auf 37 °C wurden sie, auf ein Microarray pipettiert und luftblasenfrei mit einem Deckglas abgedeckt. Das Microarray wurde in eine Hybridisierungskammer gelegt, in deren Reservoiren sich je 10 µl 3 x SSC befanden, die Kammern verschlossen und über Nacht in einem Wasserbad bei 37 °C inkubiert.

Dann wurden die Microarrays aus der Hybridisierungskammer entnommen und mit folgenden Schritten gewaschen:

1. 2 x SSC mit 0,1 % SDS bis das Deckglas abfiel
2. 2 x SSC mit 0,1 % SDS für 5 min
3. 0,1 x SSC mit 0,1 % SDS für 5 min
4. 0,1 x SSC für 5 s
5. 0,1 x SSC für 5 s
6. 0,1 x SSC für 5 s
7. 0,1 x SSC für 5 s
8. 0,1 x SSC für 5 s
9. 0,01 x SSC für 5 s

Anschließend wurden die Microarrays sofort für 10 min bei 235 x g zentrifugiert und am selben Tag gescannt.

Hybridisierungslösung (65 µl)

- 62,4 µl Hybridisierungspuffer
- 1,3 µl Fisch DNA [10 mg/ml]
- 1,3 µl Oligo dA [100 µM]

2.2.5.5 Scannen von Microarrays

Microarrays wurden mit ScanArray Express nach Anleitung gescannt. Es wurden stets mehrere Scans mit unterschiedlichen Anregungs- und Verstärkungsenergien durchgeführt. Ein Scan mit hoher Anregung diente dazu, von möglichst vielen Sondenpunkten Signale zu erhalten, wobei in Kauf genommen wurde, dass einige Punkte gesättigte Signale hatten. Ein zweiter Scan wurde mit niedriger Anregung durchgeführt. Dabei wurde die Anregungsenergie so eingestellt, dass die Sondenpunkte mit den stärksten Signalen geringfügig unterhalb der Sättigung lagen. Dieser Scan lieferte die Signale für Sondenpunkte, die im ersten Scan gesättigte Signale hatten.

Die Verstärkung wurde möglichst auf 70 bis 80 % eingestellt, um ein günstiges Signal-Rausch-Verhältnis zu erreichen. Die Anregung wurde so gewählt, dass der Cy3-Kanal und der Cy5-Kanal ähnliche Signalintensitäten hatten. Dadurch wurden die Kanäle teilweise normalisiert, was die nötigen Anpassungen bei der Normalisierung mit dem Programm R verringerte.

2.2.5.6 Quantifizieren von Microarrays

Die Microarray-Bilder der Scans mit starker und schwacher Anregung wurden mit ScanArray Express nach Anleitung des Herstellers quantifiziert. Dabei wurde die Quantifizierungsmethode *Adaptive circle* mit den Standardparametern verwendet.

Nach der Quantifizierung wurde die Punktfindung aller Sondenpunkte manuell überprüft und gegebenenfalls die Position oder der Durchmesser des quantifizierten Kreises an den

Sondenpunkt angepasst. Bei der Quantifizierung des Scans mit niedriger Anregung wurden nur die Kreise der Sondenpunkte angepasst, die im Scan mit starker Anregung gesättigt waren. Die quantifizierten Intensitäten der übrigen Punkte wurden nicht angepasst, da ihre Daten verworfen wurden.

2.2.6 *In silico* Methoden

2.2.6.1 Identifikation differentiell exprimierter Gene mit Significance Analysis of Microarrays (SAM)

Für die Identifikation der differentiell exprimierten Gene eines Microarray-Experiments wurde das Programm Significance Analysis of Microarrays (SAM) in der Version 1.21 nach Anleitung verwendet (Tusher *et al.*, 2001). Bei den Optionen, die SAM bietet, wurden größtenteils die Voreinstellungen beibehalten. Die Anzahl der Permutation wurde jedoch stets auf 1000 gesetzt und als *response type* wurde *one class* verwendet.

In SAM kann die gewünschte Stringenz der Berechnung mit dem sogenannten *delta*-Wert gewählt werden. Da SAM für jeden *delta*-Wert nicht nur die differentiell exprimierten Gene ermittelt, sondern auch die Anzahl der erwarteten falsch positiven Gene, wurde als Schwellenwert für *delta* der Wert gewählt, bei dem der *90th percentile* Wert der *false discovery rate* (FDR) minimal war.

2.2.6.2 Clusteranalyse in GeneSpring 6.1

Für Clusteranalyse in GeneSpring 6.1 wurden die nicht-logarithmischen Faktoren der Induktion und Repression sowie die Standardabweichungen verwendet. Die Daten wurden nach Anleitung des Herstellers importiert und ein neues *experiment* angelegt. Die Normalisierung in GeneSpring wurde deaktiviert, da die Daten bereits in dem Programm R normalisiert worden waren. Auch das *error-model* wurde nicht verwendet. Das *clustering* wurde mit den Methoden *Gene Tree* und *Condition Tree* durchgeführt. Bei *Gene Tree* werden die einzelnen Gene anhand ähnlicher Expressionsmuster gruppiert, während bei *Condition Tree* ganze Microarrays aufgrund der Ähnlichkeit aller ihrer Sonden gruppiert werden (siehe Abschnitt 3.4.3). Es wurden stets mehrere Korrelationsmethoden wie Standard-, Pearson- und Spearman- *correlation* für das *clustering* verwendet und die Ergebnisse verglichen, um die Robustheit der Clusteranalyse abschätzen zu können.

2.2.6.3 BLAST-Suchen

Für den Sequenzvergleich der cDNA-Sequenzen mit den bekannten Proteinsequenzen von *D. discoideum* und anderen Organismen wurden das Programm BLAST (Altschul *et al.*, 1990)

des NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) oder des EBI (<http://www.ebi.ac.uk/blast2/index.html?UniProt>) verwendet. Es gibt mehrere Möglichkeiten der BLAST-Suche. Die Funktion BLASTN vergleicht DNA-Sequenzen, BLASTP Proteinsequenzen und BLASTX translatiert eine DNA-Sequenz in alle sechs Leserahmen und vergleicht diese mit Proteinsequenzen.

3 Ergebnisse

3.1 Herstellung des *D. discoideum* DNA-Microarrays

3.1.1 Vergleich von Methoden zur Amplifikation der Sonden für die Herstellung von DNA-Microarrays

Die Herstellung von DNA-Microarrays für *D. discoideum*, die einen großen Teil des Genoms repräsentieren, erfordert die Produktion mehrerer tausend Sonden. Die in dieser Arbeit verwendeten Microarrays tragen 5423 Sonden des japanischen cDNA Projektes, 450 Sonden für publizierte Gene und 33 Kontrollen. Das Ausgangsmaterial für die Sondenproduktion waren im Fall der japanischen cDNA Plasmide, welche die jeweilige cDNA enthielten, und im Fall der publizierten Gene und einiger Kontrollen genomische DNA und genspezifische Oligonukleotide. Die Amplifikation der Sonden erfolgte mittels PCR und erforderte erhebliche Mittel und Zeit.

Cheung *et al.* (1999) berichten 1999, dass eine Sonden-Konzentration von 500 ng/μl nötig ist, um Microarrays zu drucken. Eine so hohe Konzentration bei einem Mindestvolumen von 10 μl wäre jedoch nur durch mehrere PCR-Ansätze je Sonde zu erreichen, was den Aufwand vervielfachen würde. Um Kosten und Aufwand so gering wie möglich zu halten, wurde die minimale Anzahl von PCR-Amplifikationen ermittelt, die benötigt wird, um eine ausreichende Sondenmenge für das Drucken von Microarrays herzustellen. Außerdem wurde ermittelt, welches Reinigungsverfahren für die PCR-Ansätze bei minimalem Aufwand und Kosten einen genügenden Reinheitsgrad der Sonden für das Drucken auf Microarrays liefert. Hegde *et al.* (2000) empfehlen die Verwendung einer Filtermethode, um Nukleotide und Oligonukleotide zu entfernen. Daher wurde die Verwendung eines kommerziellen PCR-Reinigungssystems mit der Ethanolpräzipitation verglichen.

Es wurden folgende drei Methoden zur Amplifikation und Reinigung von 90 Sonden für *D. discoideum* Gene eingesetzt:

1. Amplifikation mit einem PCR-Ansatz und Reinigung mittels Ethanolpräzipitation
2. Amplifikation mit fünf PCR-Ansätzen je Sonde und Reinigung durch Ethanolpräzipitation der vereinigten Ansätze
3. Amplifikation mit einem PCR-Ansatz und Reinigung mit dem QIAvac PCR Purification Kit.

Nach den PCR-Reaktionen wurde die korrekte Amplifikation der Sonden mittels Gelelektrophorese überprüft. Die mittlere Länge der Sonden betrug 509 bp. Um die maximale Konzentration zu erreichen, wurden die bereits gefällten Sonden der Methoden eins und zwei direkt, die QIAvac-gereinigten Sonden nach Ethanolpräzipitation, in 10 µl TE mit 50 % DMSO (10 mM Tris-HCl, pH 8,0, 1 mM EDTA, 50 % DMSO) gelöst. Die gereinigten Sonden wurden in Duplikaten auf Objektträger gedruckt und zur Hybridisierung mit markierten cDNA-Proben verwendet. Die Microarrays wurden je 10 min mit 1 x SSC, 0,1 % SDS und 0,2 x SSC gewaschen, kurz in Reinstwasser und Ethanol getaucht und durch Zentrifugation bei 235 x g getrocknet. Bilder der Microarrays wurden für die Cy3- und Cy5-Fluoreszenz mit ScanArray aufgenommen und mit QuantArray quantifiziert. Für drei Microarrays wurde die totale Signalintensität aller 90 Doppelpunkte einer Amplifikations- und Reinigungsmethode sowie die Standardabweichung berechnet

Der Vergleich der totalen Signalintensität aller 90 Doppelpunkte einer Amplifikations- und Reinigungsmethode zeigt, dass die Unterschiede bei Cy3 wie auch bei Cy5 gering sind. Zwischen der einfachen und fünffachen PCR-Amplifikation sind ebenfalls keine signifikanten Unterschiede feststellbar. Eine Erhöhung der Sondenkonzentration führt also nicht zu höheren Signalintensitäten. Die mit QIAvac gereinigten Sonden liefern für Cy3 eine 18,4 % und für Cy5 eine 18,9 % höhere Signalintensität (Abb. 6).

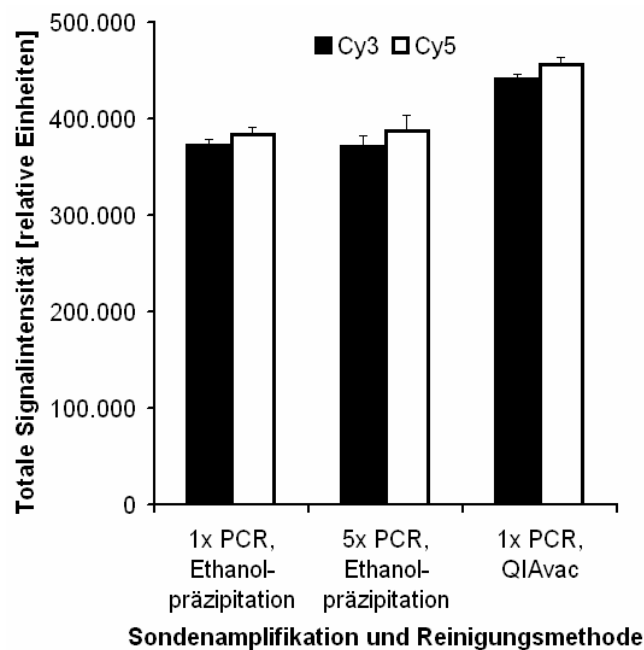


Abb. 6: Totale Signalintensität. Totale Signalintensität der 90 Sonden im Cy3- und Cy5-Kanal für die drei verschiedenen Amplifikations- und Reinigungsmethoden. Die totale Intensität wurde durch Summation der Einzelintensitäten der 90 Doppelpunkte berechnet. Dargestellt sind die Mittelwerte und Standardabweichungen von drei Microarrays.

Da die totale Signalintensität hauptsächlich von den Sonden bestimmt wird, die eine hohe Intensität zeigen, wurden die Signalintensitäten über das gesamte Intensitätsspektrum verglichen (Abb. 7). Die Signalintensitäten sind über annähernd zwei Größenordnungen fast gleich. Die Intensitäten der einfach PCR-amplifizierten, ethanolpräzipitierten und mittels QIAvac gereinigten Sonden sind über weite Teile des Graphen annähernd gleich. Die fünffach-amplifizierten Sonden weisen dagegen bei niedrigen Signalintensitäten etwas höhere Werte auf als die Sonden der anderen beiden Methoden. Die in Abb. 6 gezeigte, höhere Totalintensität der QIAvac gereinigten Sonden kommt durch wenige Sonden mit sehr hoher Intensität zustande (Abb. 7). Sie könnte eventuell durch das Entfernen von Nukleotiden und Oligonukleotiden erklärt werden. Diese fallen bei der Ethanolpräzipitation mit aus und setzen möglicherweise die DNA-Bindekapazität der Microarray-Oberfläche herab. Ein Effekt der drei Methoden auf die Morphologie der Sondenpunkte konnte nicht festgestellt werden. Die von QuantArray berechnete Uniformität der Punkte betrug im Mittel jeweils 96 %.

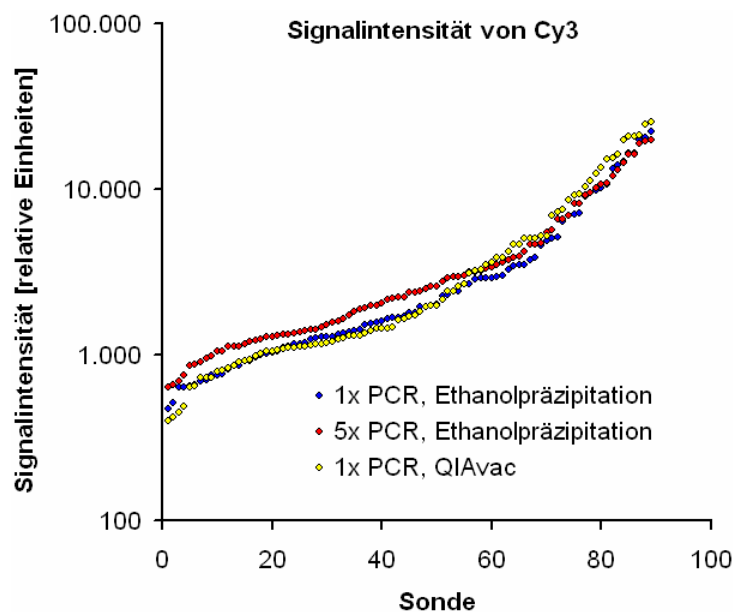


Abb. 7: Cy3-Signalintensitäten der 90 Sonden. Cy3-Signalintensitäten der 90 Sonden für die drei verschiedenen Amplifikations- und Reinigungsmethoden. Abgebildet sind die mittleren Signalintensitäten der 90 Sonden der drei Microarrays, sortiert nach Intensität.

3.1.2 Amplifikation der Sonden

Die Sonden für das *D. discoideum* Microarray wurden, wie in Material und Methoden (2.2.1) beschrieben, amplifiziert und die korrekte Amplifikation aller Sonden mittels Gelelektrophorese überprüft. Abb. 8 zeigt dies am Beispiel der dritten 96-Loch-Mikrotiterplatte. Jede Sonde ist paarweise nach PCR und nach Konzentrierung mittels Ethanolpräzipitation aufgetragen worden. Wie man sieht, wurden bis auf zwei Sonden stets

PCR-Produkte erhalten und nur bei wenigen Sonden ist ein geringer Verlust bei der Konzentrierung aufgetreten. Durch die Gelelektrophorese war auch ein Vergleich der erwarteten und beobachteten Größe möglich und es konnte überprüft werden, ob ausschließlich Produkte einer Länge gebildet wurden.

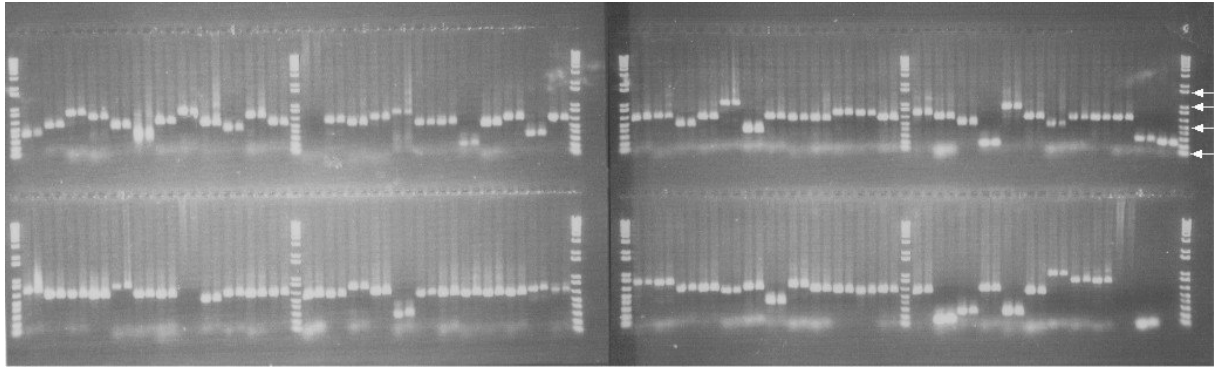


Abb. 8: DNA Gelelektrophorese der amplifizierten Sonden der dritten 96-Loch-Mikrotiterplatte. Die PCR-Produkte nach Auftrennung in einem 1 %igen Agarosegel in TBE-Puffer sind gezeigt. Das links abgebildete Gel enthält die ersten 48 Sonden (Nr. 181 bis 228) der Mikrotiterplatte; das rechte die verbleibenden 42 (Nr. 229 bis 270) sowie drei positive und drei negative Kontrollen. Aufgetragen sind die Sonden jeweils zweimal. Die linke Spur enthält 10 μ l des 100 μ l umfassenden PCR-Ansatzes, die rechte 1 μ l, der nach Ethanolpräzipitation in 11 μ l Microarray-Drucklösung gelösten Sonde. Ganz rechts unten befinden sich die sechs Spuren der drei negativen Kontrollen, links daneben die sechs Spuren der positiven Kontrollen. Es wurde die 1 kb Plus DNA Ladder als Längenstandard verwendet (1., 26. und 51. Spur). Für den Längenstandard oben rechts sind vier Fragmente markiert. Ihre Größen betragen 1650, 1000, 500 und 100 bp.

3.1.3 Transfer der Sonden aus 96-Loch- in 384-Loch-Mikrotiterplatten

Die Sonden wurden aus Platten im 96-Loch-Format, in denen sie amplifiziert und durch Fällung konzentriert worden waren, in 384-Loch-Platten überführt, um auf Microarrays gedruckt zu werden. Dazu wurde ein Proteinerer dp Pipettier-Roboter verwendet. Um sicher zustellen, dass bei diesem Transfer keine Kontamination der Sonden untereinander stattfindet, wurde überprüft, ob die Waschprozedur des Proteinerer den Übertrag von DNA verhindert.

Dazu wurde eine 96-Loch-Mikrotiterplatte in vier Spalten mit je 10 μ l Microarray-Drucklösung (25 % DMSO) befüllt, die als Leerwert dienten. Weitere acht Spalten wurden abwechselnd mit Puffer mit und ohne 100 ng/ μ l Fisch DNA befüllt, um zu bestimmen ob DNA von dem DNA-haltigen Puffer auf den DNA-freien Puffer übertragen wird. Der Proteinerer transferierte dann den Inhalt der 96-Loch-Mikrotiterplatte in eine 384-Loch-Mikrotiterplatte, wobei zwischen den Pipettiervorgängen im ersten Versuch dreimal und im zweiten Versuch fünfmal mit Reinstwasser gewaschen wurde. Anschließend wurden die transferierten acht mal 10 μ l einer Spalte vereinigt und für eine DNA-

Konzentrationsbestimmung mit PicoGreen verwendet. Die Nachweisgrenze der PicoGreen Methode liegt bei 25 pg/ml. Da zur Messung 80 µl verwendet wurden, die im Reaktionsansatz auf 500 µl mit Puffer aufgefüllt werden mußten, liegt die tatsächliche Nachweisgrenze bei 160 pg/ml, was 1,6 ppm der eingesetzten 100 ng/µl Fisch DNA entspricht.

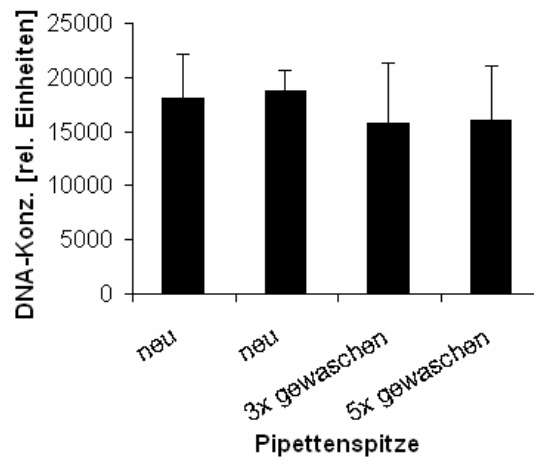


Abb. 9: Überprüfung der Pipettenspitzen-Reinigung des Proteineer. Relative DNA-Konzentration der Microarray-Drucklösung vor (Pipettenspitze neu) und nach (Pipettenspitze gewaschen) Kontakt der Pipettenspitzen mit DNA-haltigem Puffer. Dargestellt sind die Mittelwerte und Standardabweichungen von vier Messungen. Zwischen den Pipettiervorgängen wurde dreimal und fünfmal mit Reinstwasser gewaschen. Die relative DNA-Konzentration der Drucklösung mit 100 ng/µl Fisch DNA betrug $3,3 \times 10^6$ (nicht dargestellt).

Die Messwerte stimmen etwa mit den Leerwerten überein. In der pipettierten Drucklösung konnte also sowohl bei fünfmaligem, als auch bei dreimaligem Waschen der Pipettenspitzen keine DNA nachgewiesen werden (Abb. 9).

3.1.4 Druck der Microarrays

Auch beim Druck der Microarrays mußte überprüft werden, ob es zu einer Kontamination der Sonden untereinander kommt. Dazu wurden die Signalintensitäten der negativen Kontrolle, der Drucklösung, vor und nach dem Drucken einer positiven Kontrolle, der Sonde für das Aktin-Gen, verglichen. Zunächst wurde mit Ultraschall-gereinigten Microspot 2500-Stiften dreimal die negative Kontrolle gedruckt, anschließend die positive Kontrolle und dann wieder dreimal die negative Kontrolle. Zwischen jedem Druckvorgang wurden die Stifte zweimal mit der Waschprozedur des MicroGrid II Druckers gereinigt. Nach der Post-Prozessierung wurde das Microarray mit Cy3- und Cy5-markierter cDNA hybridisiert. In Abb. 10 sieht man die Mediane der quantifizierten und normalisierten Signalintensitäten der drei Druckvorgänge vor und nach der positiven Kontrolle. Die Intensitäten sind etwa gleich hoch.

Die Intensitäten der Druckvorgänge fünf, sechs und sieben sind nicht höher als die Vorgänge eins, zwei und drei, welche der Hintergrundintensität des Microarrays entsprechen. Mit der verwendeten Waschprozedur findet also kein Übertrag von Sondenmaterial statt.

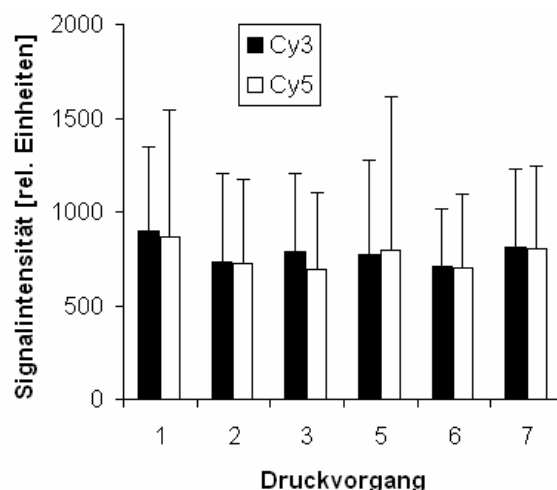


Abb. 10: Überprüfung der Drucknadel-Reinigung des MicroGrid II. Mediane Signalintensitäten und Standardabweichungen von je 32 Sondenpunkten eines Druckvorgangs nach Hybridisierung des Microarrays mit Cy3- und Cy5-markierter cDNA, die mit RNA aus vegetativen *D. discoideum* Zellen hergestellt wurde. In den Druckvorgängen eins, zwei und drei wurde die Negativkontrolle mit gereinigten Stiften gedruckt; in Druckvorgang vier die Positivkontrolle (Signalintensität nicht gezeigt) und anschließend erneut die Negativkontrolle in fünf, sechs und sieben.

3.1.5 Das *D. discoideum* DNA-Microarray

Das *D. discoideum* DNA-Microarray, besteht aus 5873 genspezifischen Sonden und 33 Kontrollen. Die Sonden sind jeweils zweimal nebeneinander gedruckt, während die Kontrollen häufiger vorkommen. Die Positivkontrollen sind je 24-mal, die Negativkontrollen sechs-, 24- oder 286-mal und die SpotReport Kontrollen je 96-mal gedruckt. Insgesamt besteht das Microarray aus 14 620 Punkten (Tabelle 1). Aufgebaut ist das Microarray aus drei Blöcken mit je 16 Unter-Microarrays, die wiederum aus 18 x 18 Punkten bestehen. Die Blöcke werden nacheinander von einem 16 Stifte umfassenden Druckkopf und jedes Unter-Microarray folglich von einem einzelnen Stift gedruckt (Abb. 11).

Tabelle 1: Sonden des *D. discoideum* Microarrays

	Sonden		Punkte	
	Anzahl	Prozent	Anzahl	Prozent
publizierte Gene	450	8 %	900	6 %
cDNAs	5423	92 %	10846	74 %
Kontrollen	33	1 %	2874	20 %
Summe	5906		14620	

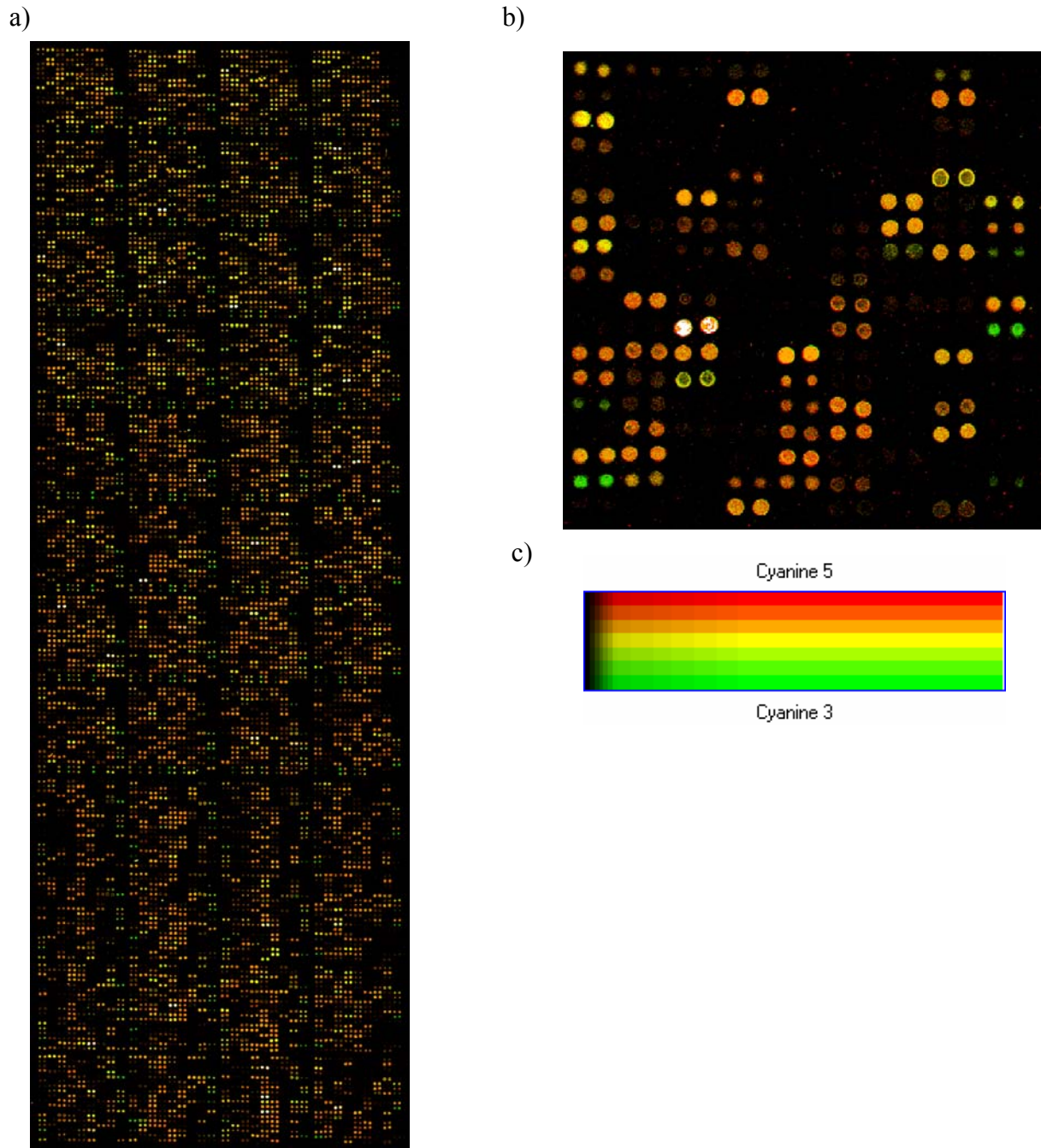


Abb. 11: Das *D. discoideum* Microarray. a) Ein mit Cy3- und Cy5-markierter cDNA hybridisiertes Microarray nach dem Scan. b) Vergrößerung des unteren rechten Unter-Microarrays. Da es sich um Rohwerte handelt und die Intensitäten nicht normalisiert sind, entspricht das Verhältnis von Cy3- zu Cy5-Intensität möglicherweise nicht dem der cDNA-Konzentrationen. c) Falschfarben-Spektrum. Die Farben in a und b kommen durch die Überlagerung des roten Falschfarben-Spektrums für Cy3 mit dem grünen Spektrum für Cy5 zustande. Je heller die Farben sind, desto höher ist die gemessene Fluoreszenz-Intensität. Weiße Punkte zeigen Sonden an, deren Signal in der Sättigung liegt.

3.2 Quantifizierung der gemessenen Signalintensitäten

Die Bilder der Fluoreszenz-Signale von Cy3 und Cy5 wurden mit dem Scan- und Quantifizierungsprogramm ScanArray Express quantifiziert. Dabei wurde die Fläche jedes Sondenpunktes automatisch mit einem Kreis markiert, dessen Radius variabel war und sich an den Punkt anpasste (Abb. 12).

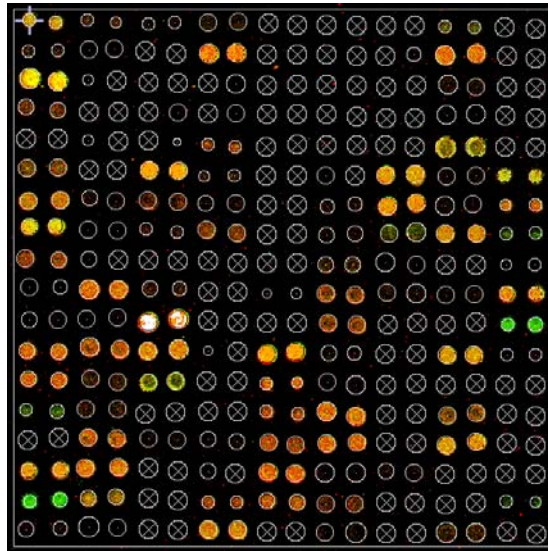


Abb. 12: Quantifizierung der Microarray-Bilder. Dargestellt ist dasselbe Unter-Microarray, wie in Abb. 11 b). Kreise markieren die Flächen der Sondenpunkte. Leere Kreise zeigen gefundene Punkte an; Kreise mit Kreuzen zeigen, dass kein Punkt gefunden wurde. Letzteres tritt dann auf, wenn kein Fluoreszenz-Signal vorliegt.

Störend für die Quantifizierung ist, dass die Microarray-Bilder keine perfekten Bilder sind, wo nur die Sondenpunkte leuchten und die Intensitäten der Punkte homogen sind. Häufig ist eine geringe Fluoreszenz der unbedruckten Microarray-Oberfläche zu beobachten, es kommen lokale Verunreinigungen vor und die Intensität der Sondenpunkte ist inhomogen. Dabei ist die Hintergrundfluoreszenz auf dem gesamten Microarray nicht immer an allen Orten gleich groß (Hegde *et al.*, 2000) und der Hintergrund des unbedruckten Microarray entspricht nicht dem Hintergrund eines Sondenpunktes (Martinez *et al.*, 2003).

Bei Microarrays mit stellenweise erhöhtem Hintergrund, der möglicherweise durch unvollständig gebundenes Sondenmaterial und damit einzelnen „ausgebluteten“ Punkten, oder auch durch ungebundenen Fluoreszenz-Farbstoff verursacht sein könnte, fiel auf, dass dieser Hintergrund nicht die Intensität der Punkte überlagert. Vielmehr sieht es so aus, als ob die DNA der Sondenpunkte das Binden des Hintergrundmaterials verhindert (Abb. 13).

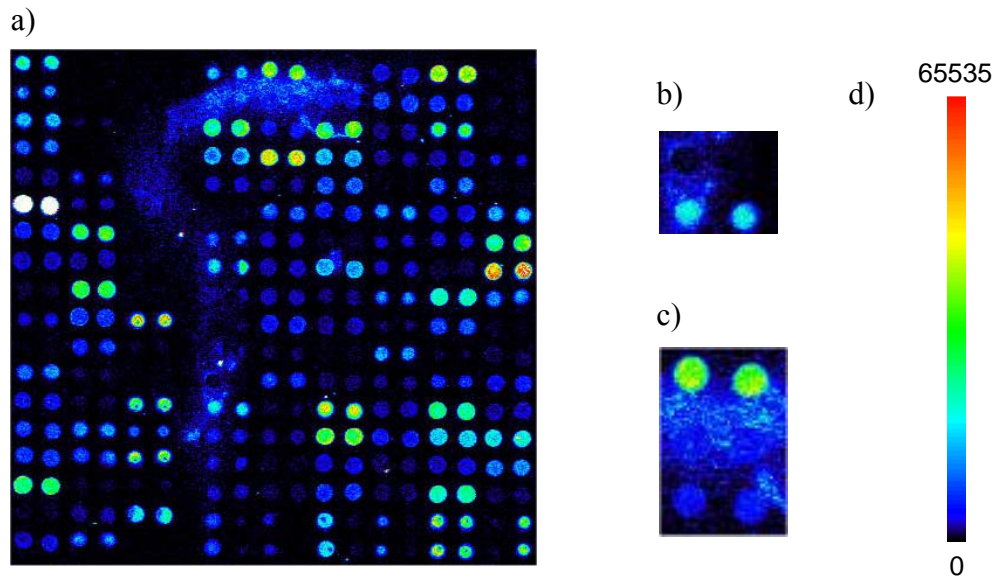


Abb. 13: Hintergrund in Microarray-Bildern. a) Ausschnitt eines Microarray-Bildes des Cy3-Kanals in Falschfarben-Darstellung, der eine „Wolke“ lokalen Hintergrunds zeigt. b) Ausschnittsvergrößerung von zwei Sondenpunkt-Paaren (Reihe 12, Spalte 7). c) Wie b) (Reihe 1, Spalte 9). d) Falschfarben-Spektrum. Die Intensitäten werden von ScanArray Express mit 16 bit Auflösung (0 bis 65535) quantifiziert.

In Abb. 13 b) sieht man, dass der Sondenpunkt oben links etwa genauso wenig Fluoreszenz zeigt, wie sein Nachbarpunkt (oben rechts). Die beiden unteren Punkte haben ebenfalls eine ähnliche Intensität, obwohl der linke Punkt von Hintergrund umgeben ist. Auch die beiden Punkte in der mittleren Reihe von c) heben sich vor dem Hintergrund mit höherer Intensität ab.

Da der Hintergrund nicht additiv ist, wurde der lokale Hintergrund, den ScanArray Express für jeden Punkt anhand der Umgebung des Punktes berechnet, nicht von der Signalintensität des Punktes subtrahiert. Die Quantifizierung wurde bei allen Microarrays manuell kontrolliert. In Fällen, wo die Sondenpunkte nicht von lokalem Hintergrund unterschieden werden konnten, wurden die Punkte von der weiteren Analyse ausgeschlossen.

Die Inhomogenität der Sonden signale wurde durch die Verwendung des Median kompensiert. Der Median ist ein robuster Mittelwert (Kesel *et al.*, 1999).

3.3 Entwicklung eines automatisierten Auswerteverfahrens

Microarray-Experimente liefern meist sehr große Datensätze. Ein typisches Experiment mit dem *D. discoideum* Microarray, das 14 620 Sondenpunkte trägt, erzeugt bei drei unabhängigen RNA-Isolationen und der Verwendung von *dye-swaps* auf insgesamt sechs Microarrays jeweils 87 720 Intensitätswerte für den Cy3- und den Cy5-Kanal. Da eine

manuelle Auswertung nicht möglich ist, wurde ein automatisiertes Auswerteverfahren entwickelt (Abb. 14).

Ausgehend von den quantifizierten Signalintensitäten der einzelnen Sondenpunkte, die von dem Programm ScanArray Express bereitgestellt werden, sollten primär die differentiell exprimierten Gene identifiziert werden. Dazu müssen zuvor die SpotReport- und einige andere Kontrollen sowie die in ScanArray Express nicht gefundenen oder als schlecht markierten Punkte heraus gefiltert werden. Dann müssen der Cy3- und der Cy5-Kanal normalisiert werden. Für die Analyse einer Zeitreihe der Infektion von *D. discoideum* mit *L. pneumophila* sollte auch das Clustering von Genen über die verschiedenen Zeitpunkte möglich sein.

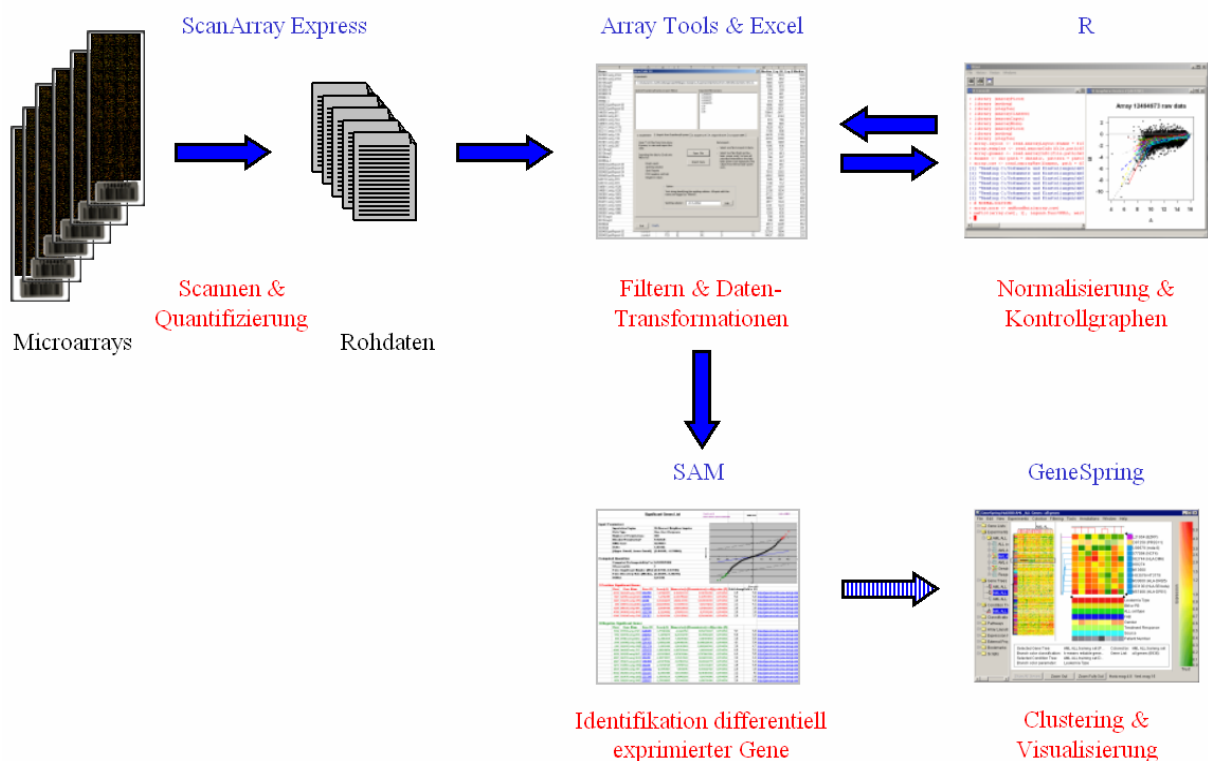


Abb. 14: Das automatisierte Auswerteverfahren für die *D. discoideum* Microarrays. Die Rohdaten der Microarrays, also die Signalintensitäten der Sondenpunkte, werden mit dem Programm ScanArray Express erzeugt. Array Tools, das *Add-in* für Excel dient dazu, die Daten so zu filtern und zu transformieren, das sie von dem Statistikprogramm R normalisiert werden können. Anschließend werden die normalisierten Daten von Array Tools erneut aufbereitet, damit dann die differentiell exprimierten Gene vom Excel *Add-in* SAM identifiziert werden können. Zusätzlich können Daten, wie die der Zeitreihe, mit dem Microarray-Analyseprogramm GeneSpring untersucht werden. Dieser Schritt wurde nicht automatisiert.

3.3.1 Array Tools

Array Tools ist ein *Add-in* für Microsoft Excel und wurde in *Visual Basic* geschrieben (Abb. 15). Excel ermöglicht die einfache Manipulation von Daten in Tabellenform und erlaubt das Erstellen von Graphen. Der Import und Export von Daten im *column separated values* (CSV) Format, wie sie von ScanArray Express erzeugt werden, ist möglich, und auch das Programm *Significance Analysis of Microarrays* (SAM) ist ein *Add-in*, das in Excel läuft. Array Tools ist auf der Internet-Seite des Instituts für Biochemie I verfügbar (<http://www.uni-koeln.de/med-fak/biochemie/transcriptomics/>).

Array Tools filtert und transformiert die Microarray-Rohdaten, um sie von R normalisieren zu lassen. Anschließend werden die Daten für SAM aufbereitet. Die einzelnen Schritte sind im Folgenden beschrieben.

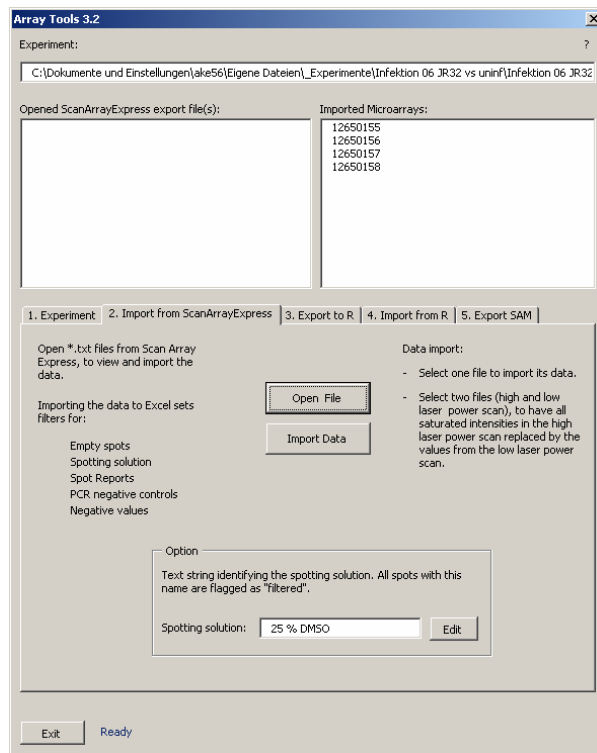


Abb. 15: Array Tools. Die Bedienoberfläche von Array Tools zeigt im oberen Drittel die Microarrays des Experimentes und gliedert sich darunter in fünf Karteikarten, mit denen sich folgende Schritte durchführen lassen:

1. Experiment anlegen
2. Import von ScanArray Express
3. Export nach R
4. Import von R
5. Export nach SAM

1. Experiment

Ein Microarray-Experiment besteht meist aus mehreren Microarrays, die der Messwiederholung dienen.

Um die Wiederholungen nutzen zu können, sollten die Microarrays untereinander skaliert sein, d. h. die Häufigkeitsverteilungen der Faktoren der differentiellen Genexpression sollten aneinander angeglichen sein (Yang *et al.*, 2002). Dies erfolgt während der Normalisierung in

R. Für die Identifikation der differentiell exprimierten Gene nutzt SAM die zwei Punkte je Sonde auf den Microarrays sowie alle Microarrays eines Experiments als Messwerte. Daher, und um die Gleichbehandlung der Microarrays zu gewährleisten, werden alle Microarrays eines Experiments in Array Tools in einer Container-Datei gespeichert und zusammen verarbeitet.

2. Import der Daten von ScanArray Express

Microarrays erlauben die Messung der Genexpression tausender Gene, doch die Messspanne ist nicht sehr groß. Gemessen wird die Genexpression in Form von Fluoreszenz-Intensität der markierten cDNA. Der Messbereich kann durch Wahl der Anregungsintensität und der Signalverstärkung gewählt werden. Dabei beobachtet man meistens, dass die stärksten Signale bereits in der Sättigung liegen, während die schwächsten noch nicht sichtbar sind. Um diesen Informationsverlust zu verringern, wurden die Microarrays mit zwei Messbereichen gescannt. Das erste Bildpaar wurde mit starker Anregung aufgenommen. Dabei waren die meisten Sondenpunkte gut sichtbar und auch von den Punkten mit sehr geringer Fluoreszenz waren viele zu erkennen (Abb. 16 a). Die Punkte mit der stärksten Intensität (in Abb. 16 a) sind es vier) hatten gesättigte Signale und konnten nicht verwendet werden.

Das zweite Bildpaar wurde mit niedrigerer Anregungsenergie aufgenommen, um auch die Sonden mit starken Signalen messen zu können. Wie man in Abb. 16 b) sieht, sind die Intensitäten aller Punkte unterhalb der Sättigung, gleichzeitig sind jedoch viele der Punkte mit geringer Fluoreszenz nicht mehr sichtbar.

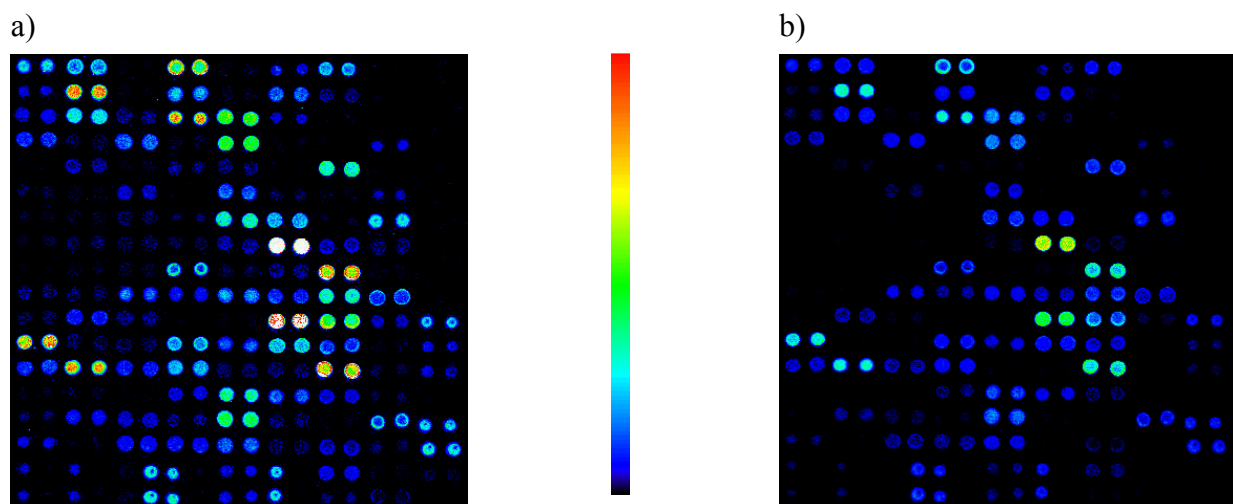


Abb. 16: Bilder eines Unter-Microarrays nach Anregung mit hoher und niedriger Energie. a) Fluoreszenzbild des Cy3-Kanals nach Anregung mit hoher Energie. Es sind vier Punkte mit gesättigten Signalen, dargestellt in weiß, zu sehen. b) Fluoreszenzbild des Cy3-Kanals nach Anregung mit niedriger Energie. Die gesättigten Signale aus a) liegen innerhalb der Messspanne. Abgebildet ist auch die Farbskala.

Anschließend wurden beide Bildpaare quantifiziert und die Messwerte mit Array Tools importiert. Da die meisten Punkte in den Bildern mit starker Anregung keine gesättigten Signale hatten, konnten ihre Messwerte aus diesen Bildern für die Datenanalyse verwendet werden. Für die in diesen Bildern gesättigten Punkte wurden automatisch die Messwerte der Bilder mit geringerer Anregung verwendet. Als Grenzwert wurde ein Gesamtanteil von 10 % gesättigter Bildpunkte im Cy3- und Cy5-Kanal des Punktes festgelegt. Die Anteile gesättigter Bildpunkte werden von ScanArray Express berechnet.

Die von ScanArray Express berechneten Rohdaten enthalten für alle Punkte, außer den Signalintensitäten und dem Anteil der gesättigten Bildpunkte, noch weitere Werte, die als Qualitätskriterien dienen können. Ein Wert ist die *flag* (zu dt. Bitschalter), ein Qualitäts-Kennzeichen, mit dem ScanArray Express Punkte markiert, die gefunden oder nicht gefunden wurden. Diese Kennzeichen kann man auch manuell setzen, und wurden während der Kontrolle der Quantifizierung dazu verwendet, um schlechte Punkte zu markieren. Schlechte Punkte treten z.B. dann auf, wenn Staubkörner auf einzelne Punkte gefallen sind. In Array Tools werden alle Punkte mit den Kennzeichen „not found“, „absent“ und „bad“ entfernt. Ausnahmen sind die Punkte, die nur in einem Kanal nicht detektierbar sind. Die Gene dieser Sonden werden nur einem experimentellen Zustand exprimiert und sind daher interessant. Um das „Rauschen“ des Hintergrunds zu kompensieren, müssen die Punkte im jeweils anderen Kanal eine Signal haben, das zehnmal so groß ist, wie der Hintergrund. Dann wird die Intensität für den Kanal ohne Signal auf eins gesetzt, um Quotienten berechnen zu können.

Zusätzlich werden alle Punkte entfernt, die folgende Filter nicht passieren:

- a) Die Signalintensität muss größer als null sein. Die Gene von Sondenpunkten ohne Signal werden nicht exprimiert und sind daher uninteressant.
- b) SpotReport-Sonden werden entfernt, da sie künstliche Expressionsfaktoren haben, die die Normalisierung stören könnten. Die Faktoren können einseitig sein und die SpotReport-Sonden kommen häufig auf dem Microarray vor.
- c) Negative Kontrollen werden entfernt, da für sie das Gleiche gilt wie für Sonden ohne Signal.

3. Export nach R

Die Normalisierung der Microarrays findet in dem Programm R statt (siehe Abschnitte 2.1. und 3.3.2). R ist ein Programm, das Daten in Form von Dateien importiert und exportiert und durch Kommandozeilenbefehle oder Skripten gesteuert werden kann. Array Tools erzeugt

daher aus den Microarray-Daten geeignete Dateien für R und ein Skript, das die Normalisierung in R steuert.

4. Import von R

Die normalisierten Daten der Microarrays werden in die Container-Datei in Excel importiert.

5. Export nach SAM

SAM ist ein Excel *Add-in* und kann damit die Daten in der Container-Datei nutzen. Dazu werden die Werte von Array Tools in einer für SAM nutzbaren Form angeordnet. SAM verwendet für die Identifikation der differentiell exprimierten Gene alle Messwerte einer Sonde. Um sicherzustellen, dass für einzelne Gene nicht sehr viel weniger Messwerte vorhanden sind, als für die meisten Gene, werden die Gene entfernt, bei denen mehr als die Hälfte der Werte nicht vorhanden sind.

3.3.2 Normalisierung der Microarray-Daten in R

R ist eine Programmiersprache, die zur Analyse und graphischen Darstellung von Daten verwendet werden kann und durch Module erweiterbar ist (Ihaka und Gentleman, 1996) (*The R Project for Statistical Computing*, <http://www.r-project.org/>). Das BioConductor-Projekt stellt zahlreiche Module für biologische Fragestellungen bereit, darunter auch Module zur Normalisierung von Microarray-Daten (Dudoit *et al.*, 2002; Dudoit und Yang, 2003) (BioConductor, <http://www.bioconductor.org/>). Für die Normalisierung der Microarray-Daten wurde R 1.6.2 mit den Modulen `marrayClasses`, `marrayInput`, `marrayNorm`, `marrayPlots` und `Biobase` verwendet.

Da sich mit R auch Graphen erzeugen lassen, werden automatisch für jedes Microarray vor und nach Normalisierung Kontrollgraphen erzeugt (Abb. 17). Dazu wird das Bland-Altman-Diagramm verwendet, bei dem für jeden Sondenpunkt die logarithmierten Unterschiede zwischen dem Cy3- und dem Cy5-Wert gegen den logarithmierten Mittelwert aufgetragen werden (Bland und Altman, 1986). Diese Auftragung wird auch als MA-Diagramm bezeichnet (Smyth *et al.*, 2002). Formeln 1 und 2 zeigen die Berechnung von M und A.

$$M = \log_2 \frac{\text{Intensität}_{\text{Experiment}}}{\text{Intensität}_{\text{Kontrolle}}} \quad (\text{Formel 1})$$

$$A = \log_2 \sqrt{\text{Intensität}_{\text{Experiment}} \times \text{Intensität}_{\text{Kontrolle}}} \quad (\text{Formel 2})$$

Abb. 17 zeigt Kontrollgraphen vor (a) und nach (b) Normalisierung. Vor der Normalisierung ist die Wolke der Datenpunkte zu negativen M-Werten hin verschoben, was auf eine insgesamt höhere Fluoreszenz-Intensität der Kontrolle zurückzuführen ist. Außerdem ist die Datenwolke gekrümmt. Dies kann mit unterschiedlichen Verhältnissen der Cy3- zur Cy5-Fluoreszenz über den Messbereich erklärt werden. Nach der Normalisierung ist die Datenwolke auf der Nulllinie von M zentriert und weist keine Krümmung mehr auf.

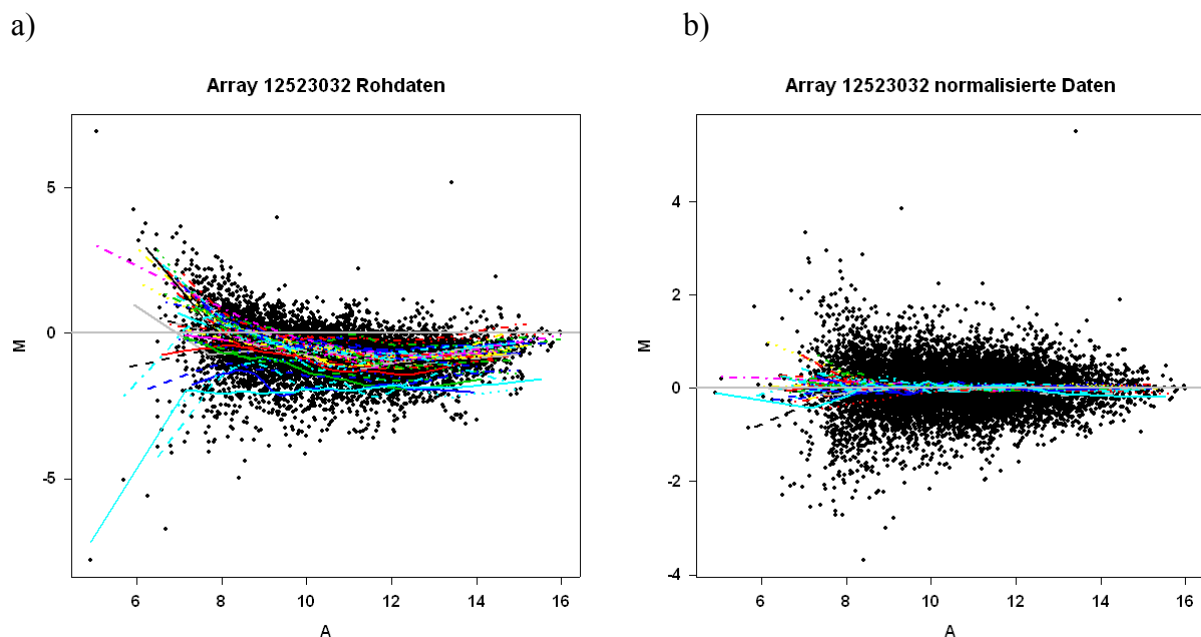


Abb. 17: Kontrollgraphen vor Normalisierung und nach Normalisierung. Aufgetragen ist der logarithmierte Unterschied (M) zwischen dem Cy3- und dem Cy5-Wert gegen den logarithmierten Mittelwert (A) für jeden Sondenpunkt. Die Punkte mit M-Werten größer Null zeigen induzierte Gene, während M-Werte kleiner Null reprimierte Gene bedeuten. Hohe A-Werte entsprechen Genen, die im Mittel hoch exprimiert sind, niedrige dem Gegenteil. Die farbigen Linien stellen die Regressionskurven der LOWESS-Normalisierung dar. Da die Normalisierung für jedes Unter-Microarray einzeln berechnet wird, sind 48 Kurven dargestellt. a) Vor Normalisierung. b) Nach Normalisierung.

Die Normalisierung wird in R für alle Microarrays eines Experiments zusammen durchgeführt und die normalisierten Daten werden anschließend skaliert, so dass nicht nur die Verteilung der M-Werte auf Null zentriert ist, sondern die Verteilungen auch annähernd gleich breit sind (Abb. 18).

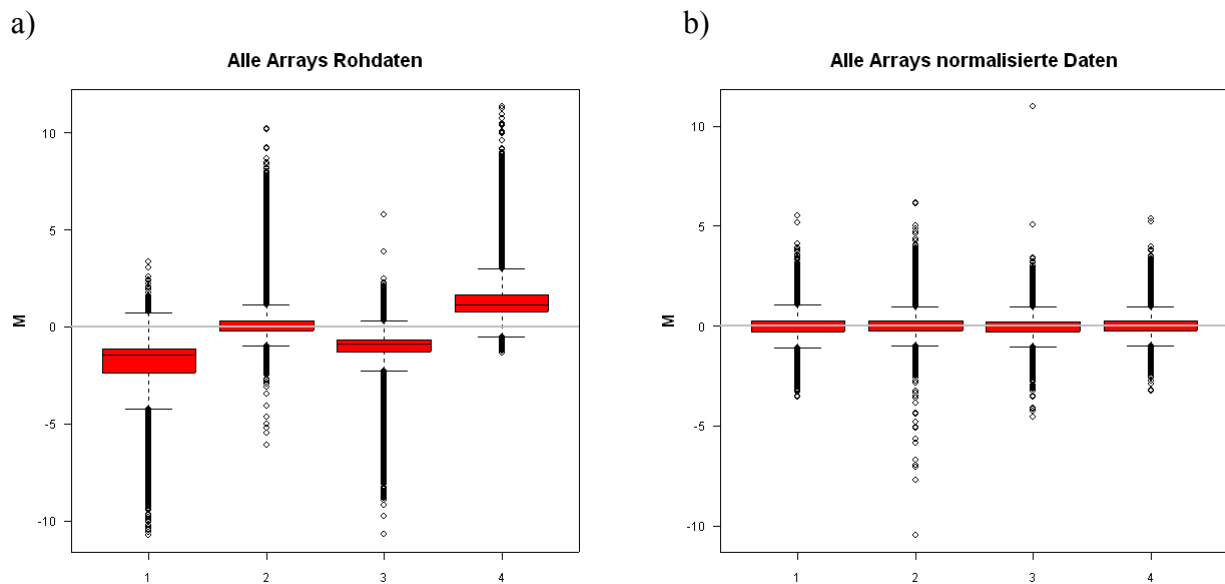


Abb. 18: Normalisierung und Skalierung der Microarrays eines Experiments. Dargestellt sind die Verteilung von vier Microarrays vor (a) und nach Normalisierung (b). Die roten Rechtecke umfassen 50 % der Datenpunkte eines Microarrays und der horizontale Strich im Rechteck zeigt den Median der Verteilung an. Die beiden horizontalen Striche ober- und unterhalb des roten Rechtecks begrenzen die Verteilung; Punkte jenseits dieser Striche sind Ausreißer (Tukey, 1977). Während der Normalisierung werden die Verteilungen der Arrays auf einen M-Wert von Null verschoben und anschließend die Höhe der Verteilungen angeglichen.

3.3.3 Identifikation differentiell exprimierter Gene mit SAM

Zur Identifikation der differentiell exprimierten Gene wurde das Programm SAM verwendet (siehe Abschnitte 2.1.9, 3.4.2 und 3.6.2).

3.4 Validierung des Microarray-Systems

3.4.1 Das SpotReport System

Das Microarray trägt die Sonden des SpotReport – 10 Array Validation Systems, mit denen die Faktoren der Induktion bzw. Repression der Genexpression, sowie die Nachweisgrenze des Microarrays überprüft wurden. Das SpotReport System besteht aus zehn DNA Sonden und den zugehörigen RNAs von *Arabidopsis thaliana* Genen. Diese Gene kodieren für Proteine der Photosynthese und ihre Sequenz wurde mittels BLAST mit den bekannten Sequenzen des *D. discoideum* Genoms verglichen. Die *A. thaliana* Gene sind von allen bekannten *D. discoideum* Genen so verschieden, dass keine Hybridisierung möglich sein sollte.

Da diese DNA-Sonden und RNA-Proben definierte Konzentrationen haben, können Sondenpunkte mit identischer Konzentration gedruckt werden. Zu der *D. discoideum* Proben-RNA können dann beliebige Mengen *Arabidopsis*-RNA gegeben werden, die anschließend

zusammen in cDNA transkribiert, farbmarkiert und hybridisiert werden. Die RNAs können zu gleichen Teilen oder in beliebigen Verhältnissen zum Experiment und zur Kontrolle gegeben werden. Dadurch können mit beliebigen Faktoren induzierte und reprimierte Gene simuliert werden und die im Microarray-Experiment gemessenen Faktoren mit den bekannten verglichen werden. Dabei können in einem Experiment zehn verschiedene Faktoren gleichzeitig verwendet werden. Außerdem kann durch Zusatz von geringen Mengen Probe die Nachweisgrenze des Systems ermittelt werden.

3.4.1.1 Faktoren der Induktion und Repression

Die *A. thaliana* RNAs wurden in Verhältnissen von 1:1, 2:1, 5:1 und 10:1 der RNA des Experiments und der Kontrolle eines Microarrays zugesetzt und die gemessenen Verhältnisse mit den tatsächlichen verglichen (Abb. 19).

Die gemessenen Faktoren spiegeln die tatsächlichen wider, sind jedoch stets zu gering. So wird zum Beispiel statt eines log-Faktors von 3,32 (entspricht 10) nur einer von 2,95 (entspricht 7,73) gemessen. Im Fall des 1:1 Verhältnisses wird ein Wert von 0,03 statt 0,00 gemessen. Ursachen könnten die Messunsicherheit sowie das verwendete Normalisierungsverfahren sein. Die Differenz von 0,03 ist jedoch so gering, dass sie praktisch bedeutungslos ist.

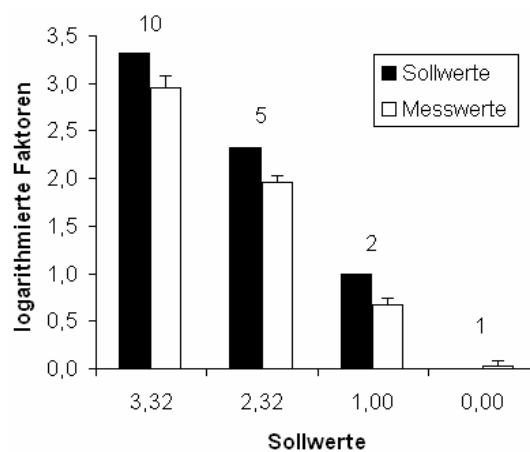


Abb. 19: Vergleich der tatsächlichen mit den gemessenen Faktoren der SpotReport Kontrollen. Dargestellt sind die \log_2 -Werte der Faktoren. Die nicht-logarithmierten Faktoren und die Standardfehler sind über den Balken angegeben.

3.4.1.2 Nachweisgrenze des Microarrays

Die Nachweisgrenze des Microarrays wurde bestimmt, indem 1 bis 400 pg der verschiedenen SpotReport RNAs der RNA von *D. discoideum* zugesetzt wurden. Es wurde beobachtet, bis zu welcher Menge ein Signal detektierbar war. In Abb. 20 a) lassen sich die Sondenpunkte bis zu einer Menge von 40 pg gut erkennen. Bei 5 pg ist noch etwas Fluoreszenz sichtbar, ein

Kreis lässt sich jedoch nicht mehr erkennen. Innerhalb der markierten Kreisfläche treten die blauen Punkte zwar etwas dichter auf als außerhalb, eine Unterscheidung vom Hintergrund ist jedoch kaum mehr möglich. Bei 1 und 2 pg sind keine Sondenpunkte mehr sichtbar. Die Nachweisgrenze des Microarray-Systems liegt also etwa bei 5 pg.

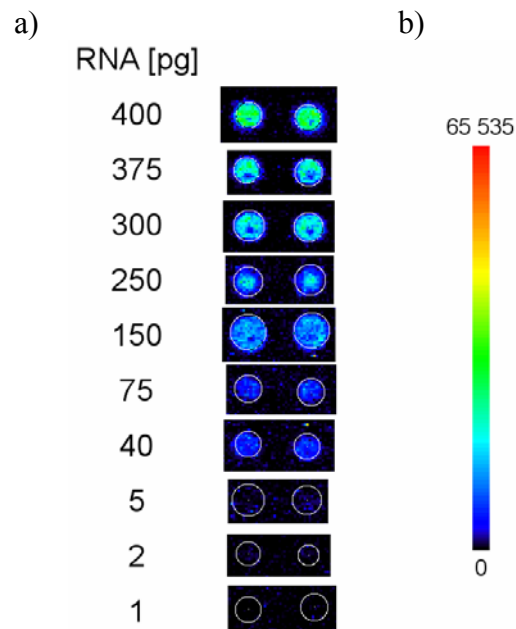


Abb. 20: Nachweisgrenze des Microarrays. a) Paare der zehn SpotReport-Sondenpunkte nach Hybridisierung mit 1 bis 400 pg der zugehörigen RNA. Die weißen Kreise markieren Positionen der Punkte. b) Falschfarben-Spektrum

Da jedoch die Signalintensitäten an der Nachweisgrenze definitionsgemäß an der Grenze des detektierbaren liegen, können bei so geringen RNA-Mengen keine präzisen Genexpressionsmessungen durchgeführt werden. Dies ist erst mit größeren Mengen möglich. Lügen z.B. für ein Gen im Experiment 40 pg und in der Kontrolle 5 pg RNA vor, so könnte zwar die Induktion des Gens festgestellt werden, doch der gemessene Faktor der Induktion wäre mit einer großen Messunsicherheit behaftet.

Anhand der für Säugerzellen bekannten Häufigkeiten der mRNA Klassen pro Zelle (Tabelle 2) kann abgeschätzt werden, für welche RNA-Klassen Genexpressionsänderungen mit dem Microarray messbar sind.

Tabelle 2: mRNA Populationen in einer Säugerzelle. Die mRNAs der Zelle sind in drei Klassen eingeteilt. Solche, die häufig, also in hoher Kopienzahl, in der Zelle vorkommen, mRNAs mit mittlerer Häufigkeit und seltene RNAs. Von den häufigen mRNAs gibt es nur vier verschiedene Spezies, während die seltenen mRNAs mit 11 000 Spezies 96 % aller mRNAs stellen. (modifiziert aus (Alberts *et al.*, 1994))

Klassen	Kopien jeder mRNA pro Zelle	x	Anzahl unterschiedlicher mRNAs	=	Anzahl der mRNAs jeder Klasse
Häufige mRNAs	12 000	x	4	=	48 000
Mittlere mRNAs	300	x	500	=	15 000
Seltene mRNAs	15	x	11 000	=	165 000
Summe					363 000

Der Anteil der mRNA an der gesamten cytoplasmatischen RNA beträgt 3 % bis 5 % (Alberts *et al.*, 1994). In einem Reaktionsansatz mit 20 µg totaler RNA sind im Mittel also 800 ng mRNA enthalten. Daher betragen die Mengen seltener, mittlerer und häufiger mRNA-Spezies:

$$\text{Häufige mRNAs: } \frac{12\,000 \text{ Kopien pro Zelle}}{363\,000 \text{ Kopien pro Zelle}} \times 800 \text{ ng} = 26 \text{ ng}$$

$$\text{Mittlere mRNAs: } \frac{300 \text{ Kopien pro Zelle}}{363\,000 \text{ Kopien pro Zelle}} \times 800 \text{ ng} = 661 \text{ pg}$$

$$\text{Seltene mRNAs: } \frac{15 \text{ Kopien pro Zelle}}{363\,000 \text{ Kopien pro Zelle}} \times 800 \text{ ng} = 33 \text{ pg}$$

Geht man davon aus, dass die mRNA-Häufigkeiten in *D. discoideum* ähnlich zu denen der Säugerzelle sind, so kann die differentielle Genexpression für häufige und mittlere mRNAs problemlos bestimmt werden. Die mittlere Mengen seltener mRNAs ist nahe an der Nachweisgrenze. Ein großer Teil dieser RNAs sollte zwar detektierbar sein, doch ihre Messung ist weniger genau als die der häufigen und mittleren mRNAs.

3.4.2 Einfluss der Anzahl von Experimentwiederholungen auf die Messunsicherheit

Die wiederholte Messung der Genexpression eines Experiments ist sehr wichtig, weil erst dadurch Aussagen über die Signifikanz der beobachteten Induktionen und Repressionen der Genexpression möglich werden (Lee *et al.*, 2000). Ein häufig verwendeter Signifikanztest für normalverteilte Daten ist der t-Test nach Student. Dieser Test ist jedoch nur für 20 oder mehr Messwerte geeignet (Kesel *et al.*, 1999). Efron *et al.* (2001) beschreiben einen t-Test, der einen empirisch ermittelten Wertungsfaktor enthält. Auch das Programm *Significance*

Analysis of Microarrays (SAM) von Tusher *et al.* (2001) verwendet einen modifizierten t-Test, um die signifikanten Gene und die Anzahl der falsch positiven Gene zu ermitteln.

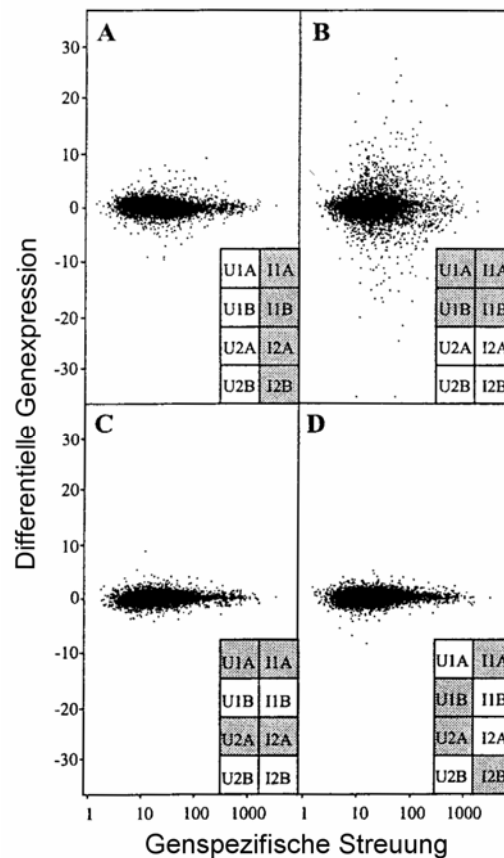


Abb. 21: Veranschaulichung des Permutationsverfahrens von SAM in Tusher *et al.* Aufgetragen ist der Faktor der Induktion oder Repression der Genexpression gegen die genspezifische Streuung für alle Sonden der verwendeten Microarrays. Die genspezifische Streuung entspricht der Standardabweichung der wiederholten Messungen. Es wurde RNA aus bestrahlten (I) und unbestrahlten (U) humane Lymphoblastoidzelllinien (1 und 2) revers transkribiert, mit Farbstoff markiert und mit acht Microarrays hybridisiert (A und B) und verglichen. A) Vergleich der bestrahlten und unbestrahlten Zelllinien. B) Vergleich der Zelllinien 1 und 2. C) Vergleich der Hybridisierungen A und B. D) Eine der 36 möglichen Permutationen, bei denen gleich viele Microarrays der Zelllinien 1 und 2 als Kontrolle dienen. Bei den Permutationen werden sowohl die bestrahlten, als auch die unbestrahlten Zellen als Kontrolle verwendet. Grau: Experiment, weiß: Kontrolle (aus Tusher *et al.*, 2001).

SAM berechnet für jedes Gen einen Signifikanzwert mit Hilfe eines Permutationstests. Dabei wird der gemessene Faktor der Induktion oder Repression mit der Messunsicherheit verglichen. Die unspezifischen Schwankungen der Genexpression werden bestimmt, indem alle möglichen Kombinationen von Experiment und Kontrolle, Experiment und Experiment sowie Kontrolle und Kontrolle miteinander verglichen werden (Abb. 21). Die Genauigkeit der Berechnung nimmt mit der Zahl der Permutationen zu. Die maximale Anzahl möglicher Permutationen kann mit dem Binomialkoeffizienten bestimmt werden. Der Binomialkoeffizient (Formel 3) berechnet sich aus den k Messungen des Experiments und der Gesamtzahl der Messungen N , also Experiment und Kontrolle.

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \quad (\text{Formel 3})$$

Werden nur wenige Messungen, $k = 1$ bis 3, durchgeführt, so sind nur wenige Permutationen möglich, was sich negativ auf die Präzision des ermittelten genspezifischen Streuwertes auswirkt. Mit zunehmender Zahl von Messungen nimmt die Anzahl der möglichen Permutationen stark zu (Tabelle 3), was die Aussagekraft der von SAM durchgeführten Berechnungen deutlich erhöht.

Tabelle 3: Entwicklung des Binomialkoeffizienten. Beispiele der maximalen Permutationszahl für verschiedene N (Summe der Messungen von Kontrolle und Experiment) und k (Messungen des Experiments). Dabei wird stets $k = \frac{1}{2} N$ gewählt, da typischerweise die Hälfte der Messungen mit der RNA der experimentellen Bedingung durchgeführt wird.

k	N	Binomialkoeffizient; entspricht den maximal möglichen Permutationen
1	2	2
2	4	6
3	6	20
4	8	70
5	10	252
6	12	924

Durch eine ausreichend große Zahl von Wiederholungen werden nicht nur statistische Tests und Permutationsanalysen ermöglicht, auch der mittlere Fehler des arithmetischen Mittels (= Standardfehler, Formel 4) verringert sich. Je mehr Messwerte zur Verfügung stehen, um so geringer ist der Standardfehler und umso genauer kann der Faktor der differentiellen Expression eines Gens angegeben werden. Dies ist vor allem dann bedeutsam, wenn der Faktor klein ist, da sich sonst eine reale Induktion oder Repression nicht von der Messunsicherheit unterscheiden läßt. Der Standardfehler s_0 berechnet sich aus der Standardabweichung s und der Anzahl der Werte n .

$$\text{Standardabweichung: } s = + \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (\text{Formel 4})$$

$$\text{Standardfehler: } s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (\text{Formel 5})$$

Der Standardfehler nimmt mit den ersten zwölf Messwerten etwa auf die Hälfte des Wertes mit zwei Werten ab, bei einer weiteren Erhöhung der Messwerte nimmt er zunehmend langsamer ab (Kesel *et al.*, 1999).

Um zu überprüfen ob der Standardfehler bei einer Erhöhung der Anzahl verwendeter Microarrays wie erwartet abnimmt, wurde ein Permutationstest durchgeführt. Möglich wäre, dass der Standardfehler nicht wie erwartet abnimmt, da die Standardabweichung der Messungen durch die verschiedenen Arten von technischen (Doppelpunkte, Microarrays und *dye-swaps*) und biologischen (unabhängige RNA-Isolationen) Wiederholungen stark zunimmt.

Es wurden die Microarrays von zwei Experimenten sowie ein simulierter Datensatz verwendet. Der erste Datensatz bestand aus jeweils dem ersten Punkt der Doppelpunkte je Sonde der ersten 16 Microarrays des Infektionsexperiments mit *L. pneumophila* (siehe Tabelle 4). Die Beschränkung auf 16 Messwerte resultierte daraus, dass hierfür bereits die Standardfehler von 65519 Microarray-Kombinationen berechnet werden mußten (Summe der Binomialkoeffizienten mit $N = 16$ und $k = 2$ bis 16). Die $2,7 \times 10^{11}$ möglichen Kombination bei 19 Microarrays mit Doppelpunkten hätten zu viel Rechenzeit beansprucht.

Der zweite Datensatz bestand aus zehn Microarrays, mit denen die $\Delta limD$ -Mutante von *D. discoideum* mit dem Wildtyp verglichen wurde (von D. Bakthavatsalam zur Verfügung gestellt). Zum Vergleich wurde ein Datensatz aus Zufallszahlen für 16 Microarrays mit je 6000 Datenpunkten verwendet, wobei die Zufallszahlen wie die logarithmierten-Faktoren der realen Experimente um Null streuten.

Ausgehend von dem vollständigen Datensatz mit zehn bzw. 16 Microarrays wurden zunächst alle Microarray-Kombinationen der Teilmengen mit 15, 14, 13 usw. Microarrays berechnet. Für jede Sonde auf den Microarrays wurde dann der Standardfehler der Messwerte einer Microarray-Kombination berechnet. Anschließend wurden die Standardfehler der Sonden gemittelt. In einem zweiten Schritt wurden die mittleren Standardfehler aller Microarray-Kombinationen einer Teilmenge gemittelt. Dadurch konnte berechnet werden, wie groß der mittlere Standardfehler gewesen wäre, wenn statt 16 nur 15 oder weniger Microarrays verwendet worden wären. Für die Darstellung in Abb. 22 wurden die Standardfehler des $\Delta limD$ -Experiments und der Simulation auf das Infektionsexperiment normiert, um sie vergleichbar zu machen. Dazu wurde der Standardfehler der Teilmenge mit zwei Microarrays gleichgesetzt. Das nicht normierte $\Delta limD$ -Experiment wies höhere Standardfehler als das Infektionsexperiment auf.

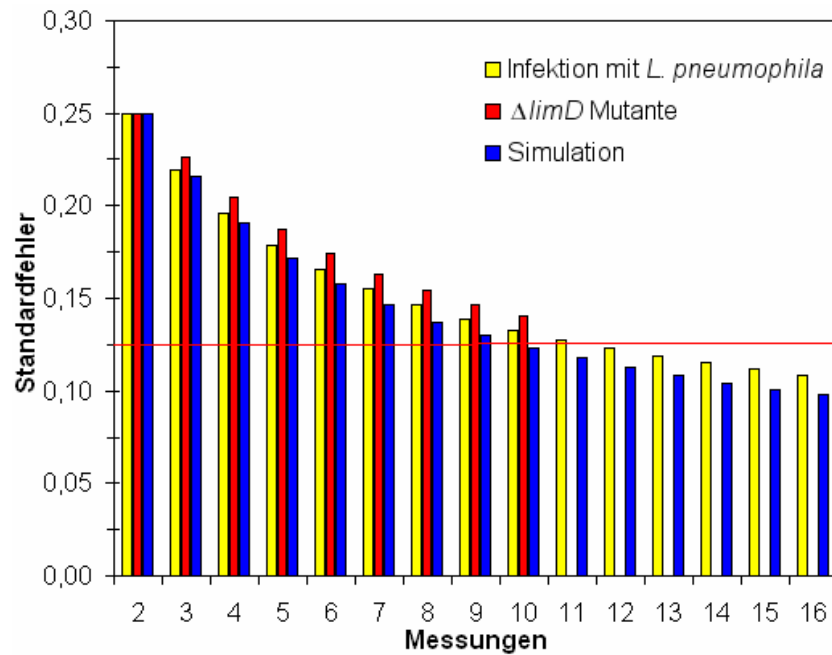


Abb. 22: Entwicklung des Standardfehlers mit zunehmender Anzahl von Messungen. Der Standardfehler aller drei Datensätze nimmt mit zunehmender Microarrayzahl ab und unterschreitet ab elf bzw. zwölf Messungen die Hälfte des maximalen Werts (rote Linie).

Die Entwicklung des Standardfehlers des $\Delta limD$ - und des Infektionsexperiments folgen dem Verlauf des simulierten Datensatzes und nehmen kontinuierlich ab. Zunächst nimmt der Standardfehler stark ab, unterschreitet bei zwölf Microarrays die Hälfte seines initialen Wertes und nimmt mit zunehmender Microarrayzahl weiter, jedoch immer langsamer, ab. Dies stimmt mit dem in Kesel *et al.* (1999) beschriebenen Verlauf überein.

Die Standardfehler der beiden realen Experimente nehmen nicht ganz so stark ab wie die der Simulation mit Zufallszahlen. Dies könnte damit erklärt werden, dass durch das Hinzufügen von Microarrays zum Experiment nicht nur weitere Messwerte zur Verfügung stehen, über die gemittelt werden kann, sondern dass mit unabhängigen RNA-Isolationen und *dye-swaps* auch zusätzliche Variablen hinzu kommen.

Aufgrund der Tatsache, dass statistische Tests erst mit einer genügend großen Zahl von Messwiederholungen zuverlässige Ergebnisse liefern, die Zahl der möglichen Permutationen durch das Programm *Significance Analysis of Microarrays* erst ab fünf bis sechs Microarrays hohe Werte erreicht und der mittlere Standardfehler ab ca. zwölf Messwerten halbiert werden kann, wurden für die folgenden Experimente nach Möglichkeit sechs Microarrays verwendet. Diese liefern aufgrund ihrer Doppelpunkte zwölf Messwerte pro Gen.

3.4.3 Faktoren, die Einfluss auf die detektierten Genexpressionsänderungen haben

Bei der wiederholten Messung der Genexpressionsänderungen gibt es mehrere Formen von Messwiederholungen, die verwendet werden können.

Eine Möglichkeit ist, dass in unabhängigen Kulturen parallel mehrere Infektionen durchgeführt werden, aus den *D. discoideum* Zellen jeder Kultur RNA isoliert wird und mit jeder RNA ein Microarray Experiment durchgeführt wird. Hierbei wird die biologische Variabilität bestimmt. Sie ist eine Eigenschaft aller lebenden Organismen und unvermeidbar, doch durch ihre Messung ist man in der Lage, die Signifikanz der gefundenen Genexpressionsänderungen anzugeben (Churchill, 2002).

Die zweite Art der Wiederholung erfolgt auf dem *D. discoideum* DNA-Microarray, da für jede Sonde zwei und im Fall der Kontrollen mehrere Punkte auf dem Microarray existieren.

Die dritte Möglichkeit, die wiederholte Messung zu variieren, ist das Vertauschen der Farbmarkierung von Experiment- und Kontroll-cDNA. Da es sich bei Cy3 und Cy5 um zwei Farbstoffe mit geringfügig unterschiedlicher Struktur und teilweise deutlich verschiedenen Fluoreszenz-Eigenschaften handelt, ist es möglich, dass die Farbstoffe einen großen Einfluss auf die gemessenen Fluoreszenz-Intensitäten ausüben. Die Tatsache, dass häufig eine der beiden Probenmischungen generell stärkere Fluoreszenz emittiert als die andere, wird durch die Normalisierung ausgeglichen. Um einen möglichen Gen-spezifischen Effekt der Farbstoffe zu minimieren, wird der sogenannte *dye-swap* verwendet (Smyth *et al.*, 2002). Die Wiederholungen vom Typ zwei und drei werden auch als *technical replicates* bezeichnet (Churchill, 2002).

Für die Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich mit uninfizierten Zellen nach 24 Stunden Koinkubation wurden die Einflüsse der drei beschriebenen Wiederholungsmöglichkeiten untersucht. Dieses Experiment war dazu am besten geeignet, da hier die größte Anzahl von Wiederholungen vorhanden war. Es gab sieben unabhängige Total-RNA-Extrakte, die zu vier verschiedenen Zeitpunkten isoliert worden waren und unter Vertauschung der Farbmarkierung von Experiment- und Kontroll-cDNA markiert und mit insgesamt 19 Microarrays hybridisiert wurden (Tabelle 4).

Die gemessenen Genexpressionsquotienten aller Gene der 19 Microarrays wurden in GeneSpring 6.1 mittels Clusteranalyse verglichen. Dabei stellen die jeweils doppelt aufgedruckten Sonden der 19 Microarrays die experimentellen Bedingungen dar.

Das Ergebnis einer hierarchischen Clusteranalyse sind Cluster und die Korrelationskoeffizienten der Cluster. Die Cluster können sowohl für die Gene (vergleiche Abb. 23), als auch für die experimentellen Bedingungen berechnet werden. Im ersten Fall erhält man Gruppierungen von Genen, die über die experimentellen Bedingungen hinweg ähnlich reguliert werden. Im zweiten Fall wird ermittelt, unter welchen Bedingungen das gesamte Transkriptom ähnlich reguliert wird. Dargestellt werden die Cluster in Form von Baumstrukturen (Dendrogrammen), bei denen die Blätter den einzelnen Genen oder Bedingungen entsprechen und die Länge des Astes zwischen zwei Verzweigungspunkten von den Korrelationskoeffizienten bestimmt wird. Führen zum Beispiel zwei Bedingungen zu annähernd identischen Änderungen der Genexpression, so ist zu erwarten, dass diese Bedingungen einen hohen Korrelationskoeffizienten haben und sie daher schon nach einer kurzen Astlänge einen Cluster bilden.

Die Clusteranalyse bietet zahlreiche Möglichkeiten für die Berechnung der Cluster. Es gibt eine Vielzahl von Algorithmen zur Bestimmung der Korrelation von Datenpunkten. Wichtig ist dabei, dass die mathematische Korrelation die Ähnlichkeit von koordiniert regulierten Genen gut abbildet. Eisen *et al.* beobachteten, dass der Standard-Korrelationskoeffizient gute Ergebnisse liefert, da hier die Ähnlichkeit der Form von Expressionsprofilen erfasst wird, während die Höhe des Profils, d. h. die RNA-Kopiezahl, eine untergeordnete Rolle spielt (Eisen *et al.*, 1998).

Da der verwendete Cluster-Algorithmus einen großen Einfluss auf das Ergebnis haben kann wurde neben der Standardkorrelation auch die Pearson- und die Spearman-Korrelation verwendet und die Ergebnisse verglichen. Alle 38 Messungen je Gen wurden bei den drei Methoden über dieselben Knotenpunkte verknüpft. Unterschiedlich waren lediglich die Astlängen und die Anordnung des Dendrogramms. Das bedeutet, dass die Frage, welche Messungen am stärksten übereinstimmen, in allen Fällen zum gleichen Ergebnis führt. Dies könnte dadurch erklärt werden, dass diese Zuordnung sehr robust erfolgt, weil die Daten in Form von Clustern vorliegen, die selbst mit unterschiedlichen Korrelationsmethoden nur auf eine Weise gruppiert werden können.

Der Grad der Ähnlichkeit wird von den drei Methoden unterschiedlich bewertet, hat aber keinen Einfluss auf die Verknüpfungen der Cluster. Die Anordnung des Dendrogramms hat meist keine Bedeutung, da für ein Dendrogramm mit n Elementen 2^{n-1} lineare Anordnungen möglich sind (Morris *et al.*, 2003). Häufig wird zwar versucht, eine Anordnung zu erzeugen,

welche die Beziehungen gut wieder gibt, doch eine optimale Anordnung wird aus praktischen Gründen meist nicht berechnet.

Die größte Korrelation liegt zwischen den Doppelpunkten eines Microarrays vor. Alle 38 Einzelpunkt-Messungen bilden auf der niedrigsten Ebene einen Cluster mit ihrem jeweiligen Nachbarpunkt aus (Abb. 23). In der Hybridisierungsreaktion eines Microarrays stellen sich bei zwei Punkten mit identischem Sondenmaterial offensichtlich sehr ähnliche Cy3/Cy5-Quotienten ein, die auch während des Quantifizierungsprozesses nicht wesentlich verschlechtert werden.

Die nächsthöhere Ebene von Clustern wird in acht Fällen aus Arrays mit gleicher Farbmarkierung und in einem Fall (Microarrays 12 49 49 77 und 12 49 49 78) mit gleicher RNA gebildet. Unterschiedliche Farbmarkierung des Experiment-Kontrolle-Paares (*dye-swap*) hat also einen größeren Einfluss auf die Genexpressionsquotienten als die möglichen Unterschiede zwischen Paaren von Sondenpunkten.

Ob unterschiedliche Microarrays oder unterschiedliche RNA-Ansätze eines Isolationsdatums eine höhere Variabilität erzeugen, kann nicht ermittelt werden, da ein RNA-Ansatz in einer Farbstoffkombination immer nur auf einem Microarray hybridisiert worden ist. Die Unterschiede zwischen RNA-Ansätzen ist jedoch geringer als zwischen den *dye-swaps*, da die Clusterbildung nach Farbstoff-Kombination und nicht nach RNA-Ansatz erfolgt.

Auf der drittniedrigsten Ebene des Dendrogramms bestehen zwei Cluster jeweils aus Microarrays, die mit RNA eines Isolationsdatums (8/03 und 5/03) hybridisiert wurden. Zwei weitere Cluster enthalten Microarrays mit der RNA aus 3/02. Einer der beiden Cluster beinhaltet sechs Microarrays, deren Experiment-cDNA mit Cy3 markiert wurde. Der andere Cluster setzt sich aus sechs Cy5-markierten und zwei Cy3-markierten Experiment-cDNAs zusammen.

Die Microarrays mit der RNA aus 7/03 bilden auf der höchsten und der zweithöchsten Ebene Cluster mit den übrigen Arrays. Dabei korrelieren die Microarrays mit Cy5-markiertem Experiment besser mit den Microarrays, die mit RNA anderer Isolationsdaten verwendet wurden, als mit den *dye-swap* Microarrays. Dabei ist jedoch zu beachten, dass die Astlängen der oberen vier Ebenen des Dendrogramms sehr kurz sind und deshalb die beobachteten Unterschiede sehr gering sind.

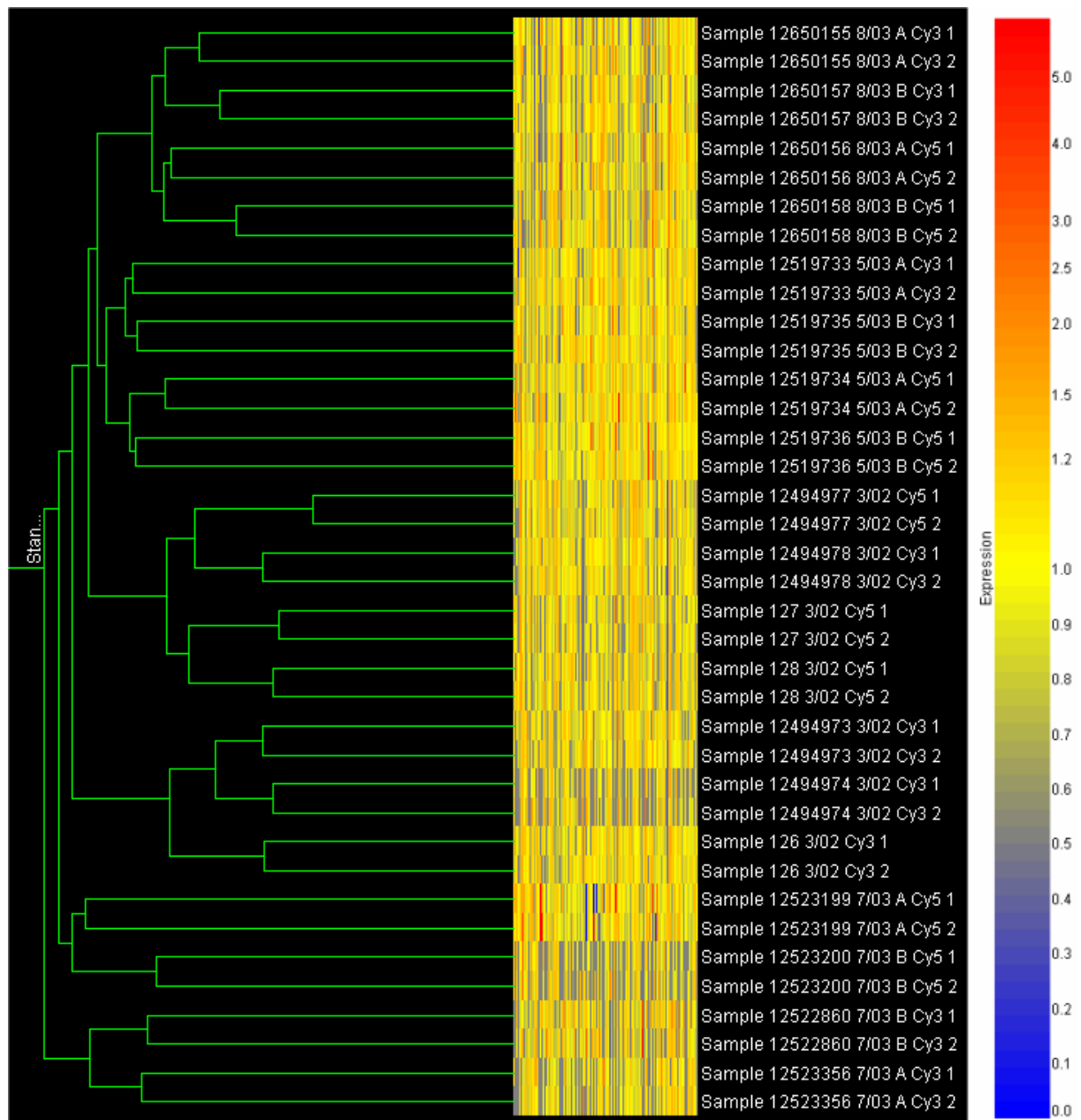


Abb. 23: Vergleich der 19 Microarrays des Infektionsexperiments. Clusteranalyse der 38 Einzelmessungen je Gen für die Infektion von *D. discoideum* mit *L. pneumophila* 24 Stunden nach Koinkubation. Jedes Blatt des Dendrogramms zeigt die Faktoren der Induktion (rot), Repression (blau) oder Unverändertheit (gelb) der Expression von 5661 Genen an. Sonden, für die kein Messwert vorliegt sind grau markiert. Die Einzelmessungen sind gekennzeichnet mit der Barcode-Nummer des Microarrays (z.B. 12650155), dem Isolationsdatum der RNA (z.B. 8/03), dem RNA-Ansatz falls zutreffend (A oder B), der Farbmarkierung des Experiments (Cy3 oder Cy5) und der Nummer des Doppelpunktes (1 oder 2). Die Farbskala für die Induktion und Repression von Genen und die zugehörigen Faktoren sind rechts dargestellt.

Insgesamt zeigt sich, dass die biologische Variabilität der RNA mit verschiedenen Isolationsdaten zusammen mit dem *dye-swap* die größte Messunsicherheit verursachen, während die Variabilität zwischen RNA-Ansätzen eines Isolationsdatums sowie technische Variabilität der Microarrays und ihrer Doppelpunkte gering ist. Diese Ergebnisse stimmen mit den Beobachtungen von Churchill (2002) sowie Yang und Speed (2002) überein.

Für die Planung von Microarray-Experimenten bedeuten diese Ergebnisse, dass es wichtig ist mehrere biologische Messwiederholungen, also unabhängige RNA-Extraktionen zu verwenden, und *dye-swaps* einzusetzen. Die biologische Variabilität ist interessant, da sie Aufschluss über die Robustheit des untersuchten Systems gibt während der *dye-swap* dazu dient, Unterschiede der eingesetzten Farbstoffe auszugleichen. Es ist nicht sinnvoll einen RNA-Extrakt für mehrere Microarrays zu verwenden, da die technische Variabilität gering ist und man über die biologische Variabilität im Ungewissen bleibt.

3.5 Identifikation der differentiell exprimierten Gene von *D. discoideum* 24 h nach Infektion mit *L. pneumophila*

Die Genexpression von *D. discoideum* als Reaktion auf die Infektion mit *L. pneumophila* wurde zunächst durch einen Vergleich von infizierten mit nicht infizierten Zellen nach 24 Stunden Koinkubation untersucht. Zwei weitere Experimente erfolgten mit *L. hackeliae* und *L. pneumophila* $\Delta dotA$ als Kontrolle.

3.5.1 Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zu uninfizierten Zellen

Für die Infektion wurde der Stamm *Legionella pneumophila* Philadelphia I JR32 verwendet. Dies ist ein virulenter *Legionella* Stamm, der aus dem Isolat aus einem Patienten hervorgegangen ist (Marra und Shuman, 1989). Der Zeitpunkt 24 Stunden nach Infektionsbeginn wurde gewählt, da zu diesem Zeitpunkt noch keine intrazelluläre Vermehrung der Legionellen und noch keine Lyse der Wirtszellen erfolgt, die Infektionsrate jedoch höher als zu früheren Zeitpunkten ist. Intrazelluläre Vermehrung ist ab 48 Stunden sichtbar, Wirtszell-Lyse tritt nach 72 bis 96 Stunden ein (Hägele *et al.*, 2000; Skriwan *et al.*, 2002). Es wurden bei diesem Vergleich sehr viele Microarrays verwendet, um die Zuverlässigkeit des Microarray-Systems zu überprüfen (siehe Abschnitt 3.8.4). Mit 19 Microarrays, die für die Untersuchung von sieben unabhängigen RNA-Paaren verwendet wurden, ist dieser Vergleich daher statistisch am besten abgesichert (Tabelle 4).

Tabelle 4: Experiment-Design. Hybridisierungen des Vergleichs von infizierten *D. discoideum*-Kulturen mit *L. pneumophila* nach 24 Stunden Koinkubation. Angegeben ist jeweils die Barcode-Nummer des Microarrays, der RNA-Extrakt mit Isolationsdatum und Ansatz, falls zwei parallele Extraktionen durchgeführt wurden, sowie der verwendete Farbstoff.

Microarray	Experiment		Kontrolle	
	Total-RNA	Farbstoff	Total-RNA	Farbstoff
12 52 33 56	L. p. infiziert 7/2003 Ansatz A	Cy3	uninfiziert 7/2003 Ansatz A	Cy5
12 52 31 99	L. p. infiziert 7/2003 Ansatz A	Cy5	uninfiziert 7/2003 Ansatz A	Cy3
12 52 28 60	L. p. infiziert 7/2003 Ansatz B	Cy3	uninfiziert 7/2003 Ansatz B	Cy5
12 52 32 00	L. p. infiziert 7/2003 Ansatz B	Cy5	uninfiziert 7/2003 Ansatz B	Cy3
1 26	L. p. infiziert 3/2002	Cy3	uninfiziert 3/2002	Cy5
1 27	L. p. infiziert 3/2002	Cy5	uninfiziert 3/2002	Cy3
1 28	L. p. infiziert 3/2002	Cy5	uninfiziert 3/2002	Cy3
12 49 49 73	L. p. infiziert 3/2002	Cy3	uninfiziert 3/2002	Cy5
12 49 49 74	L. p. infiziert 3/2002	Cy3	uninfiziert 3/2002	Cy5
12 49 49 77	L. p. infiziert 3/2002	Cy5	uninfiziert 3/2002	Cy3
12 49 49 78	L. p. infiziert 3/2002	Cy5	uninfiziert 3/2002	Cy3
12 51 97 33	L. p. infiziert 5/2003 Ansatz A	Cy3	uninfiziert 5/2003 Ansatz A	Cy5
12 51 97 34	L. p. infiziert 5/2003 Ansatz A	Cy5	uninfiziert 5/2003 Ansatz A	Cy3
12 51 97 35	L. p. infiziert 5/2003 Ansatz B	Cy3	uninfiziert 5/2003 Ansatz B	Cy5
12 51 97 36	L. p. infiziert 5/2003 Ansatz B	Cy5	uninfiziert 5/2003 Ansatz B	Cy3
12 65 01 55	L. p. infiziert 8/2003 Ansatz A	Cy3	uninfiziert 8/2003 Ansatz A	Cy5
12 65 01 56	L. p. infiziert 8/2003 Ansatz A	Cy5	uninfiziert 8/2003 Ansatz A	Cy3
12 65 01 57	L. p. infiziert 8/2003 Ansatz B	Cy3	uninfiziert 8/2003 Ansatz B	Cy5
12 65 01 58	L. p. infiziert 8/2003 Ansatz B	Cy5	uninfiziert 8/2003 Ansatz B	Cy3

Die Berechnung der signifikant differentiell exprimierten Gene mit SAM liefert 1218 Gene, bei einer Rate der falsch Positiven von 0,03 %. Es werden also 0,46 falsch positive Gene unter den 1218 gefundenen Genen erwartet. Von den 1218 Genen sind 612 induziert und 606 reprimiert. Die höchste Induktion ist 5,9fach und die stärkste Repression ist 3,2fach.

Wie man in Abb. 24 sieht, haben einige wenige Gene hohe *scores*, also einen hohen Wert der SAM-Statistik. Den höchsten Betrag hat mit 15,2 die cDNA-Sonde SSB153. Dieser Wert ist der Quotient des mittleren, logarithmischen Faktors der Induktion (2,56) und der Messunsicherheit (0,168). Der Faktor, um den das zugehörige Gen induziert ist beträgt $2^{2,56} = 5,9$. Auch reprimierte Gene haben teilweise hohe *score*-Beträge. Zwischen den Genen mit sehr hohen Signifikanzwerten liegen die Gene mit niedrigeren Werten, sowie die nicht signifikant regulierten Gene. Der Übergang von den nicht-signifikanten zu den signifikanten Genen ist fließend. Das Gen mit dem geringsten, noch als signifikant angesehenen *score* ist nur um den Faktor 1,2 induziert. Es ist zwar vorstellbar, dass eine 1,2fache Induktion eine Wirkung in *D. discoideum* hat, doch da dies weder nachgewiesen werden kann noch eine Bestätigung dieses Ergebnisses mittels Northernblot oder quantitativer PCR nicht möglich ist, werden im Folgenden nur die Gene mit hohen *scores* näher betrachtet.

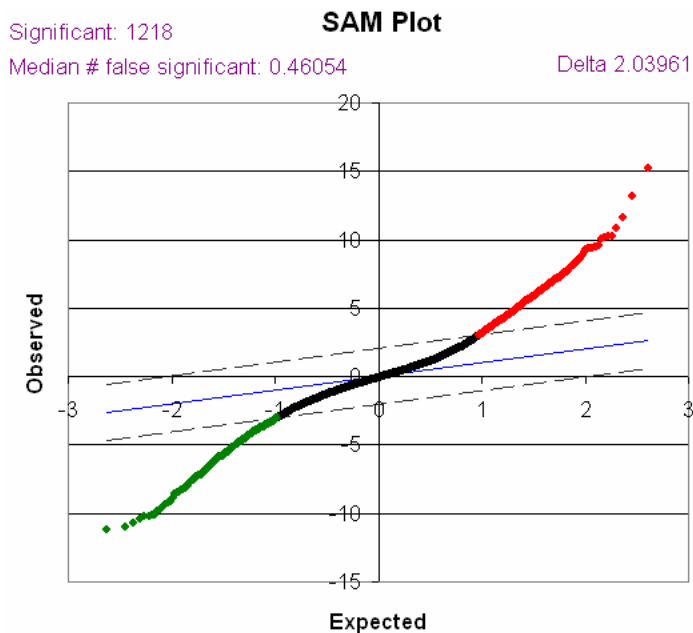


Abb. 24: SAM-Diagramm. Aufgetragen sind die *scores*, d. h. der Wert der SAM-Statistik jedes Gens (*observed*) gegen den *score*, der für das Gen aufgrund seiner Messunsicherheit erwartet wird (*expected*). Ein positiver *score* zeigt hochregulierte Gene, wobei signifikante Gene rot markiert sind. Runterregulierte Gene haben einen negativen *score*; die signifikanten sind grün markiert. Angegeben ist auch der delta-Wert. Das ist der Schwellenwert der Signifikanz und sein Betrag ist der Schnittpunkt der gestrichelten Linien mit der Ordinate im positiven und negativen Bereich. Außerdem wird die Anzahl der signifikanten Gene und die der erwarteten falsch positiven angezeigt.

Es wurden die Sonden mit den zehn höchsten und zehn niedrigsten *score*-Werten, sowie die Gene mit den zehn höchsten Faktoren der Induktion und Repression ausgewählt, um sie zu annotieren. Insgesamt wurden cDNA-Sequenzen von 16 induzierten und 13 reprimierten Genen annotiert.

Da für die Sonden des japanischen cDNA-Projekts nur die DNA-Sequenz vorlag, wurde mit dem Programm BLAST (Altschul *et al.*, 1990) nach Proteinen bekannter Funktion in *D. discoideum* und anderen Organismen gesucht, um Hinweise zu erhalten, welche zelluläre Funktion die zu den Sonden gehörenden Gene haben könnten. Je nach Länge der Sondensequenz wurde diese entweder direkt für eine Suche mit BLASTX gegen die UniProt-Datenbank oder die *non-redundant* Datenbank des NCBI verwendet, oder zunächst mit einer BLASTN-Suche gegen das *D. discoideum*-Genom der zugehörige offene Leserahmen gesucht und dessen Aminosäuresequenz anschließend für eine BLASTP-Suche verwendet (siehe Abschnitt 2.2.6.3).

3.5.1.1 Annotation der Sonden induzierter Gene

Für zwei der Sonden, bei denen eine Induktion gemessen wurde, konnten entsprechende *D. discoideum* Proteine identifiziert werden, die bereits charakterisiert worden waren (Tabelle 5). Das Gen *rtoA* (*ratioA*) wurde bei einer Mutagenisierungsstudie identifiziert. Die Mutanten haben während des Entwicklungszykluses einen unnatürlich hohen Anteil von Prästielzellen. RtoA ist ein 39,8 kDa großes Protein, das zehn Wiederholungen einer Serin-reichen Sequenz

mit elf Aminosäuren enthält (Wood *et al.*, 1996). Diese Wiederholungen in der Mitte der Sequenz machen 60 % des Proteins aus und Brazill *et al.* (2000) haben gezeigt, dass dieser Abschnitt eine zufällige Windung (*random coil*) aufweist. Der Serin-reiche Teil von RtoA, das S-Fragment, katalysiert *in vitro* die Fusion von Phosphatidylcholin-Vesikeln, wobei der Inhalt der Ausgangsvesikel verlustfrei in die Fusionsvesikel übergeht. *In vivo* ist RtoA für die Fusion von Endozytose-Vesikeln wichtig. Brazill *et al.* (2000) konnten zeigen, dass in $\Delta rtoA$ Zellen die Fusion von Endo- und Exozytose-Vesikeln verringert ist.

Die Sonde SLJ833 repräsentiert das Gen *manA*, das für die lysosomale α -Mannosidase kodiert. α -Mannosidase wird als ein 140 kDa großes Präprotein synthetisiert und wird proteolytisch in zwei lösliche 58 und 60 kDa große Untereinheiten gespalten (Schatzle *et al.*, 1992). α -Mannosidase ist eine in eukaryotischen Zellen ubiquitäre Exoglykosidase, die α -D-mannosidische Bindungen spaltet und dadurch N-glykosidische Oligosaccharide abbaut. In Säugern sind Mutationen im Gen der α -Mannosidase Ursache der Mannosidosis, einer Störung des Glykoproteinabbaus, der zur Akkumulation von Oligosacchariden und der Vergrößerung der Lysosomen führt (Beccari *et al.*, 1999; Sun und Wolfe, 2001). Mannose ist neben Rhamnose, Glucosamin und anderen Zuckern Teil der Lipopolysaccharide der äußeren Membran von *L. pneumophila* (Otten *et al.*, 1986).

Für sieben der Sonden wurden ähnliche, bereits charakterisierte Proteine in *D. discoideum* oder anderen Organismen gefunden. Die Sonde SSL850 hat hohe Ähnlichkeit mit dem *D. discoideum* Gen, das für Discoidin I kodiert. Das *alignment* der Aminosäuresequenzen weist an 134 von 266 Positionen die gleiche und an 167 von 266 Positionen eine ähnliche Aminosäure auf. Discoidin I ist ein Kohlenhydrat-bindendes Protein (Lectin) (Simpson *et al.*, 1974), das von *D. discoideum* während des Entwicklungszyklus gebildet wird und in Form eines Tetramers an den Discoidin-Rezeptor auf der Zelloberfläche bindet (Madley und Hames, 1981). Es dient der interzellulären Adhäsion (Breuer und Siu, 1981). Discoidin I bindet auch an *Klebsiella aerogenes* und *E. coli*, von denen sich *D. discoideum* ernährt (Breuer und Siu, 1981; Cooper *et al.*, 1983). Ob Discoidin I oder das Produkt des von SSL850 repräsentierten Gens an *L. pneumophila* bindet, ist nicht bekannt.

Tabelle 5: Gene, die nach 24 Stunden Koinkubation mit *L. pneumophila* in *D. discoideum* induziert sind. Angegeben sind die Sonden des Microarrays, der Signifikanz-Wert von SAM (*score*), der Faktor der Induktion und das BLAST-Ergebnis, mit dem UniProt oder GI *identifier* (*id*), der Beschreibung und dem p-Wertes des Treffers mit dem niedrigsten p-Wert. Fett gedruckt sind die bekannten *D. discoideum* Proteine, die einzelnen Sonden zugeordnet werden konnten. Für D13973 wurde keine BLAST-Suche durchgeführt, da es ein publiziertes Gen ist.

Sonde	score	Faktor	BLAST-Ergebnis		
			<i>id</i>	Beschreibung	p-Wert
SSB153	15,24	5,9	-	-	-
SSL850	13,18	4,3	72372	Ähnlich zu Discoidin I von <i>D. discoideum</i>	7,0 x 10 ⁻⁶²
SSK605	11,64	1,6	Q8G588	Ähnlich zu Phosphoribosylaminoimidazol Carboxylase von <i>Bifidobacterium longum</i>	1,1 x 10 ⁻⁷⁵
SLD769	10,85	2,4	39937493	Ähnlich zu Endopeptidase ClpB von <i>Rhodopseudomonas palustris</i>	0
SSL284	10,29	1,9	1710805	rtoA	1,2 x 10 ⁻⁹⁶
VSA218	10,28	2,1	28828171	Ähnlich zu <i>short-chain</i> Dehydrogenase von <i>Schizosaccharomyces pombe</i>	5,0 x 10 ⁻³²
SLJ833	10,13	2,1	283451	α-Mannosidase	0
SSM409	10,07	1,8	-	-	-
SLD718	9,59	1,8	AAR33509	Ähnlich zu Acetyl-CoA Hydrolase von <i>Geobacter sulfurreducens</i>	1,9 x 10 ⁻⁴¹
SSM847	9,51	2,4	-	-	-
SSJ163	8,40	3,0	AAQ87403	Ähnlich zu <i>short-chain</i> Dehydrogenase von <i>Rhizobium sp.</i> (Stamm NGR234)	8,6 x 10 ⁻²⁹
SSK692	8,14	2,4	Q86JS3	Hypothetisches Protein in <i>D. discoideum</i> , keine ähnlichen Proteine in anderen Organismen mit p-Wert < 0,98	1,9 x 10 ⁻⁷⁹
SSI119	7,15	2,6	Q8JHA7	Ähnlich zu Glutathion S-Transferase von <i>Xenopus laevis</i>	2,2 x 10 ⁻¹⁷
SSK512	6,59	2,4	Q8GSJ9	Ähnlich zu hypothetischem Protein in <i>Arabidopsis thaliana</i>	1,1 x 10 ⁻³⁸
SLH683	4,91	3,0	-	-	-
D13973	4,57	4,4	-	Dp87, Sorus-Matrix Protein	-

Die Sonden VSA218 und SSJ163 haben beide Ähnlichkeit mit einer *short-chain* Dehydrogenase; VSA218 mit einer von *Schizosaccharomyces pombe* und SSJ163 mit einer von *Rhizobium sp.* Die *alignments* der Aminosäuresequenzen haben 41 % und 43 % identische Aminosäuren. Untereinander haben die Sonden mit nur 15 % identischen Aminosäuren eine geringe Ähnlichkeit. *Short-chain* Dehydrogenasen gehören zur großen Familie von Enzymen mit ähnlicher Raumstruktur, die den Enzymklassen der

Oxidoreduktasen, Lyasen oder Isomerasen angehören und so unterschiedliche Verbindungen wie Steroide, Prostaglandine, Retinoide, aliphatische Alkohole umsetzen. Gemeinsam ist ihnen das Sequenzmotiv TGX₃GXG und die katalytische Tetrade N-S-Y-K (Belyaeva *et al.*, 2003; Persson *et al.*, 2003; Shi und Lin, 2004).

Die Phosphoribosylaminoimidazol Carboxylase (*N*⁵-CAIR-Synthetase) von *Bifidobacterium longum* zeigt mit 40 % Sequenzidentität die meisten Übereinstimmungen zu SSK605. Dieses Enzym katalysiert den sechsten Schritt der *De-novo*-Synthese von Purinnukleotiden ausgehend von 5-Phosphoribosyl-1-pyrophosphat (Lehninger, 2001).

SLD769 hat große Ähnlichkeit zu der Endopeptidase ClpB von *Rhodopseudomonas palustris*. Der Anteil der identischen Aminosäuren liegt bei 60 % und 77 % der Aminosäuren sind ähnlich. ClpB gehört zur Familie der molekularen Chaperone Clp/Hsp100. Anders als die *heat-shock* Proteine 60 und 70 sind die Chaperone dieser Familie in der Lage Protein-Aggregate, die durch Hitze-Schock oder andere Arten von Stress entstehen, aufzulösen (Feng und Gierasch, 1998).

Eine noch größere Übereinstimmung zeigen SLD718 und die Acetyl-CoA Hydrolase von *Geobacter sulfurreducens*. 66 % der 107 Aminosäuren des *alignments* sind identisch und 78 % ähnlich. Das Enzym katalysiert die Hydrolyse von Acetyl-CoA. Buu *et al.* (2003) spekulieren, dass es in *S. cerevisiae* die intrazellulären Acetyl-CoA- und Coenzym A-Konzentrationen reguliert. Seine Funktion ist jedoch noch unklar.

Die BLAST-Suche mit der Sonde SSI119 ergab, dass die Glutathion S-Transferase von *Xenopus laevis* die größte Ähnlichkeit hat. Der p-Wert des BLAST-Ergebnisses ist mit $2,2 \times 10^{-17}$ zwar nicht sehr hoch, jedoch verteilen sich die 69 identischen und 98 ähnlichen Aminosäuren über die gesamten 194 Aminosäuren der Glutathion S-Transferase. Dennoch ist fraglich, ob das von SSI119 repräsentierte Gen für ein Protein kodiert, das wie die Glutathion S-Transferase die Konjugation von reduziertem Glutathion mit zahlreichen Elektrophilen katalysiert. In Säugern dient die Glutathion S-Transferase dazu, die Zelle vor oxidativem Stress zu schützen, wie er z.B. von Karzinogenen und cytotoxischen Chemotherapeutika hervorgerufen wird (Hayes und Strange, 1995; Strange *et al.*, 2000; Coles und Kadlubar, 2003).

Für die Sonden SSK692 und SSK512 konnte nur Ähnlichkeit zu hypothetischen Proteinen von *D. discoideum* oder anderen Organismen gefunden werden, für die keine weiteren Informationen vorhanden waren.

Für vier der 16 Sonden wurden keine signifikanten BLAST-Treffer erhalten.

Für die Sonde D13973 war eine Annotation nicht nötig, da es sich um das publizierte Gen *Dp87* handelt (Ozaki *et al.*, 1993). *Dp87* kodiert für das 83 kDa große SP75, eines der neun häufigen Proteine der Sporenhülle. Im Entwicklungszyklus von *D. discoideum* beginnt die Transkription von *Dp87* in der Aggregationsphase und das gebildete Protein ist zunächst an das ER gebunden, bevor es später im Entwicklungszyklus über Präsporen-spezifische Vakuolen in die Sporen-Zwischenräume ausgeschieden wird (Nakao *et al.*, 1994; West *et al.*, 1996).

3.5.1.2 Annotation der Sonden reprimierter Gene

Von den Sonden der reprimierten Gene konnte nur SLE471 einem charakterisierten *D. discoideum* Gen zugeordnet werden (Tabelle 6). SLE471 ist identisch mit *cotB*, das für das *spore coat* Proteins SP70 kodiert. Die extrazelluläre Sporenhülle besteht bei *D. discoideum* aus den drei Proteinen SP96, SP70 und SP60, die von den Genen *cotA*, *cotB* und *cotC* kodiert werden. Im Entwicklungszyklus von *D. discoideum* werden die Proteine nach der Aggregation synthetisiert und während der *slug*-Phase in Vesikeln der Präsporen gelagert, bevor sie sekretiert werden (Fosnaugh und Loomis, 1993).

Die Sonden SSL818, SSL828 und SSL579 haben teilweise sehr hohe Ähnlichkeit mit den kleinen Calcium-bindenden Proteinen (CBP) 4a, 4b und 2. Dorywalska *et al.* (2000) entdeckten CBP4a und CBP4b während eines *yeast two hybrid screens* mit CBP1 als Köder. Beide Proteine sind 162 Aminosäuren lang und haben eine Aminosäure-Sequenzidentität von 90 %. Sie enthalten die Konsensus-Sequenzen für vier EF-Hand Motive und interagieren Ca^{2+} -abhängig mit CBP1. SSL818 und CBP4b haben 89 % identische und 93 % ähnliche Aminosäuren; bei SSL828 und CBP4a sind es entsprechend 84 % und 89 %. Die Unterschiede zwischen SSL579 und CBP2 sind deutlich größer. Der Anteil der identischen und gleichartigen Aminosäuren beträgt nur 36 % bzw. 54 %. Das Calcium-bindende Protein 2 ist 168 Aminosäuren lang und weist ebenfalls vier Konsensus-Sequenzen für vier EF-Hand Motive auf (Andre *et al.*, 1996). SSL579, SSL818 und SSL828 sind auch nicht identisch mit vier weiteren Calcium-bindenden Proteinen, die von Sakamoto *et al.* (2003) kürzlich beschrieben worden sind.

Die Sonde VSE168 hat große Übereinstimmungen mit β' -COP von *Bos taurus* (Stenbeck *et al.*, 1993). Von den 638 Aminosäuren des *alignments* sind 384 identisch (54 %) und 73 % ähnlich. COPs (*coat protein*) sind Proteine, die vergleichbar mit Clathrin, Vesikel umhüllen, die in den intra-Golgi Transport sowie den Transport zwischen Golgi und ER involviert sind

(Lowe und Kreis, 1998). Es sind sieben Untereinheiten (α , β , β' , γ , δ , ε und ξ) des sogenannten *coatomers* bekannt (Stenbeck *et al.*, 1993).

Tabelle 6: Gene, die nach 24 Stunden Koinkubation mit *L. pneumophila* reprimiert sind. Angegeben sind die Sonden des Microarrays, der Signifikanz-Wert von SAM (*score*), der Faktor der Repression und das BLAST-Ergebnis, mit dem UniProt oder GI *identifier* (*id*), der Beschreibung und dem p-Wertes des Treffers mit dem niedrigsten p-Wert. Fett gedruckt ist das *D. discoideum* Gen *cotB*, das eindeutig identifiziert werden konnte.

Sonde	score	Faktor	BLAST-Ergebnis		
			id	Beschreibung	p-Wert
VSE168	-11,23	2,3	P35605	Ähnlich zu β' -COP von <i>Bos taurus</i>	$5,1 \times 10^{-189}$
SSC645	-10,96	2,6	-	-	-
SSC158	-10,67	2,9	Q22254	Ähnlich zum T06D8.9 Protein von <i>C. elegans</i>	$7,5 \times 10^{-40}$
SSL828	-10,44	2,5	Q9GYE0	Ähnlich zu <i>calcium binding protein 4a</i> von <i>D. discoideum</i>	$1,9 \times 10^{-30}$
SSC374	-10,25	2,3	-	-	-
SSL818	-10,18	2,7	9864392	Ähnlich zu <i>calcium binding protein 4b</i> von <i>D. discoideum</i>	$2,0 \times 10^{-62}$
VSG118	-10,14	1,7	-	-	-
SSF803	-10,08	2,5	Q86HF1	Hypothetisches Protein von <i>D. discoideum</i>	$5,6 \times 10^{-110}$
SSB431	-9,80	3,0	2570906	Ähnlich zu Stearoyl-CoA Desaturase von <i>Mucor rouxii</i>	$1,0 \times 10^{-111}$
SSM741	-9,71	3,2	-	-	-
SLE471	-9,25	2,9	102273	<i>cotB, spore coat protein SP70</i>	0
SSL579	-8,11	2,5	1705663	Ähnlich zu <i>calcium binding protein 2</i> von <i>D. discoideum</i>	$9,0 \times 10^{-11}$
SSM558	-5,93	2,8	-	-	-

Die Stearoyl-CoA Desaturase von *Mucor rouxii* zeigt mit 49 % Sequenzidentität und 66 % ähnlichen Aminosäuren viel Übereinstimmung mit SSB431. Stearoyl-CoA Desaturasen (SCD) sind membrangebundene Nicht-Häm-Eisen Enzyme des Endoplasmatischen Retikulums, welche die Δ^9 -cis Oxidation gesättigter, Coenzym A-gekoppelter Fettsäuren katalysieren. Bevorzugte Substrate sind Palmitoyl- und Stearoyl-CoA. Die Produkte Palmitoleoyl- und Oleoyl-CoA sind Ausgangsstoffe für die Synthese von Triglyceriden, Wachsen, Cholesterylestern und Phospholipiden (Heinemann und Ozols, 2003; Miyazaki und Ntambi, 2003).

SSC158 hat 36 % Sequenzidentität zu T06D8.9, einem mutmaßlichen Membranprotein von *Caenorhabditis elegans*, über das nichts weiter bekannt ist.

Die Sonde SSF803 ist identisch mit dem hypothetischen Protein Q86HF1 von *D. discoideum*, und hat keine Ähnlichkeit zu bekannten Proteinen anderer Organismen.

Für fünf der 13 Sonden wurden keine signifikanten Treffer ($p\text{-Wert} < 0,0007$) in der Datenbank UniProt oder in der nr-Datenbank des NCBI erhalten.

3.5.1.3 Kategorisierung der differentiell exprimierten Gene

Urushihara *et al.* vom japanischen cDNA-Projekt haben eine Kategorisierung ihres nicht-redundanten cDNA-Satzes vorgenommen. Die funktionellen Kategorien basieren auf denen für *Saccharomyces cerevisiae* des Munich Information Centre for Protein Sequences (MIPS, <http://mips.gsf.de/proj/yeast/CYGD/db/index.html>). Zusätzlich wurden Kategorien für die Multizellularität von *D. discoideum* während des Entwicklungszykluses aufgenommen. Um die Sonden den Kategorien zuzuweisen, wurde zunächst eine Liste mit Schlüsselwörtern aus den Beschreibungen der BLASTX-Ergebnisse erstellt und die Schlüsselwörter Kategorien zugeordnet. Dann wurden die Sonden anhand der Schlüsselwörter ihres besten BLASTX Treffers ($p\text{-Wert} < 10^{-4}$) den Kategorien zugewiesen. Ein Drittel der Sonden konnte in zwölf funktionelle Kategorien eingeteilt werden (Urushihara *et al.*, 2004).

Tabelle 7: Funktionelle Kategorien der 5423 cDNA Sonden des *D. discoideum* Microarrays. Angegeben ist die Anzahl der Sonden in den einzelnen Kategorien und der Anteil der Kategorie am gesamten Satz der cDNAs.

Kategorie	Anzahl	Anteil [%]
1. Metabolism	757	14,0
2. Energy	64	1,2
3. Transcription	107	2,0
4. Translation	90	1,7
5. Protein destination	139	2,6
6. Cellular biogenesis and organization	53	1,0
7. Transport facilitation	96	1,8
8. Cell proliferation	63	1,2
9. Movement	67	1,2
10. Stress response and cell rescue	34	0,6
11. Signal transduction	115	2,1
12. Multicellular organization	95	1,8
13. Retrotransposon and plasmid proteins	13	0,2
14. Classification uncertain	2377	43,8
15. Unclassified proteins	1340	24,7
16. Error	13	0,2

Tabelle 7 zeigt die zwölf Kategorien und die Anzahl der Sonden des Microarrays, die in jeder Kategorie enthalten sind, sowie ihren Anteil an den 5423 Sonden. Die Kategorien sind in bis zu neun Unterkategorien unterteilt, die eine detailliertere Beschreibung der Funktion erlauben. Darüber hinaus gibt es noch vier Kategorien, die zur Erfassung schlecht oder nicht klassifizierter Proteine sowie von Fehler dienen.

Mit Hilfe dieser Kategorien können die Änderungen der Genexpression nach Infektion mit *L. pneumophila* für alle cDNA-Sonden auf einer funktionellen Ebene betrachtet werden. Von den 1218 differentiell exprimierten Genen sind 1121 cDNAs. Davon sind 566 induziert und 555 reprimiert. Es wurde ermittelt, in welchen Unterkategorien der Anteil der regulierten Gene an den induzierten bzw. reprimierten Genen mehr als zwei mal so hoch ist, wie der Anteil dieser Unterkategorie an dem gesamten cDNA-Satz. Der Schwellenwert zwei wurde willkürlich gewählt.

Tabelle 8: Überrepräsentierte Unterkategorien der 566 induzierten Gene. Aufgeführt sind die Kategorien, bei denen die induzierten Gene um mehr als den Faktor zwei überrepräsentiert sind. *der Überrepräsentation

alle cDNA-Sonden		induzierte Gene		Faktor*	Unterkategorie
Anzahl	Anteil [%]	Anzahl	Anteil [%]		
70	1,3	18	3,2	2,5	1.3 Metabolism: nucleotide metabolism
6	0,1	2	0,4	3,2	2.1 Energy: glycolysis and gluconeogenesis
9	0,2	2	0,4	2,1	4.2 Translation: translation factors
16	0,3	7	1,2	4,2	6.5 Cellular biogenesis and organization: nucleus
37	0,7	8	1,4	2,1	9.1 Movement: cytoskeletal components
14	0,3	3	0,5	2,1	10.2 Stress response and cell rescue: Stress response
3	0,1	1	0,2	3,2	10.3 Stress response and cell rescue: DNA repair
2	0,0	1	0,2	4,8	11.2 Signal transduction: Second messenger formation

Wie man in Tabelle 8 sieht, gibt es acht Unterkategorien, bei denen die induzierten Gene überrepräsentiert sind. So sind in der Unterkategorie 6.5 *Cellular biogenesis and organization: nucleus* z.B. sieben Gene induziert, was einem Anteil von 1,2 % der induzierten

Gene entspricht. Die Unterkategorie macht jedoch nur 0,3 % aller cDNA-Sonden aus. Sie ist bei den induzierten Genen um den Faktor 4,2 überrepräsentiert.

Auch das Auftreten von 18 induzierten Genen der Unterkategorie 1.3 *Metabolism: nucleotide metabolism* lässt vermuten, dass es sich hier tatsächlich um einen Stoffwechselweg handelt, der als Ganzes aktiviert worden ist, da ein Viertel der Gene induziert ist.

Tabelle 9: Überrepräsentierte Unterkategorien der 555 reprimierten Gene. Aufgeführt sind die Kategorien, bei denen die induzierten Gene um mehr als den Faktor zwei überrepräsentiert sind. *der Überrepräsentation

alle cDNA-Sonden		reprimierte Gene		Faktor*	Unterkategorie
Anzahl	Anteil [%]	Anzahl	Anteil [%]		
35	0,6	9	1,6	2,5	1.6 Metabolism: lipid, fatty-acid and isoprenoid metabolism
2	0,0	1	0,2	4,9	2.3 Energy: tricarboxylic-acid pathway
6	0,1	2	0,4	3,3	3.1 Transcription: rRNA synthesis
78	1,4	38	6,8	4,8	4.1 Translation: ribosomal proteins
1	0,0	1	0,2	9,8	4.3 Translation: tRNA-synthetases
43	0,8	9	1,6	2,0	5.2 Protein destination: protein targeting, sorting and translocation
7	0,1	3	0,5	4,2	5.3 Protein destination: protein modification
3	0,1	1	0,2	3,3	6.1 Cellular biogenesis and organization: cell wall and plasma membrane
4	0,1	1	0,2	2,4	6.3 Cellular biogenesis and organization: endoplasmic reticulum and Golgi
6	0,1	3	0,5	4,9	6.4 Cellular biogenesis and organization: intracellular transport vesicles
7	0,1	2	0,4	2,8	8.5 Cell proliferation: other
1	0,0	1	0,2	9,8	10.1 Stress response and cell rescue: Homeostasis
14	0,3	5	0,9	3,5	10.2 Stress response and cell rescue: Stress response
22	0,4	5	0,9	2,2	12.3 Multicellular organization: Morphogenesis

Die Unterkategorie 11.2 *Signal transduction: Second messenger formation* enthält insgesamt nur zwei Gene. Daher ist der Faktor der Überrepräsentation von 4,8 wenig aussagekräftig.

Auffällig ist auch, dass von der Unterkategorie 9.1 *Movement: cytoskeletal components* mehr als 20 % der cDNA-Sonden induziert sind und von der Hauptkategorie 10 *Stress response* zwei der sechs Unterkategorien induziert sind.

Bei den überrepräsentierten Kategorien der reprimierten Gene (Tabelle 9) fällt vor allem die Unterkategorie 4.1 *Translation: ribosomal proteins* auf, wo mit 38 von 78 Genen annähernd die Hälfte der Gene reprimiert ist.

Kategorie 4 *Translation* zeichnet sich insgesamt dadurch aus, dass drei der vier Unterkategorien überrepräsentiert sind; zwei aufgrund reprimierter Gene und eine aufgrund induzierter Gene. Gleiches gilt für die Kategorien 6 *Cellular biogenesis and organization* sowie 10 *Stress response and cell rescue*, wo vier von acht bzw. zwei von sechs Unterkategorien überrepräsentiert sind.

3.5.2 Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zur Koinkubation mit *L. hackeliae*

Zusätzlich zum Vergleich von *L. pneumophila*-infizierten *D. discoideum* Zellen 24 Stunden nach Beginn der Infektion wurden zu diesem Zeitpunkt zwei weitere Vergleiche durchgeführt. Zum einen wurden *L. pneumophila*-infizierte Zellen mit Zellen verglichen, die in Anwesenheit von *L. hackeliae* in Kultur gehalten wurden. Zum anderen wurden *L. pneumophila* infizierte Zellen mit Zellen verglichen, die in Anwesenheit von *L. pneumophila* $\Delta dotA$ inkubiert wurden (siehe Abschnitt 3.8).

cDNA-Microarrays, die hier wie allgemein üblich für kompetitive Hybridisierungen zweier fluoreszenzmarkierter cDNA-Mischungen verwendet wurden, zeigen stets die Unterschiede in der Zusammensetzung der cDNA-Mischungen an. Der Vergleich von *L. pneumophila*-infizierten Zellen mit uninfizierten Zellen liefert also alle Unterschiede der Genexpression von *D. discoideum*, die als Reaktion auf den Kontakt mit *L. pneumophila*, die Aufnahme der Bakterien und die eigentliche Infektion auftreten.

Bezüglich der Infektion weisen die Arbeiten von Hägele *et al.* (2000) darauf hin, dass 84 % der von *D. discoideum* aufgenommenen *L. pneumophila*-Zellen nicht verdaut werden. Anders verhält es sich bei *L. hackeliae*, einer *Legionella*-Spezies mit verminderter Pathogenität. Von den aufgenommenen Bakterien gelangen 82 % in Phagolysosomen und werden verdaut. Der Vergleich *L. pneumophila*-infizierter Zellen mit solchen, die *L. hackeliae* verdauen, sollte daher Gene liefern, die aufgrund von Infektion statt Verdau differenziell exprimiert sind.

Gene, die durch den Kontakt mit und die Aufnahme von Legionellen reguliert sind, sollten in beiden Kulturen gleichermaßen exprimiert sein und sich daher gegenseitig aufheben.

Es wurden je zwei unabhängige RNA-Präparationen verwendet, die mittels *dye-swaps* mit vier Microarrays hybridisiert wurden. Die SAM-Analyse ergab 882 regulierte Gene, von denen 428 induziert und 454 reprimiert waren. Die jeweils zehn induzierten und reprimierten Gene mit den höchsten *scores* sind in Tabelle 10 und Tabelle 11 gezeigt.

Tabelle 10: Induzierte Gene in *D. discoideum* nach Infektion mit *L. pneumophila* mit *L. hackeliae* als Kontrolle. Angegeben sind die Sonden des Microarrays, der Signifikanz-Wert von SAM (*score*), der Faktor der Induktion und das BLAST-Ergebnis, mit dem UniProt oder GI *identifier* (*id*), der Beschreibung und dem p-Wertes des Treffers mit dem niedrigsten p-Wert. Fett gedruckt ist das bekannte *D. discoideum* Protein Discoidin I, das eindeutig identifiziert werden konnte. Für J01284 wurde keine BLAST-Suche durchgeführt, da es ein publiziertes Gen ist. Die Sonden der Gene, die auch im Vergleich von *L. pneumophila*-infizierten Zellen mit nicht-infizierten Zellen unter den zehn Genen mit den höchsten *scores* gefunden wurden, sind unterstrichen. Alle Sonden dieser Tabelle finden sich unter den 612 induzierten Genen des Vergleichs mit nicht-infizierten Zellen als Kontrolle.

Sonde	score	Faktor	BLAST-Ergebnis		
			<i>id</i>	Beschreibung	p-Wert
<u>SLI845</u>	18,5	4,7	39937493	Ähnlich zu Endopeptidase ClpB von <i>Rhodopseudomonas palustris</i>	0
<u>SSL850</u>	18,4	6,1	72372	Ähnlich zu Discoidin I von <i>D. discoideum</i>	$7,0 \times 10^{-62}$
<u>SSB153</u>	18,3	13,0	-	-	-
<u>SSJ163</u>	15,6	4,6	AAQ87403	Ähnlich zu <i>short-chain</i> Dehydrogenase von <i>Rhizobium</i> sp. (Stamm NGR234)	$8,6 \times 10^{-29}$
<u>SSM847</u>	15,3	3,7	-	-	-
J01284	14,6	3,5	-	Discoidin I	-
SLJ337	13,3	2,8	-	-	-
VSE614	12,9	3,5	6118365	Ähnlich zu PqaA von <i>Salmonella typhi</i>	$4,0 \times 10^{-37}$
SSE777	12,6	3,1	-	-	-
VSI715	11,6	3,2	-	-	-

Von den induzierten und reprimierten Genen finden sich jeweils fünf auch in der entsprechenden Tabelle des Vergleichs mit uninfizierten Zellen. Bei den induzierten Genen sind es die Sonden SSL850, SSB153, SSJ163, SSM847 und SLI845, wobei SLI845 eine Sonde ist, die wie SLD769 aus Tabelle 5 größte Ähnlichkeit mit der Sequenz der Endopeptidase ClpB von *Rhodopseudomonas palustris* hat. Die Vermutung, dass es sich um zwei Sonden für dasselbe Gen handeln könnte, wurde durch die Ergebnisse einer neuen Analyse der cDNA-Sequenzen des japanischen cDNA Projekts bestätigt, bei der diese beiden Sonden zu einem Gen zusammengefasst wurden. Sowohl beim Vergleich mit uninfizierten,

als auch bei mit *L. hackeliae* koinkubierten Zellen, finden sich jeweils beide Sonden mit ähnlichen Faktoren unter den induzierten Genen. Allerdings nicht mit dem gleichen *score*.

Bei den reprimierten Genen sind es die Sonden SLE471, SSL818, SSC158, SSM558 und SSB431, die sich auch unter den reprimierten Gene des Vergleichs mit uninfizierten Zellen finden. Die Faktoren der Induktion und Repression dieser Gene sind bei beiden Experimenten nicht gleich. So ist z.B. das Gen der Sonde SSL850 beim Vergleich mit uninfizierten Zellen 4,3fach induziert, während es im Experiment mit *L. hackeliae* 6,1fach induziert ist.

Die Sonde SSH545 der reprimierten Gene konnte als cDNA-Fragment des bekannten Gens der Cystein Proteinase 5 identifiziert werden. Von den übrigen Sonden konnten nur für VSK389 und SSB307 Proteine mit ähnlichen Sequenzen gefunden werden.

Tabelle 11: Reprimierte Gene in *D. discoideum* nach Infektion mit *L. pneumophila* mit *L. hackeliae* als Kontrolle. Angegeben sind die Sonden des Microarrays, der Signifikanz-Wert von SAM (*score*), der Faktor der Repression und das BLAST-Ergebnis, mit dem UniProt oder GI *identifier* (*id*), der Beschreibung und dem p-Wertes des Treffers mit dem niedrigsten p-Wert. Fett gedruckt sind das *D. discoideum* Gen *cotB* und die Cystein Proteinase 5, die eindeutig identifiziert werden konnten. Die Sonden der Gene, die auch im Vergleich von *L. pneumophila*-infizierten Zellen mit nicht-infizierten Zellen unter den zehn Genen mit den höchsten *scores* gefunden wurden, sind unterstrichen. Alle Sonden dieser Tabelle finden sich unter den 606 reprimierten Genen des Vergleichs mit nicht-infizierten Zellen als Kontrolle.

Sonde	score	Faktor	BLAST-Ergebnis		
			id	Beschreibung	p-Wert
<u>SLE471</u>	-20,7	6,7	102273	<i>cotB, spore coat protein SP70</i>	0
<u>SSM558</u>	-18,5	8,6	-	-	-
SSB307	-16,2	4,3	19527174	Ähnlich zu Spleißfaktor 3b von <i>Mus musculus</i>	0
SSC374	-15,8	3,5	-	-	-
SSF615	-14,7	3,3	-	-	-
<u>SSC158</u>	-13,1	4,4	Q22254	Ähnlich zum T06D8.9 Protein von <i>C. elegans</i>	$7,5 \times 10^{-40}$
<u>SSL818</u>	-12,2	4,0	9864392	Ähnlich zu <i>calcium-binding protein 4b</i> von <i>D. discoideum</i>	$2,0 \times 10^{-62}$
SSH545	-11,3	2,6	1706263	Cystein Proteinase 5 Vorläufer	$1,0 \times 10^{-157}$
VSK389	-11,0	2,4	1510110	Ähnlich zu Ubiquitin-konjugierendem Enzym von <i>Paramecium tetraurelia</i>	$4,0 \times 10^{-43}$
<u>SSB431</u>	-10,9	3,7	2570906	Ähnlich zu Stearoyl-CoA Desaturase von <i>Mucor rouxii</i>	$1,0 \times 10^{-111}$

In Abschnitt 3.6 wurde gezeigt, dass durch das Vorhandensein von 19 DNA-Microarrays für den Vergleich von *L. pneumophila*-infizierten Zellen mit uninfizierten Zellen zum 24 h Zeitpunkt sehr viele Gene als differentiell exprimiert identifiziert wurden, die nur geringe

Faktoren der Induktion oder Repression haben. Daher wurde für einen Vergleich dieses Experiments mit dem Experiment, bei dem *L. hackeliae* die Kontrolle ist, ebenfalls ein Schwellenwert von $> 1,5$ für die Faktoren der Induktion und Repression verwendet. Dadurch verringerte sich die Zahl der differentiell exprimierten Gene auf 269 bzw. 588. Es gibt 197 Gene, die in beiden Experimenten reguliert sind (Abb. 25). Davon sind 120 induziert und 77 reprimiert. Diese Gene stellen wahrscheinlich die Gene dar, die durch den Infektionsvorgang reguliert sind. Bei dem Experiment mit uninfizierten Zellen als Kontrolle beeinflussen die anderen Vorgänge wie der Kontakt mit und die Aufnahme der Legionellen nur 72 Gene. Einen größeren Effekt scheint der verminderte Verdau von *L. pneumophila* im Vergleich zu *L. hackeliae* zu haben. Vermutlich werden dadurch 187 Gene induziert und 204 reprimiert, dass bei der Infektion mit *L. pneumophila* nicht 82 %, sondern nur 16 % der Bakterien verdaut werden. Im Infektionsansatz gibt es sicherlich solche *D. discoideum* Zellen, die infiziert werden und solche, die Legionellen verdauen. Messbar sind jedoch nur die mittleren mRNA-Konzentrationen.

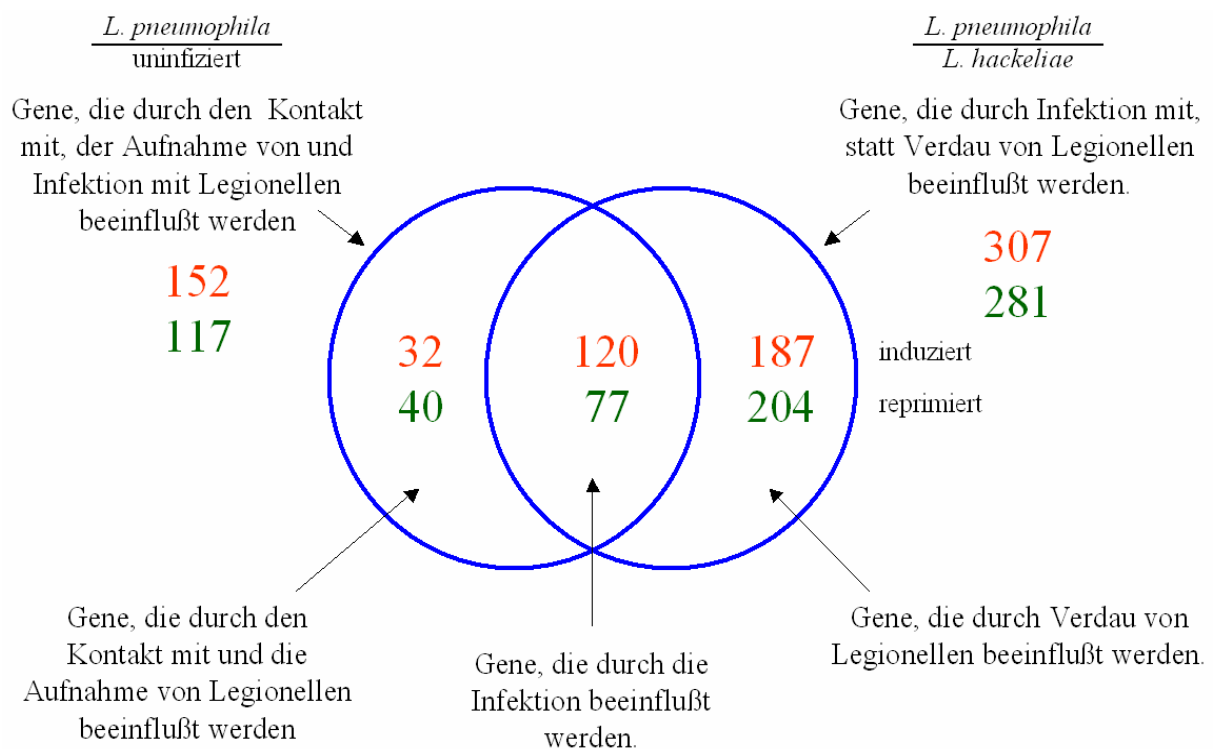


Abb. 25: Vergleich der Experimente von *L. pneumophila*-infizierten Zellen mit uninfizierten Zellen und mit *L. hackeliae* als Kontrolle. Anzahl der induzierten und reprimierten Gene, die bei beiden Infektionsexperimenten differentiell reguliert sind und solcher, die nur bei einem der Experimente zu finden sind. Es wurde ein Schwellenwert für die Faktoren der Induktion und Repression von $> 1,5$ gesetzt. Angegeben sind auch die unterschiedlichen Prozesse, von denen angenommen wird, dass sie die Genexpression beeinflussen. Oben stehen die Prozesse der beiden Experimente und unten die Prozesse der Teilmengen.

3.5.3 Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zur Koinkubation mit *L. pneumophila* Δ dotA

Der Vergleich infizierter *D. discoideum* Zellen mit Zellen, die mit *L. pneumophila* Δ dotA koinkubiert wurden, hat Gemeinsamkeiten mit dem Vergleich, bei dem *L. hackeliae* als Kontrolle verwendet wurde. Auch *L. pneumophila* Δ dotA wird von *D. discoideum* aufgenommen und verdaut. DotA ist Bestandteil des Typ IV Sekretionssystems und wichtig für die Blockade der Fusion von Phagosomen mit Lysosomen. Roy *et al.* (1998) haben gezeigt, dass sich bereits 15 Minuten nach der Aufnahme fast alle Δ dotA Legionellen in Phagolysosomen befinden. Daher wurde erwartet, dass in *D. discoideum* die Gene differentiell exprimiert sind, die von der Infektion mit Legionellen im Unterschied zum Verdau von Legionellen reguliert sind.

Tabelle 12: Induzierte Gene nach Infektion mit *L. pneumophila* mit *L. p. Δ dotA* als Kontrolle. Angegeben sind die Sonden des Microarrays, der Signifikanz-Wert von SAM (*score*), der Faktor der Induktion und das BLAST-Ergebnis, mit dem UniProt oder GI *identifier (id)*, der Beschreibung und dem p-Wertes des Treffers mit dem niedrigsten p-Wert. Fett gedruckt sind bekannte *D. discoideum* Proteine, die einzelnen Sonden zugeordnet werden konnten. Für AF140780 wurde keine BLAST-Suche durchgeführt, da es ein publiziertes Gen ist. Die Sonden der Gene, die auch im Vergleich von *L. pneumophila*-infizierten Zellen mit nicht-infizierten Zellen unter den zehn Genen mit den höchsten *scores* gefunden wurden, sind unterstrichen. Bis auf SSK833 finden sich alle Sonden dieser Tabelle unter den 612 induzierten Genen des Vergleichs mit nicht-infizierten Zellen als Kontrolle.

Sonde	score	Faktor	BLAST-Ergebnis		
			id	Beschreibung	p-Wert
VSH181	28,8	4,2	30679394	Ähnlich zu exprimiertem Protein von <i>Arabidopsis thaliana</i>	$1,0 \times 10^{-129}$
SSK833	26,7	5,3	Q8MP80	Ähnlich zu hypothetischem Protein von <i>D. discoideum</i>	$1,0 \times 10^{-67}$
AF140780	25,3	3,0	-	Countin	
VSE427	25,2	4,1	Q95US4	Ähnlich zu Glykoprotein 130 von <i>D. discoideum</i>	$9,7 \times 10^{-30}$
<u>SSB153</u>	22,4	9,2	-	-	-
SSE890	22,2	3,3	Q815U0	Ähnlich zu vermeintlich Kinesin-verwandtem Protein von <i>Plasmodium falciparum</i>	$4,5 \times 10^{-5}$
SLG851	22,0	8,0	Q04503	Präsporen Protein Dp87 Vorläufer	$1,0 \times 10^{-234}$
SSK761	21,3	6,2	SYHH_HUMAN	Ähnlich zu einem Histidyl-tRNA Synthetase Homolog von <i>Homo sapiens</i>	$1,4 \times 10^{-77}$
SLB478	19,9	5,5	RN14_MOUSE	Ähnlich zu RING Finger Protein 14 von <i>Mus musculus</i>	$2,1 \times 10^{-26}$
SSL437	19,9	5,0	-	-	-

Auch für dieses Experiment wurden je zwei unabhängige RNA-Präparationen verwendet, die mittels *dye-swaps* mit vier Microarrays hybridisiert wurden. Die SAM-Analyse ergab 1268 differentiell exprimierte Gene, von denen 584 induziert und 684 reprimiert waren. Die jeweils zehn induzierten und reprimierten Gene mit den höchsten *scores* sind in Tabelle 12 und Tabelle 13 gezeigt.

Tabelle 13: Induzierte Gene nach Infektion mit *L. pneumophila* mit *L. p. ΔdotA* als Kontrolle. Angegeben sind die Sonden des Microarrays, der Signifikanz-Wert von SAM (*score*), der Faktor der Repression und das BLAST-Ergebnis, mit dem UniProt oder GI *identifier* (*id*), der Beschreibung und dem p-Wertes des Treffers mit dem niedrigsten p-Wert. Fett gedruckt sind bekannte *D. discoideum* Proteine, die einzelnen Sonden zugeordnet werden konnten. Die Sonden der Gene, die auch im Vergleich von *L. pneumophila*-infizierten Zellen mit nicht-infizierten Zellen unter den zehn Genen mit den höchsten *scores* gefunden wurden, sind unterstrichen. Alle Sonden dieser Tabelle finden sich unter den 606 reprimierten Genen des Vergleichs mit nicht-infizierten Zellen als Kontrolle.

Sonde	score	Faktor	BLAST-Ergebnis		
			id	Beschreibung	p-Wert
SLJ376	-21,7	3,4	Q9UAG7	Flavohämoglobin	$7,3 \times 10^{-209}$
SSL389	-20,0	2,5	-	-	-
SSM424	-19,0	3,6	Q7N194	Ähnlich zu Xaa-Pro Aminopeptidase von <i>Photorhabdus luminescens</i>	$6,9 \times 10^{-62}$
SSL845	-18,1	4,1	-	-	-
VSD222	-16,8	2,0	Q91652	Ähnlich zu Gen 16 Protein von <i>Xenopus laevis</i>	$8,1 \times 10^{-36}$
VSH732	-16,4	3,5	Q9ABM7	Ähnlich zu einer Oxidoreduktase der Eisen/Ascorbat Familie von <i>Caulobacter crescentus</i>	$2,9 \times 10^{-9}$
VSH352	-16,3	2,0	Q9IA75	Ähnlich zu ribosomalem Protein L17 von <i>Paralichthys olivaceus</i>	$1,9 \times 10^{-43}$
<u>SSC158</u>	-16,3	4,7	Q22254	Ähnlich zum T06D8.9 Protein von <i>C. elegans</i>	$7,5 \times 10^{-40}$
VSI401	-15,4	3,7	Q8T690	ABC Transporter AbcG3	0
SLG489	-15,2	2,2	O21042	Cytochrom Oxidase Untereinheit 1 und 2	$1,0 \times 10^{-203}$

Es gibt es nur zwei Übereinstimmungen mit den zehn induzierten und reprimierten Genen mit den höchsten *scores* des Experiments mit uninfizierten Zellen als Kontrolle, nämlich Sonde SSB153 bei den induzierten und SSC158 bei den reprimierten Genen. Zwei der Sonden induzierter Gene konnten bekannten *D. discoideum* Proteinen zugeordnet werden. Countin, ein sekretierter Faktor (Brock und Gomer, 1999), der die Zellzahl der Aggregate während des Entwicklungszyklus beeinflusst und Dp87, ein Protein aus Präsporen, das Ähnlichkeit zu den Proteinen SP70 und SP96 der Sporenhülle hat (Ozaki *et al.*, 1993). Von den reprimierten Genen konnten die Sonden SLJ376, VSI401 und SLG489 als cDNA-Sequenzen des

Flavohämoglobins, des ABC-Transporters G3 und der Cytochrom Oxidase identifiziert werden.

Der Vergleich der Experimente mit uninfizierten Zellen und *L. pneumophila* $\Delta dotA$ als Kontrolle wurde wie in Abschnitt 3.6 mit einem Schwellenwert von 1,5 durchgeführt. Das Ergebnis ist ähnlich zu dem in Abb. 25 gezeigten. Es gibt 193 Gene, die bei beiden Experimenten differentiell exprimiert sind und 76 Gene, die spezifisch für uninfizierte Zellen als Kontrolle sind. Mit 649 sind es deutlich mehr Gene, die für *L. pneumophila* $\Delta dotA$ als Kontrolle spezifisch sind (Abb. 26).

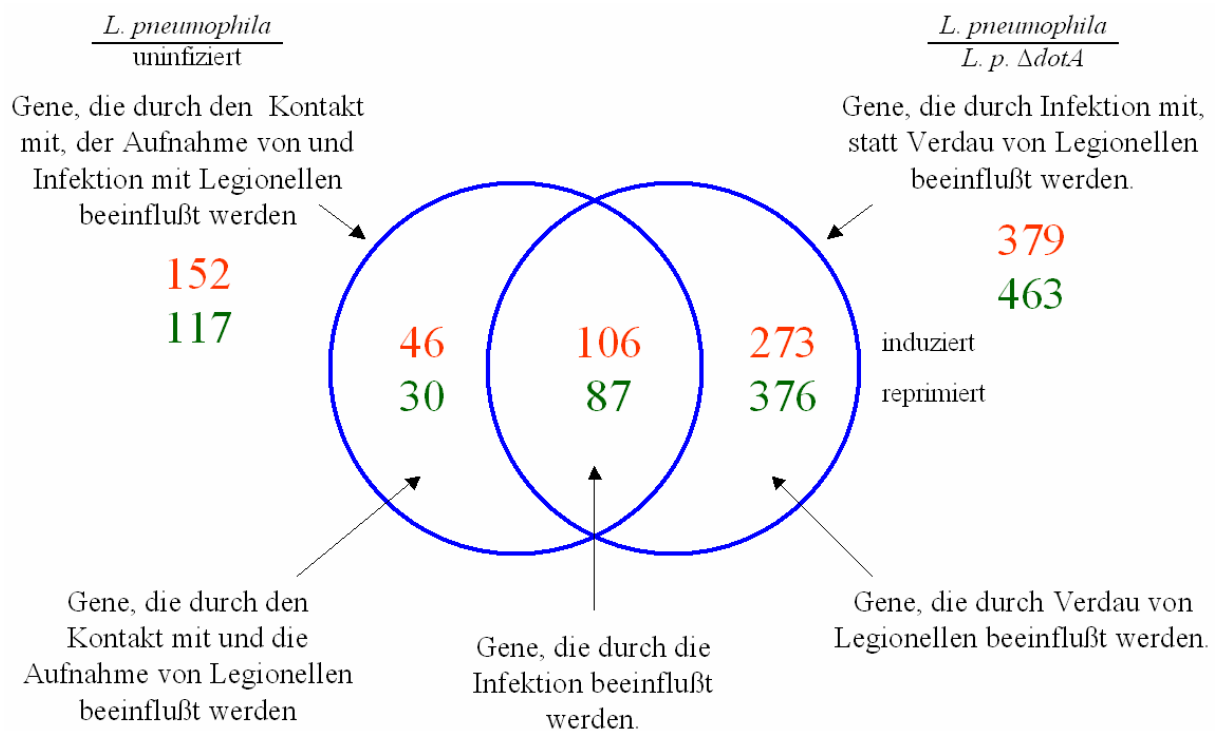


Abb. 26: Vergleich der Experimente von *L. pneumophila*-infizierten Zellen mit uninfizierten Zellen und mit *L. p. $\Delta dotA$ als Kontrolle.* Anzahl der induzierten und reprimierten Gene, die bei beiden Infektionsexperimenten reguliert sind und solcher, die nur bei einem der Experimente zu finden sind. Es wurde ein Schwellenwert für die Faktoren der Induktion und Repression von $> 1,5$ gesetzt. Angegeben sind auch die unterschiedlichen Prozesse, von denen angenommen wird, dass sie die Genexpression beeinflussen. Oben stehen die Prozesse der beiden Experimente und unten die Prozesse der Teilmengen.

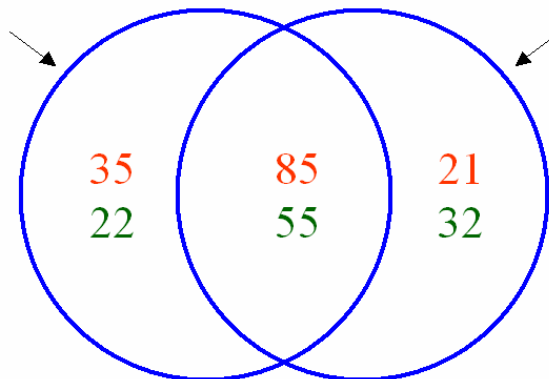
In den Vergleichen mit uninfizierten Zellen und *L. pneumophila* $\Delta dotA$ als Kontrolle wurde angenommen, dass jeweils die Schnittmengen die Gene darstellen, die spezifisch für den eigentlichen Infektionsvorgang sind. Daher wurden die Gene dieser Schnittmengen verglichen um zu ermitteln, ob sie gleich sind.

Schnittmenge der Vergleiche:

$$\frac{L. pneumophila}{\text{uninfiziert}}$$

$$\frac{L. pneumophila}{L. p. \Delta dotA}$$

106
87



Schnittmenge der Vergleiche:

$$\frac{L. pneumophila}{\text{uninfiziert}}$$

$$\frac{L. pneumophila}{L. hackeliae}$$

120
77

induziert
reprimiert

Abb. 27: Vergleich der Experimenten mit uninfizierten Zellen, *L. hackeliae* und *L. p. ΔdotA* als Kontrolle. Angegeben ist jeweils die Anzahl der zwischen zwei oder allen drei Experimenten übereinstimmenden, induzierten und reprimierten Gene.

Wie man in Abb. 27 sieht, sind die Schnittmengen aus Abb. 25 und Abb. 26 nicht gleich. Von den 141 induzierten Genen sind 85 (60 %) gleich, während es bei den reprimierten nur 55 von 109 (50 %) sind. Da diese Gene in allen drei Infektionsexperimenten mit *L. pneumophila* und uninfizierten Zellen, *L. hackeliae* und *L. pneumophila ΔdotA* als Kontrolle, differentiell exprimiert sind, könnten sie den minimalen Satz von Genen darstellen, die in *D. discoideum* stets als Reaktion auf eine Infektion mit *L. pneumophila* reguliert werden. Für 35 der induzierten und 18 der reprimierten Gene liegen funktionelle Kategorien vor. Siebzehn der induzierten Gene gehören der Unterkategorie 1.1 *Metabolism: amino-acid metabolism* an, was einer zweifachen Überrepräsentation dieser Unterkategorie gegenüber dem cDNA-Satz entspricht. Auffällig ist auch, dass in dieser Unterkategorie cDNA-Sonden für die Asparaginy-, Histidyl-, Prolyl- und Seryl-tRNA-Synthetase vorhanden sind. Darüber hinaus treten keine Kategorien gehäuft auf.

3.6 Zeitreihe der Infektion von *D. discoideum* mit *L. pneumophila*

Die Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zu uninfizierten Zellen wurde außerdem mit einer Zeitreihe untersucht, die zusätzlich zum Zeitpunkt 24 Stunden nach Infektion auch die Zeitpunkte 1 h, 3 h, 6 h und 48 h enthielt. Für jeden der vier zusätzlichen Zeitpunkte wurden drei RNA-Isolationen verwendet, deren fluoreszenzmarkierte cDNAs in Form von *dye-swaps* mit sechs Microarrays hybridisiert wurden. Für die Analyse des 24

Stunden Zeitpunkts wurden nur die 15 Microarrays verwendet, die damals zur Verfügung standen.

Die Identifikation der differentiell exprimierten Gene zeigt zunächst, dass es zu allen Zeitpunkten Änderungen der Genexpression gibt, jedoch unterschiedlich viele Gene betroffen sind (Tabelle 14). Eine Stunde nach Beginn der Infektion wurden nur 29 differentiell exprimierte Gene gefunden. Nach drei Stunden sind es mit 253 deutlich mehr und am 24 h Zeitpunkt wurde mit 852 Genen die größte Anzahl gefunden. Nach sechs und nach 48 h sind es mit 112 bzw. 157 deutlich weniger. Dieser Zahlenvergleich unterliegt jedoch der Einschränkung, dass für den 24 h Zeitpunkt Daten von 15 statt sechs Arrays, wie bei den anderen Zeitpunkten, verwendet wurden. Eine höhere Zahl von Messwiederholungen kann zu mehr, als signifikant bezeichneten, Genen führen. Da die 15 Microarrays jedoch vorlagen, wurden ihre Daten für die Analyse verwendet, da die Ergebnisse dadurch besser abgesichert sind.

Um den Effekt der vielen Microarrays des 24 h Zeitpunktes zu untersuchen, wurden nach Ermittlung der signifikant differentiell regulierten Gene verschiedene Schwellenwerte für die Faktoren der Induktion oder Repression gesetzt. Es wurde vermutet, dass die Genliste des 24 h Zeitpunktes viele schwach induzierte und reprimierte Gene enthält, die nur aufgrund der zahlreichen Microarrays als differentiell exprimiert identifiziert wurden. Tatsächlich nimmt die Genzahl des 24 h Zeitpunktes stärker ab, als die der anderen Zeitpunkte. Bei einem Schwellenwert von 1,5 sinkt die Anzahl der erhaltenen Gene beim 24 h Zeitpunkt auf 189. Dies sind 22 % der Gene ohne Schwellenwert. Bei den übrigen Zeitpunkten wirkt sich der Schwellenwert weniger stark aus. Die Genanzahl sinkt nur auf 66 bis 49 %. Die meisten Änderungen der Genexpression erfolgen also nach 24 Stunden; etwas weniger nach 3 Stunden. Deutlich weniger Gene sind nach 6 Stunden und 48 Stunden reguliert, während nach einer Stunde kaum Änderungen feststellbar sind (Tabelle 14).

Tabelle 14: Anzahl der differentiell exprimierten Gene je Zeitpunkt der Zeitreihe. Gezeigt ist die Anzahl der differentiell exprimierten Gene der fünf Zeitpunkte, wie sie mit SAM identifiziert wurden (ohne Schwellenwert) und mit einem zusätzlichen Schwellenwert für den Faktor der Induktion oder Repression von $> 1,5$. Angegeben ist auch, wie viel Prozent der von SAM gefundenen Gene mit dem Schwellenwert erhalten werden.

Zeitpunkt [h]	Anzahl der differentiell exprimierten Gene	
	ohne Schwellenwert	Schwellenwert 1,5*
1	29	19 (66 %)
3	253	129 (51 %)
6	112	50 (45 %)
24	852	189 (22 %)
48	157	77 (49 %)

Die differentiell exprimierten Gene wurden für eine Clusteranalyse verwendet, um zu ermitteln, welche Gene im Verlauf der Infektion in der gleichen Weise reguliert werden. In der Clusteranalyse werden Messwerte aufgrund ihrer Ähnlichkeit untereinander zu Haufen (engl. *cluster*) gruppiert und diese Cluster dann als Repräsentanten der in ihnen enthaltenen Messwerte verwendet. Der Vorteil dieser Methode ist eine teilweise erhebliche Reduktion der Datenmenge, wodurch oft eine Interpretation erst möglich wird. Die Annahme ist dabei, dass Gene, die ähnlich reguliert sind, an denselben Prozessen beteiligt sein könnten.

Die Clusteranalyse wird häufig dann bei Microarray-Daten eingesetzt, wenn mehr als eine experimentelle Bedingung im Vergleich zu einer Kontrolle untersucht wird. Dies ist zum Beispiel bei einer Zeitreihe der Fall (Iyer *et al.*, 1999). Dabei liegen für jedes Gen, das auf dem Microarrays repräsentiert ist, die Faktoren der Induktion oder Repression zum jeweiligen Zeitpunkt vor. Die Faktoren der einzelnen Gene können dann farbkodiert gegen die experimentelle Bedingung aufgetragen werden (siehe Abb. 28).

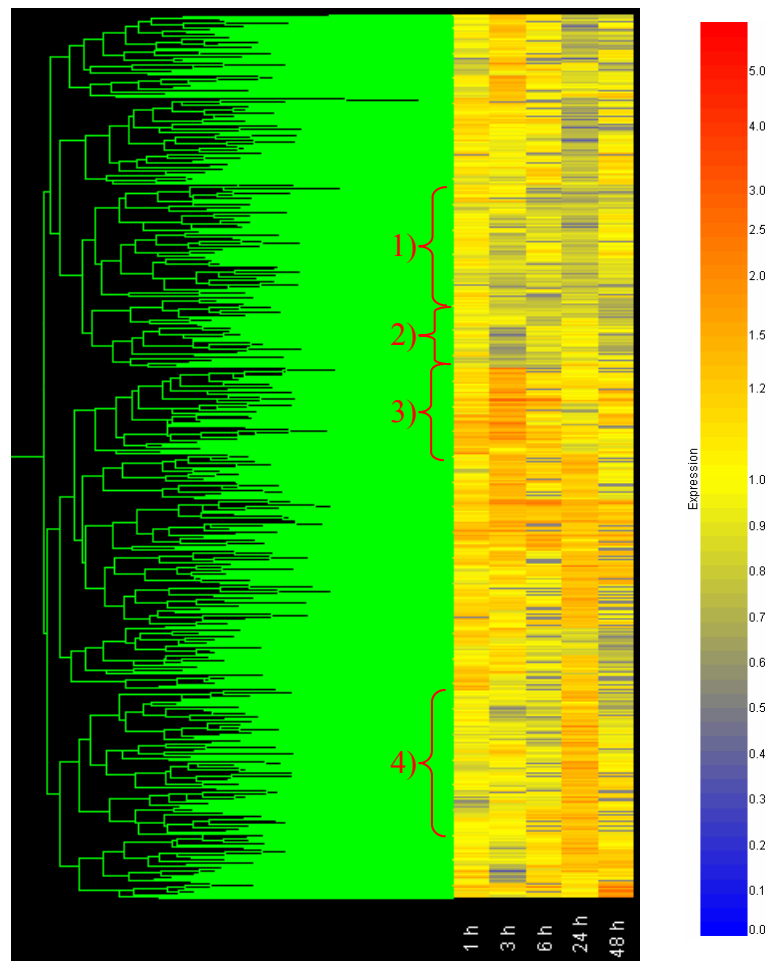


Abb. 28: Clustering der differentiell exprimierten Gene der Zeitreihe. Standard Korrelation von GeneSpring 6.1. Die 1133 Gene sind als farbige Striche dargestellt, wobei die Farbe den Faktor der Induktion (rot) oder Repression (blau) kodiert, die man der nicht-logarithmischen Farbskala entnehmen kann. Nicht regulierte Gene sind gelb dargestellt. Auf der linken Seite ist in grün ein Dendrogramm der Gene dargestellt. Die Zahlen markieren Cluster auf die im Text eingegangen wird.

Von den insgesamt 1403 Genen, die zu den fünf Zeitpunkten gefunden wurden, sind einige zu mehreren Zeitpunkten reguliert. Daher reduziert sich die Zahl der Gene auf 1133. In Abb. 28 sieht man mehrere große Cluster von Genen, die sich teilweise deutlich von den benachbarten Clustern unterscheiden, teilweise aber auch fließend in einander übergehen.

So fällt zum Beispiel in der oberen Hälfte der Abbildung ein Cluster auf, dessen Gene zu den Zeitpunkten 3 h, 6 h, 24 h und 48 h reprimiert sind (1). Darunter liegt ein Cluster bei dem vorrangig 3 Stunden nach Infektionsbeginn Gene induziert wird (2). Dieser Cluster geht fließend in einen über, wo zusätzlich auch beim 24 h Zeitpunkt Gene induziert sind (3). Einige Gene sind sogar zu allen Zeitpunkten induziert.

Der vierte Cluster ist durch induzierte Gene nach 24 Stunden gekennzeichnet (4); zu den anderen Zeitpunkten sind die Gene teil induziert, teils reprimiert und oft nicht reguliert.

Die Hauptcluster, sowie Untercluster wurden auf Häufung von Genen gleicher Kategorien untersucht. Dazu wurden auch die Beschreibungen der BLAST Treffer der Sonden in der Datenbank TrEMBL verwendet. Dabei konnten jedoch keine Cluster gefunden werden, in denen auffällig viele Gene einer Kategorie vorkommen. Ein Grund dafür könnte sein, dass von den 1133 Genen nur 317 (28 %) einer der Kategorien eins bis zwölf zugeordnet werden konnten und nur 179 (16 %) von diesen einen Treffer in TrEMBL hatten.

Wie beim 24 Stunden Zeitpunkt wurde daher ermittelt, welche funktionellen Kategorien in der Zeitreihe überrepräsentiert sind (Tabelle 15). Insgesamt wurden 18 Unterkategorien identifiziert, deren Gene im Verlauf der Zeitreihe überproportional häufig differentiell exprimiert sind. Eine Unterscheidung der induzierten und reprimierten Gene kann hier nicht vorgenommen werden, da einige Gene in der Zeitreihe sowohl induziert, als auch reprimiert werden.

Vergleicht man die regulierten Kategorien der gesamten Zeitreihe mit denen des 24 h Zeitpunktes, so zeigen sich viele Übereinstimmungen. Zwölf der 18 Kategorien finden sich in beiden Listen. Dies ist auch darauf zurückzuführen, dass die Gene des 24 h Zeitpunktes einen hohen Anteil an den regulierten Genen der Zeitreihe haben. Unterschiede können darauf beruhen, dass einzelne überrepräsentierte Kategorien des 24 h Zeitpunktes in der gesamten Zeitreihe an Gewicht verlieren und daher unter den Schwellenwert von zwei für überrepräsentierte Gene fallen.

Tabelle 15: Überrepräsentierte Kategorien der Zeitreihe. Aufgeführt sind die Kategorien, deren Anteil an den differentiell exprimierten Genen mehr als doppelt so hoch ist, wie an den gesamten cDNA Sonden. *der Überrepräsentation

alle cDNA-Sonden		differentiell exprimierte Gene		Faktor*	Kategorie
Anzahl	Anteil [%]	Anzahl	Anteil [%]		
35	0,65	14	1,69	2,6	1.6 Metabolism: lipid, fatty-acid and isoprenoid metabolism
6	0,11	2	0,24	2,2	2.1 Energy: glycolysis and gluconeogenesis
6	0,11	2	0,24	2,2	3.1 Transcription: rRNA synthesis
9	0,17	3	0,36	2,2	4.2 Translation: translation factors
28	0,52	10	1,21	2,3	5.1 Protein destination: Protein folding and stabilization
7	0,13	3	0,36	2,8	5.3 Protein destination: protein modification
3	0,06	1	0,12	2,2	6.1 Cellular biogenesis and organization: cell wall and plasma membrane
3	0,06	1	0,12	2,2	6.2 Cellular biogenesis and organization: centromere and chromosome structure
6	0,11	2	0,24	2,2	6.4 Cellular biogenesis and organization: intracellular transport vesicles
16	0,30	5	0,60	2,0	6.5 Cellular biogenesis and organization: nucleus
8	0,15	3	0,36	2,5	6.6 Cellular biogenesis and organization: mitochondrion
3	0,06	1	0,12	2,2	7.7 Transport facilitation: Drug
9	0,17	3	0,36	2,2	8.4 Cell proliferation: meiosis and sexual cycle
37	0,68	13	1,57	2,3	9.1 Movement: cytoskeletal components
14	0,26	8	0,97	3,7	10.2 Stress response and cell rescue: Stress response
3	0,06	1	0,12	2,2	10.3 Stress response and cell rescue: DNA repair
2	0,04	1	0,12	3,3	10.6 Stress response and cell rescue: Other
22	0,41	11	1,33	3,3	12.3 Multicellular organization: Morphogenesis

3.6.1 Annotation der Sonden der Kategorien mit regulierten Genen

Die insgesamt 84 Sonden der neun Hauptkategorien mit auffällig vielen regulierten Genen wurden mit BLAST-Suchen, wie in Abschnitt 3.6.1.1 beschrieben annotiert. Die Ergebnisse sind in Tabelle 16 dargestellt.

Die Gene von sechs Hauptkategorien zeigen gemischte Expressionsmuster, in dreien dominieren induzierte bzw. reprimierte Gene. Überwiegend Induktion findet sich in den Kategorien 2.1 (*Energy: glycolysis and gluconeogenesis*) und 9.1 (*Movement: cytoskeletal components*). In der Kategorie 5 (*Proteine destination*) sind die Gene hauptsächlich reprimiert.

Ähnliche Expression während der Infektion innerhalb einer Kategorie zeigen die Sonden SLE371 und SLE621 aus Kategorie 1.6, sowie VSF392, VSG238 und VSG622 aus 6.6. Bei VSF392 und VSG238 handelt es sich um zwei Sonden, die beide Teil des offenen Leserahmens (ORF) 1740 darstellen, jedoch 973 bp auseinander liegen. Identische Expression wird also erwartet. Dies trifft nicht exakt zu. Während die Werte für 1 h, 3 h und 24 h annähernd übereinstimmen, weichen die Werte für 6 h und 48 h in der Höhe der Induktion etwas voneinander ab.

Für 23 Sonden wurde kein BLAST Ergebnis oder nur eines mit hohem p-Wert erhalten.

Tabelle 16: Annotation und Expression der Gene der regulierten Kategorien der Zeitreihe. Die Sonden sind nach Hauptkategorien gruppiert und die BLAST Ergebnisse angegeben. In rot sind die Sonden markiert, deren Expression mit quantitativer PCR überprüft wurde (siehe Abschnitt 3.9.1). Für die fünf Zeitpunkte sind die nicht logarithmischen Faktoren der Induktion bzw. Repression angegeben. Der farbige Hintergrund stellt die Faktoren zusätzlich farbkodiert dar. Nicht oder schwach regulierte Gene sind gelb, induzierte rot und reprimierte blau dargestellt. Farbkodierung: Rot: 3,59 – 5,99, Dunkelorange: 2,15 – 3,58, Orange: 1,29 – 2,15, Gelb: 0,77 – 1,28, Hellbau: 0,46 – 0,76, Blau: 0,28 – 0,45, Dunkelblau: 0,17 – 0,27.

Sonde	Kategorie	BLAST Ergebnis	Faktoren				
			1 h	3 h	6 h	24 h	48 h
SLE371	1.6	Ähnlich zu SapA	1,05	0,75	0,74	0,79	0,66
SLE621	1.6	Ähnlich zu Fettsäure Δ -5 Desaturase von <i>D. discoideum</i>	1,01	0,76	0,76	0,78	0,64
SLH284	1.6	Ähnlich zu Cholesterol Transport Protein (Niemann-Pick C Protein)	1,07	0,85	1,01	0,79	0,96
SSA403	1.6	Ähnlich zu 1-Phosphatidylinositol-4,5-Bisphosphatphosphodiesterase	1,68	0,70	0,74	0,59	0,79
SSB431	1.6	Ähnlich zu Stearoyl-CoA desaturase	1,14	1,10	0,81	0,30	0,80
SSC802	1.6	Ähnlich zu npc1 Protein (Niemann-Pick C Protein)	1,04	0,76	0,90	1,17	0,65
SSF825	1.6	Ähnlich zu Enoyl-CoA Hydratase/Isomerase	1,08	0,97	0,95	1,15	1,09
SSI311	1.6	Ähnlich zu Phosphatidylinositol-Glycan-spezifische Phospholipase d 2	0,91	0,98	0,93	0,84	0,95
SSJ362	1.6	Ähnlich zu PksB, <i>short-chain</i> Alkohol Dehydrogenase	0,82	0,89	1,09	0,85	1,09
SSK576	1.6	Ähnlich zu 1-Phosphatidylinositol-4,5-Bisphosphatphosphodiesterase	1,23	0,93	0,74	0,64	0,74
SSK895	1.6	Ähnlich zu Lipoxigenase LipA	1,74	1,46	1,19	1,32	0,89
VSC157	1.6	-	0,92	1,08	0,88	0,75	0,79
VSH849	1.6	Ähnlich zu Acetyl/Propionyl-CoA Carboxylase	0,58	0,92	1,07	1,25	1,04
VSK246	1.6	Ähnlich Saplip c Protein	0,80	0,89	1,21	1,18	1,07
SSF806	2.1	Ähnlich zu GTPase-Aktivator Protein für Ras-ähnliche GTPasen	1,07	1,38	1,12	1,55	0,97
VSC214	2.1	-	0,86	0,99	1,54	1,01	1,22

Tabelle 16: Annotation und Expression der regulierten Kategorien der Zeitreihe (Fortsetzung).

Sonde	Kategorie	BLAST Ergebnis	Faktoren				
			1 h	3 h	6 h	24 h	48 h
SSB379	3.1	Ähnlich zu vermeintlich RNA-bindendem Protein	0,90	1,05	0,95	0,70	1,31
VSC366	3.1	Ähnlich zu <i>nhp2-like protein 1</i> , ribosomales Protein	1,02	0,97	0,75	0,80	1,05
SLA327	4.2	-	0,92	1,04	0,90	0,88	0,86
SLB758	4.2	Ähnlich zu Elongations Faktor 2 (EF-2)	1,12	0,94	0,77	0,93	0,66
SSK543	4.2	<i>Heat shock cognate protein</i>	1,09	1,37	1,13	1,24	1,10
SLB321	5.1	Chaperonin 60	1,04	0,91	1,05	1,28	1,07
SLC177	5.1	-	1,02	0,90	0,85	0,59	0,88
SLC355	5.1	-	0,89	1,05	1,14	1,27	1,03
SLC492	5.1	Ähnlich zu Protein Disulfid Isomerase	1,14	0,83	0,94	0,76	1,04
SLC786	5.1	Ähnlich zu cct Chaperonin θ -Untereinheit	0,99	0,91	0,94	1,38	0,76
SLD433	5.1	Ähnlich zu RpgG, ribosomales Protein	0,97	0,64	0,95	1,01	0,72
SLH760	5.1	Ähnlich zu Protein Disulfid Isomerase	1,07	0,68	0,99	0,69	0,91
SSD292	5.1	Ähnlich Peptidyl-Prolyl cis-trans Isomerase	1,00	0,70	0,85	0,71	0,89
SSI735	5.1	Ähnlich zu T-complex Protein 1, α -Untereinheit, Chaperon	1,23	0,90	0,91	1,25	0,86
VSA634	5.1	Ähnlich zu Protein Disulfid Isomerase	0,95	0,90	0,94	0,60	1,20
SSC481	5.3	Ähnlich zu ADP-Ribosylierungsfaktor 1	0,95	0,95	0,77	1,22	1,32
SSI257	5.3	Ähnlich zu Protein Disulfid Isomerase	1,07	0,92	1,01	0,66	1,01
SSL641	5.3	Ähnlich zu ADP-Ribosylierungsfaktor 1	0,83	0,90	0,98	0,74	1,00
SSI248	6.1	-	0,31	0,47	0,70	0,40	0,91
SLB666	6.2	Ähnlich zu Ankyrin 1	1,28	0,94	0,92	1,17	0,96
SSK661	6.4	-	0,85	1,01	1,05	0,75	1,07
VSD222	6.4	Ähnlich zu vermeintlichem WD40 Protein	0,93	0,87	0,87	0,85	0,75
SLD769	6.5	Ähnlich zu Endopeptidase ClpB protein	0,95	1,42	1,22	2,30	1,59
SLJ685	6.5	Ähnlich zu COP 9 Komplex Untereinheit 7a	1,06	1,20	1,25	1,25	1,26
SSB743	6.5	Ähnlich zu Histon H2b	0,98	0,96	0,90	1,34	1,07
SSB856	6.5	Ähnlich zu Histon H2b	0,95	0,76	0,91	1,00	1,13
SSF827	6.5	Ähnlich zu <i>nonhistone chromosomal protein 6b</i>	1,02	0,82	0,88	1,32	0,92
VSF392	6.6	ORF 1740	1,26	2,24	1,30	1,06	1,81
VSG238	6.6	ORF 1740	1,25	2,28	1,14	0,98	1,56
VSG622	6.6	ORF 425	1,97	2,36	1,06	0,98	1,76
SSD743	7.7	-	1,00	0,95	1,13	0,63	1,21
SSB529	8.4	-	1,07	1,01	1,04	1,22	0,81
SSI303	8.4	Ähnlich zu <i>Schizosaccharomyces pombe</i> psi Protein	0,96	0,88	0,94	0,83	0,90
VSJ507	8.4	-	1,42	1,20	1,34	1,15	1,86
SLA211	9.1	Ähnlich zu Tubulin β -Kette	1,07	0,96	0,85	1,18	0,93
SLC782	9.1	Ähnlich zu Myosin II schwere Kette	1,66	1,96	1,20	1,20	0,98
SLD534	9.1	-	1,05	0,71	0,93	1,10	0,98
SLF203	9.1	Ähnlich zu Tubulin β -Kette	0,98	1,06	0,95	1,29	0,78
SLJ557	9.1	Ähnlich zu Tubulin β -Kette	1,02	1,50	1,21	1,37	0,97
SSA163	9.1	-	1,03	0,79	0,90	0,82	0,86
SSC105	9.1	-	1,39	1,28	1,41	0,85	1,25
SSF572	9.1	-	1,08	1,14	1,05	1,35	0,87
SSH277	9.1	Ähnlich zu Myosin leichte Kette	0,96	0,89	0,93	1,20	1,02
SSH278	9.1	Ähnlich zu Protovillin, 100 kDa Aktin-bindendes Protein	0,99	0,93	1,04	1,25	0,82
SSK875	9.1	Ähnlich zu Myosin I D schwere Kette	0,91	1,08	0,94	1,24	0,91
SSL735	9.1	Ähnlich zu Cortexillin	1,02	1,67	1,07	1,19	0,91
VSD139	9.1	Ähnlich zu hypothetischem 127 kDa Protein ähnlich zu Tenascin	0,92	0,96	0,82	1,50	1,01
SLA176	10.2	Ähnlich zu GRP94 (Endoplasmic)	1,06	0,87	1,09	1,47	0,93
SLC380	10.2	Ähnlich zu Hämoglobin-ähnlichem Flavoprotein	1,39	1,06	1,12	1,34	0,91
SLC470	10.2	Ähnlich zu Chaperonen GRP78/BiP/KAR2, HSP70 Superfamilie	1,20	1,08	0,86	0,77	0,77
SLD225	10.2	Ähnlich zu Hitze-Schock Protein 70	0,94	0,94	0,68	1,04	0,79
SLG278	10.2	Ähnlich zu 32 kDa Heat Shock Protein	1,30	1,43	1,24	1,26	1,09
SLI842	10.2	-	0,98	0,82	0,73	0,71	0,87
SLJ376	10.2	-	1,17	0,89	0,90	0,61	1,01
SSH225	10.2	-	0,88	1,05	1,01	0,66	0,95
SSF626	10.3	-	1,14	0,86	0,88	1,25	1,03
SLI623	10.6	-	0,88	0,72	0,92	1,01	1,06
SLA566	12.3	Ähnlich zu TipD (WD40 Protein)	0,69	0,48	0,91	0,77	0,79
SLB855	12.3	Ähnlich zu Discoidin I, A chain	0,96	0,68	0,58	1,61	0,92
SLE690	12.3	Ähnlich TipA	1,06	1,05	0,79	0,80	0,74
SLF308	12.3	-	0,81	0,83	0,88	0,90	1,01
SSB372	12.3	-	1,04	0,83	1,21	0,83	0,95
SSD491	12.3	-	0,77	0,73	0,92	0,96	0,70
SSF514	12.3	-	0,98	0,80	1,15	0,81	1,17
SSG451	12.3	Ähnlich zu TipC	0,94	1,04	0,89	1,20	1,07
SSK163	12.3	Ähnlich zu smlA (small aggregate formation protein)	0,84	0,94	0,81	2,20	0,83
SSL850	12.3	Ähnlich Discoidin I	0,94	1,07	1,50	4,46	1,59
VSD289	12.3	Ähnlich zu Discoidin II	0,94	0,71	0,75	1,34	0,71

3.6.2 Vergleich von SAM mit dem Lord-Test

Zur Identifikation der differentiell exprimierten Gene eines Experiments wurde das Programm Significance Analysis of Microarrays (SAM) verwendet, da es speziell für Microarrays entwickelt wurde und die Rate der erwarteten falsch Positiven berechnet. Um zu sehen, welche Ergebnisse SAM im Vergleich zu einem klassischen statistischen Test liefert, wurde für die Zeitreihe der Infektion die Genliste von SAM mit der Liste eines Lord-Tests verglichen. Die klassische Analyse wurde nach dem von Kesel *et al.* (1999) beschriebenen Schema durchgeführt (Abb. 29).

3.6.2.1 Statistische Analyse

Die fünf Schritte der statistischen Analyse sind im Folgenden aufgelistet. Dabei werden für jedes Gen die beiden Zeitpunkte miteinander verglichen, zwischen denen die größte Änderung der Genexpression stattfindet:

1. Die gemessenen Genexpressionswerte sind unverbunden, da jeweils zwei unabhängige Kulturen von *D. discoideum* miteinander verglichen werden.
2. Es wird der David-Test durchgeführt, da für den χ^2 -Test zwölf Wiederholungen ein zu geringer Stichprobenumfang sind. Bei einer Irrtumswahrscheinlichkeit von 0,001 wird festgestellt, dass die Messwerte normalverteilt sind.
3. Da die Messwerte normalverteilt sind, wird ein F-Test zur Prüfung auf Gleichheit der Varianzen mit einer Irrtumswahrscheinlichkeit von 0,001 durchgeführt. Bei 83 % der Gene liegt eine Gleichheit der Varianzen vor. Daher wird der weitere Test so durchgeführt, als ob absolute Gleichheit der Varianzen vorläge.
4. Der Stichprobenumfang beträgt zwölf und ist somit klein.
5. Es wird der Lord-Test mit einer Irrtumswahrscheinlichkeit von 0,005 durchgeführt. Es werden 1423 Gene identifiziert, die signifikant differentiell exprimiert wurden.

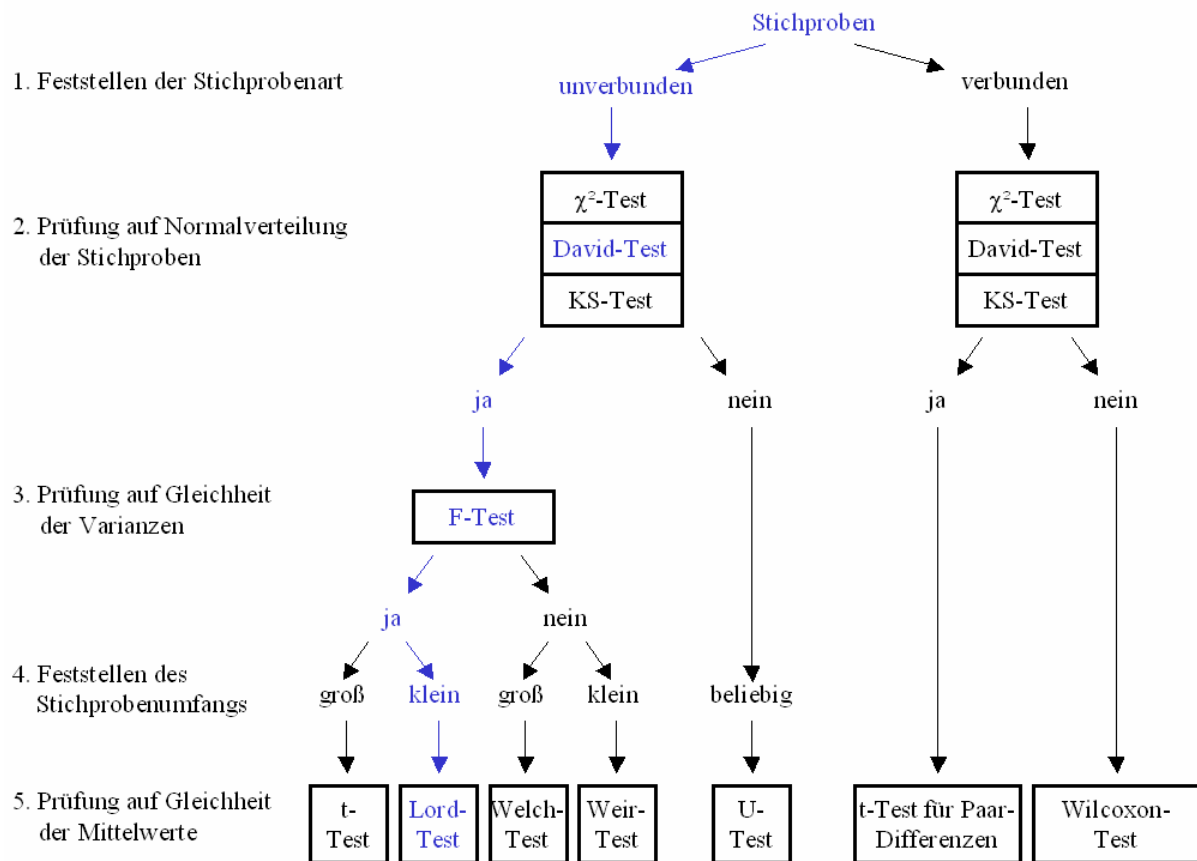


Abb. 29: Schema des statistischen Tests. Der eigentliche Signifikanztest findet im 5. Schritt statt. Die vorherigen Tests und Entscheidungen dienen dazu einen geeigneten Test auszuwählen. In Blau ist der Pfad markiert, dem beim Test gefolgt wurde (aus Kesel *et al.*, 1999).

Die Liste der 1423 Gene, die der Lord-Test lieferte, wurde mit der Genliste von SAM verglichen. Die Genliste von SAM besteht aus 1133 Genen, die zu einem der fünf untersuchten Zeitpunkte als signifikant differentiell exprimiert identifiziert wurden. Die Grundlage, auf der SAM und der Lord-Test die Gene identifizierten, ist dabei unterschiedlich. Mit dem Lord-Test wurde bestimmt, ob die Faktoren der Induktion oder Repression zweier Zeitpunkte sich signifikant unterscheiden. Dafür wurden jedes Gen die beiden Zeitpunkte mit maximaler Induktion und Repression ausgewählt. SAM hingegen ermittelt für jeden Zeitpunkt, welche Gene differentiell exprimiert sind. Je nachdem, wie die Gene reguliert sind, kann das einen Einfluss darauf haben, ob die beiden Verfahren die Gene als differentiell exprimiert identifizieren.

SAM findet in der Zeitreihe 1133 differentiell exprimierte Gene, von denen 591 auch vom Lord-Test identifiziert wurden (Tabelle 17).

Tabelle 17: Vergleich der Genlisten von SAM und dem Lord-Test. Es gibt 591 Gene, die in beiden Listen zu finden sind.

	SAM	Lord-Test
Lord-Test	591	1423
SAM	1133	591

Das sind mehr als doppelt so viele wie die 272 Übereinstimmungen, die durch eine rein zufällige Auswahl der Genlisten zu erwarten gewesen wären. Die Unterschiede können außer von den unterschiedlichen Ausgangsdaten auch durch die unterschiedlichen Bewertungskriterien der beiden Tests begründet sein. Statistische Tests wie der Lord- und der t-Test haben die Tendenz, eine geringe Abweichung der Messwerte untereinander stark zu gewichten, da die Division durch sehr kleine Standardabweichungen (t-Test) oder Spannweiten (Lord-Test) einen sehr großen Wert liefern. Bei Microarray-Experimenten sollte jedoch der Faktor der differentiellen Genexpression stärker gewichtet werden als die Reproduzierbarkeit (Smyth *et al.*, 2002). Da dies bei SAM der Fall ist, wurde die Genliste von SAM für die weitere Analyse verwendet.

3.7 Validierung der Microarray-Ergebnisse mit quantitativer PCR und Northernblots

Um die mit Microarrays gemessenen Induktionen und Repressionen von Genen zu überprüfen, wurden stichprobenartig Gene des *D. discoideum*-Microarrays ausgewählt. Es wurden Gene der Vergleiche von infizierten Zellen mit uninfizierten Zellen sowie Koinkubation mit *L. hackeliae* für quantitative PCR (qPCR) Experimente und Northernblots verwendet. Auch die Expressionsfaktoren der Zeitreihe wurden so verifiziert.

In Tabelle 18 sieht man, dass die Messungen mit qPCR in 25 von 30 Messungen die gleiche Richtung der differentiellen Expression, Induktion oder Repression zeigen. In zwei Fällen, nämlich beim 6 Stunden und 24 Stunden Wert von VSG622 weichen die Ergebnisse der beiden Methoden deutlich voneinander ab. Während mit dem Microarray mit Werten von 1,06 und 0,98 praktisch keine differentielle Expression gemessen wurde, ist mit qPCR bei 6 Stunden eine Repression um 35 % auf 0,65 und bei 24 Stunden eine Induktion um 57 % auf 1,57 gemessen worden. Die Ergebnisse des Microarrays sind hier sicherlich mit einem größeren Messfehler behaftet, da die gemessenen Faktoren bei den Wiederholungen einzelner Microarrays stärker schwankten als die unabhängigen Bestimmungen der qPCR.

Tabelle 18: Überprüfung der Microarray-Ergebnisse mittels quantitativer PCR. Vergleich der Microarray-Ergebnisse mit denen der quantitativen PCR (qPCR) für die Zeitreihe der Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zu uninfizierten Zellen. Angegeben sind die nicht-logarithmierten Mittelwerte der Faktoren der Induktion (Werte > 1) und Repression (Werte < 1) von drei Messungen für den jeweiligen Zeitpunkt.

Zeitpunkt [h]	SSL850		SSB431		SSF806	
	Microarray	qPCR	Microarray	qPCR	Microarray	qPCR
1	0,94	0,89	1,14	1,49	1,07	1,24
3	1,07	0,99	1,10	0,95	1,38	1,16
6	1,50	1,74	0,81	0,79	1,12	1,34
24	4,46	6,47	0,30	0,30	1,55	1,29
48	1,59	2,81	0,80	0,95	0,97	0,72

Zeitpunkt [h]	SSK576		SLD769		VSG622	
	Microarray	qPCR	Microarray	qPCR	Microarray	qPCR
1	1,23	1,07	0,95	1,00	1,97	1,42
3	0,93	0,82	1,42	1,89	2,36	1,51
6	0,74	0,81	1,22	1,51	1,06	0,65
24	0,64	0,66	2,30	2,79	0,98	1,57
48	0,74	0,75	1,59	1,24	1,76	1,22


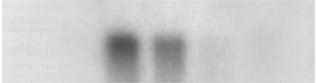

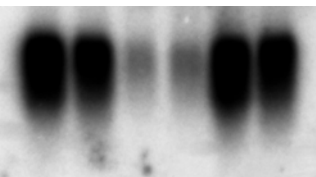
Die übrigen drei Abweichungen zwischen den Microarray- und qPCR-Ergebnissen, bei SSL850 und SSB431 drei Stunden sowie bei SLD769 eine Stunde nach Infektionsbeginn, sind auf die Messunsicherheit der beiden Methoden zurückzuführen. Die Messwerte sind stets geringfügig größer oder kleiner als eins. Die Gene sind zu diesen Zeitpunkten also vermutlich nicht reguliert.

Bei differentiell exprimierten Genen unterscheiden sich die Microarray- und qPCR-Ergebnisse häufig im Faktor der Induktion oder Repression. In 13 Fällen werden mit der qPCR höhere Faktoren gemessen. Die Abweichung der Microarray-Ergebnisse von den qPCR-Ergebnissen beträgt im Mittel 25 %. Die größte Abweichung mit 76 % gibt es beim 48 Stunden Zeitpunkt von SSL850. In neun Fällen wurde mit Microarrays höhere Faktoren gemessen; im Mittel beträgt die Abweichung 21 %.

In drei Fällen wurden mit Microarrays und qPCR exakt oder annähernd gleiche Werte gemessen.

Für die Vergleiche von *L. pneumophila* infizierten Zellen mit uninfizierten und mit *L. hackeliae* infizierten Zellen sind die Ergebnisse von qPCR und Northernblot Experimenten in Tabelle 19 gezeigt. Die Gene der Sonden SSB153 und SSJ163 sind in den mit *L. pneumophila* infizierten Zellen induziert. Dies zeigen alle drei Methoden. Lediglich die Faktoren der Induktion sind bei den Microarrays stets niedriger als bei der qPCR.

Tabelle 19: Überprüfung der Microarray-Ergebnisse. Vergleich der Microarray-Ergebnisse mit denen der quantitativen PCR (qPCR) und Northernblots für die Vergleiche von *L. pneumophila*-infizierten mit uninfizierten Zellen sowie *L. pneumophila*-infizierten mit *L. hackeliae*-infizierten Zellen. Bei den Northernblots wurden jeweils zwei Total-RNA Isolationen von uninfizierten Zellen (uninf.), mit *L. pneumophila* infizierten Zellen (*L. pneu.*) und mit *L. hackeliae* infizierten Zellen (*L. hack.*) nebeneinander aufgetragen. In Spalten zwei bis fünf sind die ermittelten Faktoren der Induktion und Repression dargestellt. Für SLE471 wurde keine qPCR durchgeführt.

Sonde	<i>L. pneumophila</i> infiziert / uninfiziert		<i>L. pneumophila</i> infiziert / <i>L. hackeliae</i> infiziert		Northernblot uninf. <i>L. pneu.</i> <i>L. hack.</i>
	Microarray	qPCR	Microarray	qPCR	
SSB153	11,50	15,35	13,00	21,06	
SSJ163	5,60	10,98	4,60	8,66	
SSL818	0,26	0,43	0,25	0,27	
SLE471	0,16	-	0,15	-	

Auch bei den Sonden SSL818 und SLE471 ist das Ergebnis konsistent; alle drei Methoden zeigen eine Repression in *L. pneumophila* infizierten Zellen.

4 Diskussion

4.1 Herstellung des *D. discoideum* DNA-Microarrays

4.1.1 Vergleich von Methoden zur Amplifikation der Sonden für die Herstellung von DNA-Microarrays

Vor der Amplifikation der 5423 cDNA-Sonden wurden verschiedene Methoden der Herstellung verglichen: 1. Amplifikation mit einem PCR-Ansatz und Reinigung mittels Ethanolpräzipitation 2. Amplifikation mit fünf PCR-Ansätzen je Sonde und Reinigung durch Ethanolpräzipitation der vereinigten Ansätze 3. Amplifikation mit einem PCR-Ansatz und Reinigung mit dem QIAvac PCR Purification Kit. Alle drei getesteten Methoden lieferten vergleichbare Signalintensitäten auf dem Microarray. Von den geringfügig höheren maximalen Signalintensitäten nach Reinigung mit QIAvac abgesehen, stellt die einfache PCR-Amplifikation mit anschließender Ethanolpräzipitation also ein schnelles und kostengünstiges Verfahren zur Herstellung von Microarray-Sonden dar. Der einzige Vorteil mehrerer PCRs je Sonde ist die Möglichkeit PCR-Reaktionen mit geringer Ausbeute ausgleichen zu können. Die DNA-Konzentration einer PCR lag meist bei 7 ng/μl, schwankte jedoch zwischen 2 und 15 ng/μl. Nach Konzentrierung der einzelnen bzw. der fünffachen PCR betrug die durchschnittliche Konzentration ca. 50 bzw. 250 ng/μl. Es konnte keine Begrenzung der maximal erzielbaren Signalintensität eines Sondenpunktes aufgrund der Sondenkonzentration beobachtet werden. Das könnte zum einen daran liegen, dass die DNA-Bindekapazität der Microarray-Oberfläche schon bei niedrigen Sondenkonzentrationen erschöpft ist. Wenn die DNA in Form einer dicht gepackten Schicht an die Microarray-Oberfläche gebunden ist, wie in Duggan *et al.* (1999) gezeigt (Abb. 30), so genügen ca. 2×10^7 Sondenmoleküle mit einer Länge von 500 bp, um einen Punkt mit 100 μm Durchmesser vollständig zu bedecken. Da das vom MicroGrid II gedruckte Volumen etwa 100 pL beträgt, wäre eine Konzentration von 100 ng/μl ausreichend, um die Microarray-Oberfläche zu bedecken.

Ein weiterer Effekt könnte sein, dass die Zahl der für die Hybridisierung zur Verfügung stehenden Sondenmoleküle die Zahl der Probenmoleküle auch dann übertrifft, wenn der Sondenpunkt nicht vollständig mit DNA bedeckt ist. Das liegt nicht nur an der limitierten Zahl von Probenmolekülen, die sich aus z.B. 10 μg totaler RNA herstellen lassen; auch die Diffusionsgeschwindigkeit der Proben in dem dünnen Film zwischen Microarray und Deckglas ist beschränkt. Dadurch stehen nur die Proben für eine Hybridisierung zur

Verfügung, die sich in unmittelbarer Nähe des entsprechenden Sondenpunktes befinden und die effektive Probenzahl verringert sich um das 10^4 bis 10^5 fache (Duggan *et al.*, 1999).

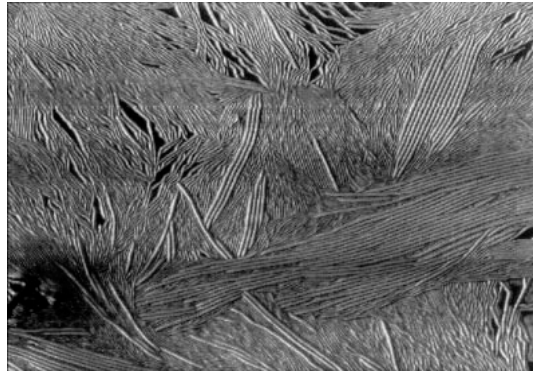


Abb. 30: Atomkraftmikroskopische Aufnahme eines hybridisierten Microarrays. Die Kantenlänge der Abbildung entspricht $2\ \mu\text{m}$ auf dem Microarray (aus Duggan *et al.*, 1999).

4.1.2 Sonden des *D. discoideum* DNA-Microarrays

Die Sonden, aus denen das *D. discoideum* Microarray besteht, basieren auf den derzeit bekannten Genen von *D. discoideum* sowie der cDNA-Bibliothek des japanischen cDNA Projekts. Zu den 450 Sonden publizierter Gene ist am meisten Information vorhanden. Außer der vollständigen Gensequenz ist für 264 Gene (59 %) auch eine Veröffentlichung vorhanden. Die 5423 cDNA-Sonden wurden vom 5'- und 3'-Ende her sequenziert. Von vielen cDNAs ist die vollständige Sequenz bekannt, da die Sequenzen beider Enden überlappten. Die 5423 Sonden stellen einen nicht redundanten Satz von cDNAs dar. Sie wurden nach Assemblierung von 45 386 cDNAs zu zusammenhängenden Sequenzen, die jeweils ein Gen repräsentieren, ausgewählt. Die cDNAs wurden aus RNA der vegetativen und der *slug*-Phase von *D. discoideum* gewonnen. Dadurch ist sichergestellt, dass zumindest ein Teil der repräsentierten Gene während der vegetativen Phase exprimiert wird, in der die Infektionsexperimente durchgeführt wurden. Die 5423 Sonden bilden zwar nur etwa die Hälfte des ca. 12 000 Gene umfassenden Genoms ab, sie repräsentieren jedoch die Gene, von denen die meisten mRNA-Kopien erzeugt wurden.

Außer den publizierten Genen und den cDNAs standen zum Zeitpunkt der Sondenherstellung für das Microarray keine weiteren Sequenzen zur Verfügung. Das Genomprojekt befand sich in der Phase der Rekonstruktion der Chromosomen und Genvorhersage. Inzwischen werden ca. 12 000 Gene für das Genom vorhergesagt. Zur Zeit werden Oligonukleotide für die Amplifikation von Sonden ausgewählt und diese Sonden für die nächste Generation des *D. discoideum* Microarrays amplifiziert.

4.2 Das automatisierte Auswerteverfahren

Durch die Entwicklung des automatisierten Auswerteverfahrens ist eine schnelle Identifikation der differentiell exprimierten Gene eines Experiments möglich. Die Normalisierung und die eigentliche Identifikation der Gene wird dabei mit der frei verfügbaren Software R, Modulen des BioConductor Projekts und SAM durchgeführt. Die gemessenen Signalintensitäten von Microarray-Experimenten müssen stets normalisiert werden, da Experiment- und Kontroll-cDNA mit zwei unterschiedlichen Farbstoffen markiert werden, die unabhängig von einander sind. Durch die Normalisierung werden Effekte wie unterschiedliche Fluoreszenz-Ausbeute und Markierungseffizienz ausgeglichen. Dazu können die Gesamtintensität des Farbkanäle, lineare Regression oder eine nicht-lineare, lokal gewichtete Methode wie LOWESS (*locally weighed least squares*) verwendet werden. Die LOWESS-Methode ist globalen oder linearen Normalisierungsmethoden überlegen, da die Unterschiede der Farbstoffe intensitätsabhängig sind (Yang *et al.*, 2002). Inzwischen stellt die LOWESS-Normalisierung die Standardmethode dar. Die von SAM verwendete Statistik zur Bestimmung der differentiell exprimierten Gene wurde von Lönnstedt und Speed mit anderen Methoden verglichen und hat sich gegenüber dem t-Test als weit überlegen herausgestellt (Smyth *et al.*, 2002). Durch die Verwendung der freien Programme konnten Mittel und Zeit gespart werden. Mit Hilfe kostenloser neuer Versionen und Erweiterungen kann das Auswerteverfahren stets aktuell gehalten werden.

Die Messspanne von DNA-Microarrays ist zu gering um die relativen Konzentrationen häufiger und seltener mRNA-Spezies zu messen. Bei der Verwendung nur eines Messbereichs könnten entweder nur die Sonden mit hohen oder die Sonden mit niedrigen Signalen korrekt gemessen werden. Für den jeweils anderen Teil könnte dann eine etwaige differentielle Expression nicht zuverlässig bestimmt werden. Daher wurden die Microarrays mit zwei Messbereichen, einem für hohe und einem für mittlere und niedrige Signalintensitäten, gescannt. ArrayTools wählt für jede Sonde automatisch den geeigneten Messbereich. Dadurch konnte die Messspanne des Microarray-Systems vergrößert werden.

4.3 Validierung des Microarray-Systems

4.3.1 Das SpotReport System

Mit Hilfe des SpotReport Systems konnten die gemessenen Quotienten der mRNA-Konzentrationen mit den tatsächlichen verglichen werden. Es zeigte sich, dass kein linearer Zusammenhang zwischen mRNA-Konzentration und Fluoreszenz-Intensität besteht. Die mit

dem Microarray gemessen Quotienten sind stets kleiner als die realen. Ursachen dafür könnten z.B. eine nicht-lineare Wandlung und Verstärkung der Fluoreszenz durch den Photomultiplier des Scanners oder mit der Konzentration zunehmende Fluoreszenz-Auslöschung der markierten Proben auf dem Microarray sein. Die Abweichungen der Messwerte sind jedoch im Vergleich zur Messunsicherheit nicht sehr groß. Daher wurde auf die Einführung eines Korrekturfaktors verzichtet.

Die Nachweisgrenze des Microarray-Systems liegt bei etwa 5 pg mRNA. Sie ist damit ausreichend niedrig, um außer RNAs mit mittlerer und großer Häufigkeit auch einen Großteil der RNAs mit geringer Häufigkeit zu detektieren. Dies ist entscheidend, da die RNAs mit geringer Häufigkeit etwa 96 % der RNA-Spezies einer Zelle ausmachen.

4.3.2 Reproduzierbarkeit der Microarray-Ergebnisse

Die wiederholte Messung der Genexpression mit mehreren Microarrays verringert den Standardfehler des Ergebnisses. Mit Hilfe von etwa zwölf Messungen kann der Standardfehler halbiert werden. Er verringert sich annähernd in dem Maße, wie in Kesel *et al.* (1999) beschrieben. Der Fehler, den man durch Messung unterschiedlicher Punkte derselben Sonde auf demselben Microarray erhält, ist dabei am geringsten. Gering ist auch der Fehler zwischen verschiedenen Microarrays, die mit identisch markierter cDNA (z.B. Experiment jeweils mit Cy3, Kontrolle jeweils mit Cy5) aus ein und demselben Experiment hybridisiert wurden (Abb. 23). Größer sind die Schwankungen, die durch *dye-swaps* verursacht werden. Die Verwendung von *dye-swaps* ist jedoch wichtig, da so artifizielle Faktoren der Induktion und Repression aufgehoben werden können. Wichtig für die Aussagekraft der Microarray-Experimente ist die wiederholte Messung mit unabhängigem biologischen Material, da hier die größten Schwankungen beobachtet wurden. Oft sind es die natürlichen Unterschiede zwischen Zellpopulationen, die am stärksten zur Messunsicherheit beitragen (Churchill, 2002).

Beim Infektionsexperiment mit *L. pneumophila* zum 24 h Zeitpunkt sind interessanterweise die Unterschiede zwischen den unabhängigen Infektionsansätzen eines Tages gering, während die Unterschiede zwischen Infektionen an verschiedenen Tagen die größte Messunsicherheit hervorrufen. Für die Infektionsansätze eines Tages wurde von *D. discoideum* stets eine gemeinsame Vorkultur inokuliert, mit der die einzelnen Kulturen für die Infektion beimpft wurden. Die Legionellen wurden aus Glycerolstöcken angezogen. Bei parallelem Wachstum der Dictyostelien und Legionellen, ausgehend von je einem Stock, sind die Änderungen der

Genexpression in *D. discoideum* am ähnlichsten. Beim Vergleich mit völlig unabhängigen Infektionen wurden die größten Unterschiede festgestellt (Abb. 23).

Durch die Verwendung vieler Microarrays, wie bei diesem Experiment, kann die biologische Variabilität bestimmt und die Messunsicherheit verringert werden. Durch die geringe Messunsicherheit werden jedoch auch sehr viele Gene von SAM als differentiell exprimiert identifiziert. Im Vergleich zu den Zeitpunkten 1, 3, 6 und 48 h nach Infektionsbeginn, die mit jeweils sechs Microarrays untersucht wurden, wies der 24 h Zeitpunkt viele Gene mit geringfügig geänderter Expression auf. Um die Experimente vergleichbar zu machen, mußte ein Schwellenwert für die Faktoren der Induktion und Repression gesetzt werden. Da die gefundenen Genexpressionsänderungen ohnehin erst oberhalb eines Faktors von ca. 1,5 reproduzierbar mit quantitativer PCR validiert werden können und auch die biologische Relevanz geringfügiger Expressionsänderungen unklar ist, ist solch ein zusätzlicher Schwellenwert für Microarray-Ergebnisse sinnvoll.

4.3.3 Validierung der Microarray-Ergebnisse mit quantitativer PCR und Northernblots

Die gemessenen Änderungen der Genexpression von *D. discoideum* in Folge der Infektion mit *L. pneumophila* konnten mittels quantitativer PCR und Northernblots für ausgewählte Sonden bestätigt werden. Nur in wenigen Fällen wichen die Messergebnisse voneinander ab. Die Unterschiede lagen meist im Rahmen der Messunsicherheit.

Mit den beiden alternativen Methoden zur Bestimmung der Genexpression wurden unter anderem die gemessenen Änderungen der Expression der Homologe zu Discoidin I (Sonde SSL850), ClpB (SLD769), einem der Calcium-bindenden Proteine (SSL818), *cotB* (SLE471), einer *short-chain* Dehydrogenase (SSJ163) und der nicht annotierten Sonde SSB153 bestätigt (siehe Abschnitt 4.4.1.1).

4.4 Infektion von *D. discoideum* mit *L. pneumophila* nach 24 Stunden

4.4.1 Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zu uninfizierten Zellen

Das Microarray-Experiment der Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zu uninfizierten Zellen hat 1218 differentiell exprimierte Gene identifiziert. Diese Zahl ist sehr groß, denn sie entspricht etwa 10 % aller Gene von *D. discoideum* und etwa 20 % der auf dem Microarray repräsentierten Gene. Viele dieser 1218 Gene haben sehr geringe Faktoren der Induktion und Repression. Der minimale Faktor beträgt in beiden Fällen

1,1. Setzt man einen Schwellenwert von 1,5, so bleiben nur 152 induzierte und 117 reprimierte Gene übrig. Bei einem Schwellenwert von 2,0 sind es sogar nur 29 bzw. 26. Ursache dafür, dass SAM so viele Gene mit geringfügig veränderter Genexpression als differentiell exprimiert identifiziert, liegt an der großen Zahl von 19 Microarrays. Sie verringern die Messunsicherheit und führen zu einer höheren Genanzahl bei gleicher Rate von falsch Positiven. Prinzipiell ist es denkbar, dass auch eine Induktion um den Faktor 1,1, also eine Erhöhung der mRNA-Konzentration um 10 %, einen Effekt in der Zelle hat. Da jedoch bei den Infektionsexperimenten auch wesentlich höhere Faktoren (bis zu 22fache Induktion) gemessen wurden, ist anzunehmen, dass in *D. discoideum* bei der Regulation der Genexpression die mRNA-Konzentrationen in größerem Umfang geändert werden. Iranfar *et al.* (2003) haben bei der Induktion des Entwicklungszyklus von *D. discoideum* Faktoren der Induktion im Bereich von 2 bis 10, in einigen Fällen auch größer, festgestellt. Auch der Vergleich mit Microarray-Experimenten zum diauxischen Wachstum von *S. cerevisiae* zeigt, dass sich die Expression regulierter Gene in stärkerem Maße ändert. So waren in diesem Experiment die Gene, die für Enzyme des Citratzyklus kodieren, alle zwischen 3,0 und 9,4fach induziert (DeRisi *et al.*, 1997).

Bei der Annotation einzelner Sonden, bei denen eine differentielle Expression festgestellt wurde, ist es daher sinnvoll sich auf die mit den höchsten Faktoren zu beschränken. Eine Validierung geringfügiger Induktion oder Repression mit alternativen Methoden ist ohnehin nicht möglich.

4.4.1.1 Annotation der differentiell exprimierten Gene

Da es sich bei 92 % der Sonden des *D. discoideum* Microarrays um cDNAs handelt, mußten die meisten der differentiell exprimierten Gene mittels Homologie-Suche annotiert werden. Einige Gene wie *rtoA*, dessen Produkt an der Endo- und Exozytose beteiligt ist, oder das der α -Mannosidase konnten dabei eindeutig identifiziert werden, da es sich um bekannte *D. discoideum*-Gene handelt. Bei vielen Sonden konnte der entsprechende kodierende Chromosomenabschnitt identifiziert werden. Das war sehr nützlich, da meist aufgrund des Genomprojekts ein von Genvorhersageprogrammen erzeugtes Genmodell für diesen Abschnitt existierte, dessen Sequenz viel länger war als die der Sonde. In diesem Fall wurde die vollständige vorhergesagte Proteinsequenz für die Homologie-Suche mittels BLASTP verwendet. Dadurch konnten häufig aussagekräftigere Treffer erzielt werden als das mit der Sondensequenz möglich gewesen wäre. Dennoch wurden für etwa die Hälfte aller Sonden keine Treffer mit ausreichend niedrigem p-Wert bei der BLAST-Suche erhalten.

Von den differentiell exprimierten Genen entsprechen einige den Erwartungen. Da *L. pneumophila* bei der Infektion die Phagosomenreifung blockiert, wurde angenommen, dass in *D. discoideum* bei der Infektion die Expression von Genen verändert wird, deren Produkte am Vesikeltransport beteiligt sind.

Das Gen *rtoA* kodiert für RtoA, ein Protein, das für die Fusion von Endo- und Exozytose-Vesikeln wichtig ist (Brazill *et al.*, 2000). *L. pneumophila* blockiert in *D. discoideum* die Fusion des Phagosoms, in dem es sich befindet, mit Endosomen und Lysosomen. Möglicherweise reagiert *D. discoideum* mit induzierter Transkription von *rtoA* auf die verminderte Vesikelfusion.

Die von *manA* kodierte lysosomale α -Mannosidase spaltet α -D-mannosidische Bindungen und kann damit zum Abbau von aufgenommenen Glykoproteinen beitragen. Hägele *et al.* haben gezeigt, dass in *D. discoideum* 15,8 % der Phagosomen mit *L. pneumophila*-Zellen mit Lysosomen fusionieren (Hägele *et al.*, 2000). Mannose ist neben anderen Zuckern Bestandteil der Lipopolysaccharide der äußeren Membran von *L. pneumophila* (Otten *et al.*, 1986). Die gemessene 2,1fache Induktion der *manA*-Expression könnte also auf den Verdau von *L. pneumophila* in einem Teil der Dictyostelien zurückzuführen sein.

Discoidine sind Proteine, die von einer Multigenfamilie kodiert werden (Wetterauer *et al.*, 1993). Discoidin I, zu dem die Sonde SSL850 große Ähnlichkeit hat, ist ein Lectin, das während des Entwicklungszyklus der interzellulären Adhäsion dient (Simpson *et al.*, 1974). Es ist außerdem bekannt, dass Discoidin I an die Lipopolysaccharide der Gram-negativen Bakterien *K. aerogenes* und *E. coli* bindet (Breuer und Siu, 1981). In vegetativen Zellen wird die Expression von Discoidin I durch cAMP, Folat, *prestarvation factor* (PSF) und *conditioned media factor* (CMF) reguliert. Die Induktion beginnt am Ende des logarithmischen Wachstums etwa drei Generationen vor dem Verbrauch der Futterbakterien (Clarke *et al.*, 1987; Blusch *et al.*, 1995). Ob Discoidin I auch an die Lipopolysaccharide von *L. pneumophila* bindet, ist nicht bekannt.

Bakterien haben im Allgemeinen in Relation zu ihrer Zellmasse einen höheren DNA-Gehalt als eukaryotische Zellen. Für die intrazelluläre Vermehrung benötigt *L. pneumophila* deshalb außer Kohlenhydraten, Aminosäuren und Lipiden auch einen großen Anteil Nukleotide. Es ist nicht bekannt wie *L. pneumophila* die Nährstoffe aus der Wirtszelle bezieht. Denkbar ist jedoch, dass die jeweiligen Monomere aufgenommen werden. Dadurch verringern sich möglicherweise die Nukleotid-Konzentrationen, weshalb in *D. discoideum* eine erhöhte Expression der Gene von Nukleotid-Biosynthesewegen erforderlich sein könnte. Die Sonde

SSK605 hat große Ähnlichkeit mit der Phosphoribosylaminoimidazol Carboxylase, einem Enzym der Purin-Biosynthese von *Bifidobacterium longum*.

Auch die Induktion des Gens, das von SLD769 repräsentiert wird, scheint plausibel. Es hat große Ähnlichkeit zur Endopeptidase ClpB von *Rhodopseudomonas palustris*. Dabei handelt es sich um ein molekulares Chaperon der Clp/Hsp100 Familie. Diese Chaperone sind in der Lage Proteinaggregate aufzulösen. Möglicherweise aggregieren in *D. discoideum* Proteine, die von *L. pneumophila* übertragen werden, oder die Infektion verändert den Stoffwechsel von *D. discoideum* so, dass es zur Aggregation eigener Proteine kommt.

Auffällig ist die Repression von drei Genen, die im Fall der Sonden SSL828 und SSL818 hohe Ähnlichkeit zu den Calcium-bindenden Proteinen (CBP) 4a und 4b von *D. discoideum*, und im Fall von SSL579 immerhin 36 % identische Aminosäuren mit CBP2 von *D. discoideum* haben. Calcium dient in vielen eukaryotischen Zellen als *second messenger* und leitet z.B. die Signale von Hormonen und Neurotransmittern weiter (Rizzuto und Pozzan, 2003). Müller-Taubenberger *et al.* (2001) fanden heraus, dass in *D. discoideum* die beiden CBPs Calreticulin und Calnexin für die Phagozytose erforderlich sind, da sie an der Umstrukturierung des Aktin-Cytoskeletts beteiligt sind. Die Funktionen der drei reprimierten putativen CBPs und ihre mögliche Rolle beim Infektionsprozess ist jedoch nicht bekannt.

Die Sonde VSE168 hat die größte Ähnlichkeit mit β' -COP von *Bos taurus*. COPs (*coat protein*) umhüllen Vesikel, die in den intra-Golgi Transport sowie den Transport zwischen Golgi und ER involviert sind (Lowe und Kreis, 1998). Mohrs *et al.* (2000) haben ein β' -COP Homolog in *D. discoideum* identifiziert, das im Golgi-Apparat lokalisiert ist. Sie fanden auch Hinweise auf eine endosomale und lysosomale Lokalisation (Mohrs, 2001). Bei dem von VSE168 repräsentierten Gen handelt es sich jedoch um ein neues β' -COP Homolog. Falls dieses Protein Endo- und Lysosomen umhüllt, wäre eine Beeinflussung durch *L. pneumophila* denkbar.

Unklar für den Infektionsprozess ist die Induktion der Gene, deren Genprodukte ähnlich zu *short-chain* Dehydrogenasen (VSA218 und SSJ163), einer Acetyl-CoA Hydrolase (SLD718) und einer Glutathion S-Transferase (SSI119) sind, sowie die Repression einer Stearoyl-CoA Desaturase (SSB431). *Short-chain* Dehydrogenasen gehören zu einer großen Familie von Enzymen, welche die Klassen der Oxidoreduktasen, Lyasen oder Isomerasen umfaßt und somit eine Vielzahl von Reaktionen katalysieren können. Die bekannten oder vermuteten Funktionen der übrigen drei Proteine konnten in keinen direkten Zusammenhang zur Infektion gebracht werden.

Widersprüchlich ist die beobachtete Induktion des Gens *dp87* sowie die Repression von *cotB*. Dp87 und CotB sind beide Proteine der Sporenhülle. Sie werden natürlicherweise während des Entwicklungszyklus von *D. discoideum* exprimiert und sezerniert (Fosnaugh und Loomis, 1993; Nakao *et al.*, 1994; West *et al.*, 1996).

Die Induktion der Expression eines Discoidin I Homologs und der Gene *dp87* und *cotB* könnte vermuten lassen, dass *D. discoideum* im Infektionsexperiment seinen Entwicklungszyklus initiiert. Dagegen spricht jedoch, dass die Zellen im Infektionsansatz langsam weiter wachsen und eine Aggregation der Zellen während der 48-stündigen Zeitreihe nicht beobachtet wurde. Außerdem wurden mit dem Microarray nur die Unterschiede der Genexpression im Vergleich mit Kontrollzellen gemessen, die abgesehen von den Legionellen im gleichen Medium kultiviert wurden.

4.4.1.2 Kategorisierung der differentiell exprimierten Gene

Basierend auf den funktionellen Kategorien für Hefegene (<http://mips.gsf.de/proj/yeast/CYGD/db/index.html>) wurden von H. Urushihara entsprechende Kategorien für *D. discoideum* entwickelt (Tabelle 7). Von den 1121 differentiell exprimierten Gene, die von cDNA-Sonden repräsentiert wurden, konnten 436 (39 %) eine der funktionellen Kategorie von 1 bis 12 zugewiesen werden. Die zellulären Prozesse in *D. discoideum*, die von der Infektion mit *L. pneumophila* beeinflusst werden, wurden auf der Ebene aller differentiell exprimierten und kategorisierten Gene untersucht. Dazu wurde ermittelt, in welchen Unterkategorien überproportional viele Gene, im Vergleich zum Anteil der Unterkategorie am cDNA-Satz, reguliert sind. Das Microarray-Experiment von DeRisi *et al.* (1997) zeigt, dass es im Fall des diauxischen Wachstums von *S. cerevisiae* zur Induktion aller Citratzyklus-Gene und Repression der meisten Glykolyse-Gene kommt. In unserem Fall konnten jedoch ein so stringentes Kriterium für die Überrepräsentation der funktionellen Kategorien nicht angewandt werden, da die Unterkategorien keine einzelnen Stoffwechsel- und Signalwege widerspiegeln. Außerdem ist die Kategorisierung zum Teil fehlerhaft, wie die Analyse einzelner Sonden zeigte. Deshalb wurde als Schwellenwert der Überrepräsentation willkürlich der Wert 2 gewählt.

Bei den induzierten Genen fällt vor allem die Unterkategorie 1.3 *Metabolism: nucleotide metabolism* auf. Von den 70 Genen des cDNA-Satzes sind 18 induziert, was einer 2,5fachen Überrepräsentation entspricht. Dies passt gut zu der Hypothese, die bezüglich der Phosphoribosylaminoimidazol Carboxylase (Sonde SSK605) aufgestellt wurde. Auch die Unterkategorien 6.5 *Cellular biogenesis and organization: nucleus* und 9.1 *Movement:*

cytoskeletal components mit sieben von 16 bzw. acht von 37 induzierten Genen weisen eine deutliche Überrepräsentation auf. Bei diesen Unterkategorien ist jedoch unklar in welchem Zusammenhang die beteiligten Proteine zum Infektionsvorgang stehen. Bei den übrigen überrepräsentierten Unterkategorien ist ein Interpretationsversuch nicht sinnvoll, da sich herausstellte, dass von den kategorisierten Genen nur jeweils ein bis drei Gene induziert sind. Dies lässt keine Aussagen über die Regulation eines zellulären Prozess zu. Eventuell erhielte man jedoch ein vollständigeres Bild, wenn alle oder ein Großteil der Gene kategorisiert wären.

Die Unterkategorien der reprimierten Gene, bei denen die Expression mehrerer Gene geändert ist, sind 1.6 *Metabolism: lipid, fatty-acid and isoprenoid metabolism*, 4.1 *Translation: ribosomal proteins* und 5.2 *Protein destination: protein targeting, sorting and translocation*. Die Repression von 38 Genen der Unterkategorie *ribosomal proteins* entspricht einer 4,8fachen Überrepräsentation dieser Unterkategorie und deutet darauf hin, dass infizierte *D. discoideum*-Zellen die Proteinsynthese reduzieren. Sortierung und Transport von Proteinen findet mit Hilfe von Vesikeln des Golgi-Apparats statt. Eine Repression von neun der 43 Gene der entsprechenden Unterkategorie könnte auf eine Beeinflussung des Vesikeltransports durch *L. pneumophila* hinweisen. Die Repression von Genen des Lipid-Metabolismus schließlich, scheint in keinem Zusammenhang zur Infektion zu stehen. Anes *et al.* (2003) haben jedoch für die Infektion von murinen Makrophagen mit Mycobakterien gezeigt, dass die Lipid-Zusammensetzung von Phagosomen über die Phagosomenreifung und damit die Abtötung des Pathogens entscheiden kann. *Mycobacterium tuberculosis* und *M. avium* vermehren sich wie *L. pneumophila* in Phagosomen von Makrophagen, die nicht mit Lysosomen fusionieren und nicht ansäuern.

4.4.2 Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zur Koinkubation mit *L. hackeliae*

Da die bei einem Microarray-Experiment als differentiell exprimiert identifizierten Gene entscheidend von der Wahl der geeigneten Kontrolle bestimmt werden, wurden außer uninfizierten Zellen auch die Koinkubation mit *L. hackeliae* und mit *L. pneumophila* $\Delta dotA$ als Kontrolle verwendet. Es wurden Übereinstimmungen und Unterschiede bei den differentiell exprimierten Genen beider Experimente erwartet. Bei beiden Experimenten sollten die Gene gefunden werden, die aufgrund der Etablierung eines replikativen Phagosoms durch *L. pneumophila* in *D. discoideum* reguliert sind. Unterschiede wurden zum einen bei den Genen erwartet, deren Expression durch den Kontakt mit und die Aufnahme

von *Legionella* beeinflusst wird. Im Experiment mit uninfizierten Zellen als Kontrolle kommt es in der Kontrolle nicht zum Kontakt mit Legionellen und es findet auch keine Aufnahme von Bakterien statt. Im Microarray-Experiment mit *L. hackeliae* als Kontrolle finden diese Vorgänge sowohl bei der Kontrolle als auch beim Experiment statt. Daher sollten diesbezüglich dieselben Gene reguliert sein und keine Unterschiede messbar sein. Anders verhält es sich mit dem Verdau von Legionellen. Während mit *L. hackeliae* als Kontrolle 82 % der Bakterien verdaut werden (Hägele *et al.*, 2000), findet bei uninfizierten Zellen als Kontrolle keine Digestion von Legionellen statt.

Vergleicht man die je zehn induzierten und reprimierten Gene mit den höchsten *score*-Werten dieses Experiments mit den Genlisten des Experiments mit uninfizierten Zellen als Kontrolle, so findet man zehn der 20 Gene in beiden Listen. Der Vergleich der je zehn Gene mit der gesamten Liste der differentiell exprimierten Gene des Experiments mit uninfizierten Zellen als Kontrolle zeigt, dass alle 20 Gene in beiden Listen enthalten sind. Dies spricht dafür, dass es sich bei den Genen mit hohen *scores* tatsächlich um Gene handelt, deren Expression durch den eigentlichen Infektionsvorgang beeinflusst wird. Die unterschiedlichen *score*-Werte eines Gens in den beiden Experimenten könnten durch die Schwankungen der Messergebnisse bedingt sein.

Insgesamt wurden 882 differentiell exprimierte Gene identifiziert. Um diese Gene mit denen des Experiments mit uninfizierten Zellen als Kontrolle zu vergleichen, wurde ein Schwellenwert von 1,5 gesetzt, da in beiden Experimenten eine unterschiedliche Anzahl von Microarrays verwendet wurde. Von den verbleibenden 269 regulierten Genen im Experiment mit uninfizierten Zellen und 588 regulierten Genen im Experiment mit *L. hackeliae* sind 197 in beiden Experimenten reguliert. Damit stimmen nur 30 % der insgesamt 660 Gene überein. Das könnte zum Teil daran liegen, dass mit 588 Genen im Experiment mit *L. hackeliae* als Kontrolle mehr als doppelt so viele Gene reguliert sind wie im Experiment mit uninfizierten Zellen als Kontrolle. Beschränkt man nämlich willkürlich die Zahl der Gene im Experiment mit *L. hackeliae* als Kontrolle auf die 269 mit den höchsten *score*-Werten, so erhält man 147 Gene (55 %) in der Schnittmenge. Es wurde erwartet, dass nicht alle differentiell exprimierten Gene beider Experimente gleich sein würden. Dass je nach Wahl des Schwellenwertes die Zahl der übereinstimmenden Gene jedoch nur 30 bis 55 % beträgt, überrascht jedoch. Ein Grund könnte sein, dass ein Teil der im Experiment mit *L. hackeliae* gefundenen Gene falsch positiv ist, da in diesem Fall nur RNAs aus einem Infektionsexperiment mit mehreren parallelen Ansätzen verwendet wurden. Im Gegensatz dazu wurden beim Vergleich von *L.*

pneumophila infizierten Zellen mit uninfizierten Zellen Infektionsexperimente an vier verschiedenen Zeitpunkten durchgeführt.

4.4.3 Infektion von *D. discoideum* mit *L. pneumophila* im Vergleich zur Koinkubation mit *L. pneumophila* $\Delta dotA$

Die Verwendung von *L. pneumophila* $\Delta dotA$ als Kontrolle bietet den Vorteil, dass bis auf die Deletion des *dotA*-Gens dieser Stamm identisch mit *L. pneumophila* ist. Dadurch werden Effekte anderer möglicher Virulenzfaktoren ausgeschlossen. Der Vergleich von Pathogenitätscharakteristika verschiedener *Legionella* Stämme (Alli *et al.*, 2003) zeigt, dass das Vorhandensein von Merkmalen wie Apoptoseinduktion und Zytopathogenität nicht immer mit der intrazellulären Vermehrungsrate korreliert.

Die *dotA*-Deletion hat jedoch gravierende Konsequenzen für *L. pneumophila*. Die Mutanten werden von Makrophagen und *D. discoideum* verdaut. Daher wurden mit *L. pneumophila* $\Delta dotA$ als Kontrolle ähnliche Ergebnisse erwartet wie mit *L. hackeliae* als Kontrolle. Bei beiden Stämmen unterschiedlich ist jedoch die Lipopolysaccharid-Hülle, da *L. pneumophila* und *L. hackeliae* verschiedenen Serogruppen angehören. *L. hackeliae* verhindert in *D. discoideum* darüber hinaus in 15 % der Phagozytosen die Fusion mit Lysosomen. Für *L. pneumophila* $\Delta dotA$ ist in *D. discoideum* nicht bekannt, ob alle aufgenommen Bakterien verdaut werden. Da Roy *et al.* (1998) jedoch in murinen Makrophagen beobachteten, dass sich nach 15 min fast alle $\Delta dotA$ -Legionellen in LAMP-1 positiven Vakuolen befanden, und *D. discoideum* für *Legionella* ein schwer zu infizierender Organismus ist, wie der Vergleich mit *A. castellanii* zeigt (Hägele *et al.*, 2000), kann man vermuten, dass *L. pneumophila* $\Delta dotA$ in *D. discoideum* vollständig verdaut wird.

Der Vergleich der je zehn induzierten und reprimierten Gene mit den höchsten *score*-Werten dieses Experiments mit denen des Experiments mit uninfizierten Zellen als Kontrolle zeigt, dass nur zwei der Gene übereinstimmen. Vergleicht man die je zehn Gene jedoch mit allen regulierten Genen des anderen Experiments, so findet man bis auf die Sonde SSK833 alle Gene.

Unter den induzierten Genen sind zwei bekannte *D. discoideum* Gene. Das Präprotein des Sorus-Matrix Proteins Dp87 und Countin. Dp87 und die zugehörige cDNA-Sonde SLG851 sind in den Experimenten mit uninfizierten Zellen und der $\Delta dotA$ -Mutante als Kontrolle induziert, finden sich jedoch nicht jeweils unter den zehn Genen mit den höchsten *scores*. Countin ist ein von Brock und Gomer (1999) entdecktes Protein, dass von *D. discoideum*

sekretiert wird und erforderlich ist, damit im Entwicklungszyklus keine übergroßen Zellaggregate gebildet werden. Der Zusammenhang zwischen der Infektion und dem Sorus Protein sowie Countin ist unbekannt.

Von den reprimierten Genen konnten ein ABC-Transporter, Flavohämoglobin und Untereinheiten der Cytochrom Oxidase identifiziert werden. Auch bei diesen Genen ergibt sich kein offensichtlicher funktioneller Zusammenhang.

Beim Vergleich aller differentiell exprimierten Gene dieses Experiments mit denen des Experiments, mit uninfizierten Zellen als Kontrolle, bei einem Schwellenwert von 1,5 (Abb. 26), ergibt sich ein ähnliches Bild wie in dem Vergleich mit *L. hackeliae* (Abb. 25). Die Anzahl der gemeinsamen Gene ist ähnlich groß. Ebenfalls etwa gleich groß ist die Zahl der Gene, die für das Experiment mit uninfizierten Zellen spezifisch sind. Unterschiedlich ist die Anzahl der Gene, die für das Experiment mit *L. pneumophila* $\Delta dotA$ spezifisch sind. Ursache dafür ist die hohe Anzahl von insgesamt 842 Genen, die den Schwellenwert von 1,5 passieren. Warum mit der $\Delta dotA$ -Mutante deutlich mehr regulierte Gene gefunden wurden als mit *L. hackeliae*, ist unklar. Technische Schwankungen der Microarray-Experimente scheinen nicht die Ursache zu sein, denn die mittlere Standardabweichung der Messergebnisse war in dem Experiment mit *L. hackeliae* niedriger als in dem Experiment mit *L. pneumophila* $\Delta dotA$. Insgesamt deuten die beobachteten Genzahlen darauf hin, dass sich *L. pneumophila* $\Delta dotA$ und *L. hackeliae* in *D. discoideum* ähnlich verhalten.

Da die Schnittmengen der beiden Vergleiche jeweils die Gene beinhalten sollten, deren Expression durch den eigentlichen Infektionsvorgang beeinflusst wird, wurden sie verglichen. Von insgesamt 250 Genen sind 140 (56 %) gleich. Dieser Anteil ist überraschend niedrig. Es war erwartet worden, dass die beiden Ausgangsvergleiche von je vier und 19 Microarrays bereits eine robuste Genliste erzeugt hätten. Die 140 differentiell exprimierten Gene umfassen jedoch fast alle Gene, die in einen funktionellen Zusammenhang zur Infektion gebracht werden konnten (siehe Abschnitt 4.4.1.1). Von den induzierten Genen sind dies RtoA, die lysosomale α -Mannosidase, Discoidin I und sein Homolog sowie das ClpB Homolog. Auch die Sonden SSB153, SSM409 und SSM847, die nicht annotiert werden konnten, sind darunter.

Bei den reprimierten Genen sind es das β' -COP Homolog, die drei Homologe kleiner Calcium-bindender Proteine und *cotB*. Die Gene der Sonden SSC374, SSC645, SSM741 und VSG118, zu denen keine Homologe gefunden wurden, sind ebenfalls in allen drei Experimenten reprimiert.

4.5 Zeitreihe der Infektion von *D. discoideum* mit *L. pneumophila*

Ein Ergebnis der Zeitreihe der Infektion ist, dass in *D. discoideum* die meisten Änderungen der Genexpression erst erfolgen, nachdem *Legionella* die Phagosomenreifung blockiert hat. Eine Stunde nach Infektionsbeginn sind bei einem Schwellenwert von 1,5 nur 19 Gene differentiell exprimiert (Tabelle 14) und die Faktoren der Induktion und Repression betragen im Mittel nur 1,8. Deutlich mehr Gene sind mit 129 und 189 zu den Zeitpunkten 3 h und 24 h nach Infektion reguliert. Interessant ist, dass die Anzahl der gefundenen Gene nach sechs und nach 48 Stunden auf 50 bzw. 77 reduziert ist. Was zu den Zeitpunkten drei und 24 Stunden nach Infektionsbeginn in *D. discoideum* passiert, ist leider nicht bekannt. Vergleiche mit Makrophagen sind nur eingeschränkt möglich, da die Infektion in Makrophagen wesentlich schneller abläuft als in *D. discoideum*. Die Lyse der Wirtszelle erfolgt in Makrophagen nach 24 bis 28 Stunden (Swanson und Hammer, 2000). In *D. discoideum* konnte erst nach 48 bis 72 Stunden eine Vermehrung von *L. pneumophila* festgestellt werden (Hägele *et al.*, 2000; Solomon *et al.*, 2000). In Makrophagen ist das Phagosom nach etwa einer Stunde mit Mitochondrien assoziiert, nach vier Stunden ist es mit Ribosomen besetzt und nach zwölf bis 24 Stunden kommt es zur Fusion mit Lysosomen. Möglicherweise spiegeln die vielen differentiell exprimierten Gene des 3 h und 24 h Zeitpunkts zwei dieser drei Prozesse wider.

Die Blockierung der Phagosomenreifung wird jedoch sicherlich nicht von Genexpressionsänderungen ausgelöst. Dafür läuft dieser Prozess zu schnell ab. Roy *et al.* (1998) haben gezeigt, dass Phagosomen mit *L. pneumophila* $\Delta dotA$ bereits nach 5 min das lysosomale Protein LAMP-1 tragen, während bei der Infektion mit *L. pneumophila* eine Kollokalisierung nicht eintritt. Da Transkription, Prozessierung und Translation einer durchschnittlichen eukaryotischen mRNA länger dauert, spielen hier vermutlich Proteine oder Lipide eine Rolle.

Die Clusteranalyse der differentiell exprimierten Gene der Zeitreihe führte nicht zur Identifikation von zellulären Prozessen, die im Verlauf der Infektion transkriptionell reguliert sind. Die Grundannahme der Clusteranalyse ist, dass Gene, die in gleicher Weise über die Zeit reguliert sind, in einem funktionellen Zusammenhang stehen könnten. Dies wurde z.B. in Experimenten zur Stimulation von Fibroblasten und zum Zellzyklus von *S. cerevisiae* gezeigt (Eisen *et al.*, 1998; Spellman *et al.*, 1998). Eine Erklärung dafür könnte sein, dass bei *D. discoideum* für zu wenige Sonden des Microarrays eine Kategorisierung und bekannte Homologe vorliegen. Von den 1133 differentiell exprimierten Genen der Zeitreihe konnten nur 317 (28 %) einer funktionellen Kategorie zugeordnet werden. Nur 179 (16 %) haben

einen Treffer in der Datenbank TrEMBL. Daher wiesen viele Cluster nur einzelne annotierte Sonden auf und ein Muster konnte nicht erkannt werden.

Ersatzweise wurden die zahlenmäßig insgesamt überrepräsentierten Kategorien der Zeitreihe ermittelt. Dabei fielen vor allem die Kategorien 6 *Cellular biogenesis and organization* und 10 *Stress response and cell rescue* auf, da sie häufig vertreten sind. Da die Infektion zu einer erheblichen Umorganisation der Wirtszelle führt und diese wahrscheinlich auch eine Stress-Antwort von *D. discoideum* auslöst, erscheint die Überrepräsentation dieser Kategorien plausibel. Ein genaues Verständnis der Genexpressionsänderungen ermöglicht dies jedoch nicht.

4.6 Ausblick

Die Verwendung von DNA-Microarrays ist ein neuer Ansatz zur Untersuchung der Interaktion von Wirt und Pathogen. Auf Wirtsseite können damit Gene identifiziert werden, die für den Infektionsprozess bedeutsam sein können. Die Untersuchung der Genexpression des Pathogens wäre ebenfalls sehr interessant.

In *D. discoideum* wurden zahlreiche Gene identifiziert, die nach Infektion mit *L. pneumophila* differentiell exprimiert werden. Die veränderte Expression einige dieser Gene wurde mit quantitativer PCR bzw. Northernblots verifiziert. Nun könnten diese Gene und ihre Proteine näher untersucht werden. Zunächst sollte bestimmt werden, ob die Proteinkonzentrationen sich in gleichem Maße verändern, wie die mRNA-Konzentrationen. Dann könnten diese Gene in *D. discoideum* deletiert sowie überexprimiert werden und die Auswirkungen auf die Infektion untersucht werden.

Das *D. discoideum* DNA-Microarray repräsentiert momentan mit 5906 Sonden etwa die Hälfte aller Gene des Genoms. Die inzwischen abgeschlossene Sequenzierung des *D. discoideum* Genoms ermöglicht die *in silico*-Vorhersage nahezu aller Gene dieses Organismus. Zur Zeit werden für die vorhergesagten Gene spezifische Oligonukleotide synthetisiert und mit diesen dann Sonden für die nächste Generation des *D. discoideum* DNA-Microarrays amplifiziert. Dieses Microarray wird dann zusätzlich zu den bisher verwendeten Sonden, Sonden für annähernd alle Gene des Genoms tragen. Damit wären Untersuchungen der Genexpression möglich, bei denen nahezu alle Gene erfasst werden könnten. Dabei würden etwa die Hälfte aller Gene von zwei Sonden repräsentiert. Einer Sonde des cDNA- und einer des Genom-Projekts. Diese könnten zur Verringerung der Messunsicherheit verwendet werden. Dafür müssen die cDNA-Sequenzen den Genmodellen zugeordnet

werden. Wichtig für eine effiziente Nutzung des *D. discoideum*-Microarrays ist auch eine verbesserte Annotation der Genmodelle. Dazu sind Homologie-Suchen und Suchen nach Domänen geplant. Diese Suchen versprechen mit der Zeit immer bessere Ergebnisse zu liefern, da mehr und mehr Genome sequenziert werden.

5 Zusammenfassung

Frischwasser-Amöben stellen den natürlichen Wirt von *L. pneumophila*, den Erreger der Legionärskrankheit dar. Die soziale Amöbe *D. discoideum*, ein Modellorganismus, wurde verwendet, um die differentielle Genexpression des Wirts nach Infektion mit *L. pneumophila* zu untersuchen. Dazu wurde ein DNA-Microarray hergestellt, das mit 5906 genspezifischen Sonden etwa die Hälfte des *D. discoideum*-Genoms repräsentiert. Mit Hilfe interner Kontrollen wurden die gemessenen Faktoren der Induktion und Repression der Expression von Genen überprüft und die Nachweisgrenze des Systems bestimmt. Es wurde untersucht, wie sich die Messunsicherheit durch wiederholte Messungen verringern lässt, und welche Faktoren zur Messunsicherheit beitragen. Ein automatisiertes Auswerteverfahren wurde entwickelt, mit dem die Messspanne des Systems durch die Verwendung von zwei Messbereichen vergrößert wurde. Das Verfahren nutzt freie Software, um alle Messwerte eines Microarray-Experiments gemeinsam zu normalisieren und die differentiell exprimierten Gene zu identifizieren.

Mit diesem Microarray-System wurde die Genexpression von *D. discoideum* nach Infektion mit *L. pneumophila* untersucht. Die Experimente wurden im Vergleich zu drei Kontrollen, uninfizierten Zellen und Koinkubation mit *L. hackeliae* sowie mit *L. pneumophila* $\Delta dotA$ durchgeführt. *L. hackeliae* und *L. pneumophila* $\Delta dotA$ sind *Legionella* Stämme, die eine verminderte Pathogenität bzw. ein deletiertes Pathogenitätsgen aufweisen. Für den Zeitpunkt 24 Stunden nach Infektionsbeginn wurden 140 in *D. discoideum* differentiell exprimierte Gene identifiziert. Bei einigen handelt es sich um Gene, die für bekannte Proteine von *D. discoideum* kodieren. Dazu gehören RtoA, Discoidin I, CotB und die lysosomale α -Mannosidase. Anderen konnte mittels Homologie-Suche eine Funktion zugeordnet werden. Hierzu zählen das molekulare Chaperon ClpB, β' -COP und drei kleine Calcium-bindende Proteine. Mit Hilfe einer Kategorisierung der Sonden nach zellulären Prozessen wurden die Genexpressionsänderungen auch auf einer funktionellen Ebene analysiert. Es konnte gezeigt werden, dass besonders viele Gene reguliert sind, deren Produkte am Nukleotid-Metabolismus beteiligt sind, oder bei denen es sich um ribosomale Proteine handelt.

Die Genexpression wurde auch in einer Zeitreihe der Infektion über 48 Stunden untersucht. Dabei zeigte sich, dass drei Stunden und 24 Stunden nach Infektionsbeginn die meisten Änderungen auftreten. Die mit dem Microarray identifizierten Änderungen der Genexpression wurden für zehn Gene mittels quantitativer PCR und Northernblots bestätigt.

6 Abstract

Freshwater amoebae are the natural host of *L. pneumophila*, the causative agent of Legionnaires' disease. The social amoeba *D. discoideum*, a model organism, was used to investigate differential gene expression of the host after infection with *L. pneumophila*. For that purpose a DNA microarray was produced, that consists of 5906 gene-specific probes representing about half the genome of *D. discoideum*. Using internal controls, the measured folds of induction and repression of gene expression were checked and the detection limit of the system determined. It was examined, how the measuring inaccuracy can be reduced by repeated measurements and which effects contribute to the inaccuracy. An automated analysis procedure was developed, that increases the measuring range of the system, by utilizing two measuring ranges. The procedure uses free software, to normalize all measured values of a microarray experiment together and identify the differentially expressed genes.

This microarray system was used to analyze the gene expression of *D. discoideum* after infection with *L. pneumophila*. The experiments were performed in comparison to three controls, uninfected cells and co-incubation with *L. hackeliae* as well as *L. pneumophila* $\Delta dotA$. *L. hackeliae* and *L. pneumophila* $\Delta dotA$ are *Legionella* strains with reduced pathogenicity and a deleted pathogenicity gene respectively. For the 24-hour time point post infection, 140 differentially expressed *D. discoideum* genes were identified. A few of these genes code for well characterized *D. discoideum* proteins. Among these are RtoA, Discoidin I, CotB and the lysosomal α -mannosidase. Functions could be assigned to others by homology searches. The molecular chaperone ClpB, β' -COP and three small calcium binding proteins are some of these. With the aid of a categorization of probes by cellular processes, the alterations in gene expression were analyzed on a functional level. It could be shown, that especially many genes are regulated, whose products are involved in nucleotide metabolism or that are ribosomal proteins.

The gene expression was also investigated during a 48-hour time course of infection. This revealed, that most of the changes occur three and 24 hours after infection. The differential expression of ten genes identified with DNA microarrays was confirmed by quantitative PCR and northern blots.

7 Literaturverzeichnis

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., und Watson, J. D. (1994). *Molecular Biology Of The Cell*, 3 rd. Edition (New York: Garland Publishing).

Alli, O. A., Zink, S., Von Lackum, N. K., und Abu-Kwaik, Y. (2003). Comparative assessment of virulence traits in *Legionella* spp. *Microbiology* 149, 631-41.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., und Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-10.

Andre, B., Noegel, A. A., und Schleicher, M. (1996). *Dictyostelium discoideum* contains a family of calmodulin-related EF-hand proteins that are developmentally regulated. *FEBS Lett* 382, 198-202.

Anes, E., Kuhnel, M. P., Bos, E., Moniz-Pereira, J., Habermann, A., und Griffiths, G. (2003). Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria. *Nat Cell Biol* 5, 793-802. Epub 2003 Aug 24.

Baldauf, S. L., und Doolittle, W. F. (1997). Origin and evolution of the slime molds (Mycetozoa). *Proc Natl Acad Sci U S A* 94, 12007-12.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., und Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972-7.

Beccari, T., Stinchi, S., und Orlacchio, A. (1999). Lysosomal alpha-D-mannosidase. *Biosci Rep* 19, 157-62.

Belyaeva, O. V., Stetsenko, A. V., Nelson, P., und Kedishvili, N. Y. (2003). Properties of short-chain dehydrogenase/reductase RalR1: characterization of purified enzyme, its orientation in the microsomal membrane, and distribution in human tissues and cell lines. *Biochemistry* 42, 14838-45.

Bland, J. M., und Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307-10.

Blusch, J., Alexander, S., und Nellen, W. (1995). Multiple signal transduction pathways regulate discoidin I gene expression in *Dictyostelium discoideum*. *Differentiation* 58, 253-60.

Brazill, D. T., Caprette, D. R., Myler, H. A., Hatton, R. D., Ammann, R. R., Lindsey, D. F., Brock, D. A., und Gomer, R. H. (2000). A protein containing a serine-rich domain with vesicle fusing properties mediates cell cycle-dependent cytosolic pH regulation. *J Biol Chem* 275, 19231-40.

Brenner, D. J., Steigerwalt, A. G., Gorman, G. W., Wilkinson, H. W., Bibb, W. F., Hackel, M., Tyndall, R. L., Campbell, J., Feeley, J. C., Thacker, W. L., Skaliy, P., Martin, W. T., Brake, B. J., Fields, B. S., McEachern, H. V., und Corcoran, L. K. (1985). Ten new species of *Legionella*. *Int. J. Syst. Bacteriol.* 35, 50-9.

Breuer, W., und Siu, C. H. (1981). Identification of endogenous binding proteins for the lectin discoidin-I in *Dictyostelium discoideum*. *Proc Natl Acad Sci U S A* 78, 2115-9.

- Brock, D. A., und Gomer, R. H. (1999). A cell-counting factor regulating structure size in *Dictyostelium*. *Genes Dev* *13*, 1960-9.
- Buu, L. M., Chen, Y. C., und Lee, F. J. (2003). Functional characterization and localization of acetyl-CoA hydrolase, Ach1p, in *Saccharomyces cerevisiae*. *J Biol Chem* *278*, 17203-9. Epub 2003 Feb 26.
- Cheung, V. G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., und Childs, G. (1999). Making and reading microarrays. *Nat Genet* *21*, 15-9.
- Churchill, G. A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet* *32 Suppl*, 490-5.
- Clarke, M., Kayman, S. C., und Riley, K. (1987). Density-dependent induction of discoidin-I synthesis in exponentially growing cells of *Dictyostelium discoideum*. *Differentiation* *34*, 79-87.
- Claviez, M., Pagh, K., Maruta, H., Baltes, W., Fisher, P., und Gerisch, G. (1982). Electron microscopic mapping of monoclonal antibodies on the tail region of *Dictyostelium* myosin. *EMBO J.* *1*, 1017-22.
- Coles, B. F., und Kadlubar, F. F. (2003). Detoxification of electrophilic compounds by glutathione S-transferase catalysis: determinants of individual response to chemical carcinogens and chemotherapeutic drugs? *Biofactors* *17*, 115-30.
- Conlan, J. W., Baskerville, A., und Ashworth, L. A. (1986). Separation of *Legionella pneumophila* proteases and purification of a protease which produces lesions like those of Legionnaires' disease in guinea pig lung. *J Gen Microbiol* *132*, 1565-74.
- Cooper, D. N., Lee, S. C., und Barondes, S. H. (1983). Discoidin-binding polysaccharide from *Dictyostelium discoideum*. *J Biol Chem* *258*, 8745-50.
- Cordevant, C., Tang, J. S., Cleland, D., und Lange, M. (2003). Characterization of members of the *Legionellaceae* family by automated ribotyping. *J Clin Microbiol* *41*, 34-43.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., und Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* *14*, 457-60.
- DeRisi, J. L., Iyer, V. R., und Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* *278*, 680-686.
- Dorywalska, M., Coukell, B., und Dharamsi, A. (2000). Characterization and heterologous expression of cDNAs encoding two novel closely related Ca(2+)-binding proteins in *Dictyostelium discoideum*. *Biochim Biophys Acta* *1496*, 356-61.
- Dudoit, S., und Yang, Y. H. (2003). Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data. In *The Analysis of Gene Expression Data: Methods and Software*, G. Parmigiani, E. S. Garrett, R. A. Irizarry und S. L. Zeger, eds. (New York: Springer).

- Dudoit, S., Yang, Y. H., und Bolstad, B. (2002). Using R for the analysis of DNA microarray data. *R News* 2, 24-32.
- Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., und Trent, J. M. (1999). Expression profiling using cDNA microarrays. *Nat Genet* 21, 10-4.
- Dyrskjot, L. (2003). Classification of bladder cancer by microarray expression profiling: towards a general clinical use of microarrays in cancer diagnostics. *Expert Rev Mol Diagn* 3, 635-47.
- Efron, B., Tibshirani, R., Storey, J. D., und Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* 96, 1151-60.
- Eisen, M. B., Spellman, P. T., Brown, P. O., und Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95, 14863-8.
- Farbrother, P., Müller, S., Noegel, A. A., und Eichinger, L. (2002). Comparison of Probe Preparation Methods for DNA Microarrays. *BioTechniques* 33, 884-888.
- Feng, H. P., und Gierasch, L. M. (1998). Molecular chaperones: clamps for the Clps? *Curr Biol* 8, R464-7.
- Fosnaugh, K. L., und Loomis, W. F. (1993). Enhancer regions responsible for temporal and cell-type-specific expression of a spore coat gene in *Dictyostelium*. *Dev Biol* 157, 38-48.
- Fraser, D. W., Tsai, T. R., Orenstein, W., Parkin, W. E., Beecham, H. J., Sharrar, R. G., Harris, J., Mallison, G. F., Martin, S. M., McDade, J. E., Shepard, C. C., and Brachman, P. S. (1977). Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med* 297, 1189-97.
- Gao, L. Y., und Abu Kwaik, Y. (1999). Apoptosis in macrophages and alveolar epithelial cells during early stages of infection by *Legionella pneumophila* and its role in cytopathogenicity. *Infect Immun* 67, 862-70.
- Gao, L. Y., Harb, O. S., und Abu Kwaik, Y. (1997). Utilization of similar mechanisms by *Legionella pneumophila* to parasitize two evolutionarily distant host cells, mammalian macrophages and protozoa. *Infect Immun* 65, 4738-46.
- Gerisch, G. (1987). Cyclic AMP and other signals controlling cell development and differentiation in *Dictyostelium*. *Annu. Rev. Biochem.* 56, 853-879.
- Gerisch, G. (1965). Spezifische Zellkontakte als Mechanismen der tierischen Entwicklung. *Umschau* 13, 392-395.
- Gerisch, G., Malchow, D., Huesgen, A., Nanjundiah, V., Roos, W., Wick, U., und Hulser, D. (1975). Cyclic-AMP reception and cell recognition in *Dictyostelium discoideum*. In *Developmental biology: Pattern formation, gene regulation*. (ICN-UCLA Symp.), D. McMahon and C. F. Fox, eds. (Menlo Park, CA: W.A. Benjamin), pp. 76-88.

- Giegerich, R., Meyer, F., und Schleiermacher, C. (1996). GeneFisher - Software Support for the Detection of Postulated Genes. In Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology: AAAI Press).
- Hägele, S., Köhler, R., Merkert, H., Schleicher, M., Hacker, J., und Steinert, M. (2000). *Dictyostelium discoideum*: a new host model system for intracellular pathogens of the genus *Legionella*. Cell Microbiol 2, 165-71.
- Hayes, J. D., und Strange, R. C. (1995). Potential contribution of the glutathione S-transferase supergene family to resistance to oxidative stress. Free Radic Res 22, 193-207.
- Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N., und Quackenbush, J. (2000). A Concise Guide to cDNA Microarray Analysis. BioTechniques 29, 548-562.
- Heinemann, F. S., und Ozols, J. (2003). Stearoyl-CoA desaturase, a short-lived protein of endoplasmic reticulum with multiple control mechanisms. Prostaglandins Leukot Essent Fatty Acids 68, 123-33.
- Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D. E., und Davis, R. W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci U S A 94, 2150-5.
- Hickey, E. K., und Cianciotto, N. P. (1997). An iron- and fur-repressed *Legionella pneumophila* gene that promotes intracellular infection and encodes a protein with similarity to the *Escherichia coli* aerobactin synthetases. Infect Immun 65, 133-43.
- Horwitz, M. A. (1983). Formation of a novel phagosome by the Legionnaires' disease bacterium (*Legionella pneumophila*) in human monocytes. J Exp Med 158, 1319-31.
- Horwitz, M. A. (1983). The Legionnaires' disease bacterium (*Legionella pneumophila*) inhibits phagosome-lysosome fusion in human monocytes. J Exp Med 158, 2108-26.
- Horwitz, M. A., und Maxfield, F. R. (1984). *Legionella pneumophila* inhibits acidification of its phagosome in human monocytes. J Cell Biol 99, 1936-43.
- Ihaka, R., und Gentleman, R. (1996). R: a language for data analysis and graphics. J. Comput. Graph. Statist. 5, 299-314.
- Iranfar, N., Fuller, D., und Loomis, W. F. (2003). Genome-wide expression analyses of gene regulation during early development of *Dictyostelium discoideum*. Eukaryot Cell 2, 664-70.
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. (1999). The transcriptional program in the response of human fibroblasts to serum. Science 283, 83-7.
- Kesel, A. B., Junge, M. M., und Nachtigall, W. (1999). Einführung in die angewandte Statistik für Biowissenschaftler (Basel: Birkhäuser Verlag).

- Kessin, R. H. (2001). *Dictyostelium* - Evolution, cell biology, and the development of multicellularity. (Cambridge, UK: Cambridge Univ. Press).
- Kibler, K., Nguyen, T. L., Svetz, J., Van Driessche, N., Ibarra, M., Thompson, C., Shaw, C., und Shaulsky, G. (2003). A novel developmental mechanism in *Dictyostelium* revealed in a screen for communication mutants. *Dev Biol* 259, 193-208.
- King, C. H., Fields, B. S., Shotts, E. B., Jr., und White, E. H. (1991). Effects of cytochalasin D and methylamine on intracellular growth of *Legionella pneumophila* in amoebae and human monocyte-like cells. *Infect Immun* 59, 758-63.
- Kirby, J. E., Vogel, J. P., Andrews, H. L., und Isberg, R. R. (1998). Evidence for pore-forming ability by *Legionella pneumophila*. *Mol Microbiol* 27, 323-36.
- Ko, K. S., Hong, S. K., Lee, H. K., Park, M. Y., und Kook, Y. H. (2003). Molecular evolution of the *dotA* gene in *Legionella pneumophila*. *J Bacteriol* 185, 6269-77.
- Kuspa, A., und Loomis, W. F. (1992). Tagging developmental genes in *Dictyostelium* by restriction enzyme-mediated integration of plasmid DNA. *Proc Natl Acad Sci U S A* 89, 8803-7.
- Kwaik, Y. A. (1998). Fatal attraction of mammalian cells to *Legionella pneumophila*. *Mol Microbiol* 30, 689-95.
- Lee, M. L., Kuo, F. C., Whitmore, G. A., und Sklar, J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A* 97, 9834-9.
- Lehninger, A. L. (2001). *Lehninger Biochemie*, 3. Auflage Edition (Berlin: Springer).
- Lehrach, H., Diamond, D., Wozney, J. M., Boedtke, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16, 4743-51.
- Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., und Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat Genet* 21, 20-4.
- Lipshutz, R. J., Morris, D., Chee, M., Hubbell, E., Kozal, M. J., Shah, N., Shen, N., Yang, R., und Fodor, S. P. (1995). Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* 19, 442-7.
- Lowe, M., und Kreis, T. E. (1998). Regulation of membrane traffic in animal cells by COPI. *Biochim Biophys Acta* 1404, 53-66.
- Madley, I. C., und Hames, B. D. (1981). An analysis of discoidin I binding sites in *Dictyostelium discoideum* (NC4). *Biochem J* 200, 83-91.
- Maniatis, T., Fritsch, E. F., und Sambrook, J. (1982). *Molecular cloning: A laboratory manual* (Cold Spring Harbor: Cold Spring Harbor Laboratory).

- Marra, A., und Shuman, H. A. (1989). Isolation of a *Legionella pneumophila* restriction mutant with increased ability to act as a recipient in heterospecific matings. *J Bacteriol* 171, 2238-40.
- Martinez, M. J., Aragon, A. D., Rodriguez, A. L., Weber, J. M., Timlin, J. A., Sinclair, M. B., Haaland, D. M., und Werner-Washburne, M. (2003). Identification and removal of contaminating fluorescence from commercial and in-house printed DNA microarrays. *Nucleic Acids Res* 31, E18-E18.
- Mehdy, M. C., Ratner, D., und Firtel, R. A. (1983). Induction and modulation of cell-type-specific gene expression in *Dictyostelium*. *Cell* 32, 763-71.
- Miyazaki, M., und Ntambi, J. M. (2003). Role of stearoyl-coenzyme A desaturase in lipid metabolism. *Prostaglandins Leukot Essent Fatty Acids* 68, 113-21.
- Mohrs, M. R. (2001). Identifizierung und Charakterisierung von β -COP in *Dictyostelium discoideum*.
- Mohrs, M. R., Janssen, K. P., Kreis, T., Noegel, A. A., und Schleicher, M. (2000). Cloning and characterization of beta-COP from *Dictyostelium discoideum*. *Eur J Cell Biol* 79, 350-7.
- Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mitra, B. N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H., und Tanaka, Y. (1998). The *Dictyostelium* developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development. *DNA Res* 5, 335-40.
- Morris, S. A., Asnake, B., und Yen, G. (2003). Dendrogram seriation using simulated annealing. *Information Visualization* 2, 95-104.
- Müller-Taubenberger, A., Lupas, A. N., Li, H., Ecke, M., Simmeth, E., und Gerisch, G. (2001). Calreticulin and calnexin in the endoplasmic reticulum are important for phagocytosis. *Embo J* 20, 6772-82.
- Nakao, H., Yamamoto, A., Takeuchi, I., und Tasaka, M. (1994). *Dictyostelium* prespore-specific gene Dp87 encodes a sorus matrix protein. *J Cell Sci* 107, 397-403.
- Nellen, W., Silan, C., und Firtel, R. A. (1984). DNA-mediated transformation in *Dictyostelium discoideum*: regulated expression of an actin gene fusion. *Mol. Cell. Biol.* 4, 2890-2898.
- Otten, S., Iyer, S., Johnson, W., und Montgomery, R. (1986). Serospecific antigens of *Legionella pneumophila*. *J Bacteriol* 167, 893-904.
- Ozaki, T., Nakao, H., Orii, H., Morio, T., Takeuchi, I., und Tasaka, M. (1993). Developmental regulation of transcription of a novel prespore-specific gene (Dp87) in *Dictyostelium discoideum*. *Development* 117, 1299-308.
- Parent, C. A., und Devreotes, P. N. (1999). A cell's sense of direction. *Science* 284, 765-70.

- Persson, B., Kallberg, Y., Oppermann, U., und Jornvall, H. (2003). Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chem Biol Interact* 143-144, 271-8.
- Raper, K. B. (1935). *Dictyostelium discoideum*, a new species of slime mold from decaying forest leaves. *J. Agr. Res.* 50, 135-147.
- Rizzuto, R., und Pozzan, T. (2003). When calcium goes wrong: genetic alterations of a ubiquitous signaling route. *Nat Genet* 34, 135-41.
- Rowbotham, T. J. (1980). Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J Clin Pathol* 33, 1179-83.
- Roy, C. R., Berger, K. H., und Isberg, R. R. (1998). *Legionella pneumophila* DotA protein is required for early phagosome trafficking decisions that occur within minutes of bacterial uptake. *Mol Microbiol* 28, 663-74.
- Roy, C. R., und Isberg, R. R. (1997). Topology of *Legionella pneumophila* DotA: an inner membrane protein required for replication in macrophages. *Infect Immun* 65, 571-8.
- Sakamoto, H., Nishio, K., Tomisako, M., Kuwayama, H., Tanaka, Y., Suetake, I., Tajima, S., Ogihara, S., Coukell, B., und Maeda, M. (2003). Identification and characterization of novel calcium-binding proteins of *Dictyostelium* and their spatial expression patterns during development. *Dev Growth Differ* 45, 507-14.
- Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., 3rd, Slocombe, P. M., und Smith, M. (1978). The nucleotide sequence of bacteriophage phiX174. *J Mol Biol* 125, 225-46.
- Schatzle, J., Bush, J., und Cardelli, J. (1992). Molecular cloning and characterization of the structural gene coding for the developmentally regulated lysosomal enzyme, alpha-mannosidase, in *Dictyostelium discoideum*. *J Biol Chem* 267, 4000-7.
- Schena, M., Shalon, D., Davis, R. W., und Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-70.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., und Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* 93, 10614-9.
- Schleicher, M., und Noegel, A. A. (1992). Dynamics of the *Dictyostelium* cytoskeleton during chemotaxis. *New Biol* 4, 461-72.
- Segal, G., Purcell, M., und Shuman, H. A. (1998). Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc Natl Acad Sci U S A* 95, 1669-74.
- Segal, G., und Shuman, H. A. (1999). *Legionella pneumophila* utilizes the same genes to multiply within *Acanthamoeba castellanii* and human macrophages. *Infect Immun* 67, 2117-24.

- Shalon, D., Smith, S. J., und Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* 6, 639-45.
- Shi, R., und Lin, S. X. (2004). Cofactor hydrogen bonding onto the protein main chain is conserved in the short-chain dehydrogenase/reductase family and contributes to nicotinamide orientation. *J Biol Chem* 13, 13.
- Simpson, D. L., Rosen, S. D., und Barondes, S. H. (1974). Discoidin, a developmentally regulated carbohydrate-binding protein from *Dictyostelium discoideum*. Purification and characterization. *Biochemistry* 13, 3487-93.
- Skriwan, C., Fajardo, M., Hägele, S., Horn, M., Wagner, M., Michel, R., Krohne, G., Schleicher, M., Hacker, J., und Steinert, M. (2002). Various bacterial pathogens and symbionts infect the amoeba *Dictyostelium discoideum*. *Int J Med Microbiol* 291, 615-24.
- Smyth, G. K., Yang, Y. H., und Speed, T. (2002). Statistical issues in cDNA Microarray Data Analysis.
- Solomon, J. M., und Isberg, R. R. (2000). Growth of *Legionella pneumophila* in *Dictyostelium discoideum*: a novel system for genetic analysis of host-pathogen interactions. *Trends Microbiol* 8, 478-80.
- Solomon, J. M., Rupper, A., Cardelli, J. A., und Isberg, R. R. (2000). Intracellular growth of *Legionella pneumophila* in *Dictyostelium discoideum*, a system for genetic analysis of host-pathogen interactions. *Infect Immun* 68, 2939-47.
- Spellman, P. T., und Rubin, G. M. (2002). Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol* 1, 5.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., und Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9, 3273-97.
- Stenbeck, G., Harter, C., Brecht, A., Herrmann, D., Lottspeich, F., Orci, L., und Wieland, F. T. (1993). beta'-COP, a novel subunit of coatamer. *Embo J* 12, 2841-5.
- Strange, R. C., Jones, P. W., und Fryer, A. A. (2000). Glutathione S-transferase: genetics and role in toxicology. *Toxicol Lett* 112-113, 357-63.
- Sun, H., und Wolfe, J. H. (2001). Recent progress in lysosomal alpha-mannosidase and its deficiency. *Exp Mol Med* 33, 1-7.
- Sussman, R., und Sussman, M. (1967). Cultivation of *Dictyostelium discoideum* in axenic culture. *Biochem. Biophys. Res. Commun.* 29, 53-55.
- Swanson, M. S., und Hammer, B. K. (2000). *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. *Annu Rev Microbiol* 54, 567-613.

- Tukey, J. W. (1977). Box-and-Whisker Plots. In Explanatory Data Analysis (Reading: Addison-Wesley).
- Tusher, V. G., Tibshirani, R., und Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98, 5116-21.
- Urushihara, H., Morio, T., Saito, T., Kohara, Y., Koriki, E., Ochiai, H., Maeda, M., Williams, J. G., Takeuchi, I., und Tanaka, Y. (2004). Analyses of cDNAs from growth and slug stages of *Dictyostelium discoideum*. *Nucleic Acids Res* 32, 1647-1653. Print 2004.
- Van Driessche, N., Shaw, C., Katoh, M., Morio, T., Sucgang, R., Ibarra, M., Kuwayama, H., Saito, T., Urushihara, H., Maeda, M., Takeuchi, I., Ochiai, H., Eaton, W., Tollett, J., Halter, J., Kuspa, A., Tanaka, Y., und Shaulsky, G. (2002). A transcriptional profile of multicellular development in *Dictyostelium discoideum*. *Development* 129, 1543-52.
- Vogel, J. P., Andrews, H. L., Wong, S. K., und Isberg, R. R. (1998). Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279, 873-6.
- West, C. M., Mao, J., van der Wel, H., Erdos, G. W., und Zhang, Y. (1996). SP75 is encoded by the DP87 gene and belongs to a family of modular *Dictyostelium discoideum* outer layer spore coat proteins. *Microbiology* 142, 2227-43.
- Wetterauer, B., Jacobsen, G., Morandini, P., und MacWilliams, H. (1993). Mutants of *Dictyostelium discoideum* with defects in the regulation of discoidin I expression. *Dev Biol* 159, 184-95.
- Wiater, L. A., Dunn, K., Maxfield, F. R., und Shuman, H. A. (1998). Early events in phagosome establishment are required for intracellular survival of *Legionella pneumophila*. *Infect Immun* 66, 4450-60.
- Wilkinson, H. W., Thacker, W. L., Steigerwalt, A. G., Brenner, D. J., Ampel, N. M., und Wing, E. J. (1985). Second serogroup of *Legionella hackeliae* isolated from a patient with pneumonia. *J Clin Microbiol* 22, 488-9.
- Wintermeyer, E., Ludwig, B., Steinert, M., Schmidt, B., Fischer, G., und Hacker, J. (1995). Influence of site specifically altered Mip proteins on intracellular survival of *Legionella pneumophila* in eukaryotic cells. *Infect Immun* 63, 4576-83.
- Wintermeyer, E., Rdest, U., Ludwig, B., Debes, A., und Hacker, J. (1991). Characterization of legiolysin (lly), responsible for haemolytic activity, colour production and fluorescence of *Legionella pneumophila*. *Mol Microbiol* 5, 1135-43.
- Wood, S. A., Ammann, R. R., Brock, D. A., Li, L., Spann, T., und Gomer, R. H. (1996). RtoA links initial cell type choice to the cell cycle in *Dictyostelium*. *Development* 122, 3677-85.
- Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., und Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30, e15.
- Yang, Y. H., und Speed, T. (2002). Design issues for cDNA microarray experiments. *Nat Rev Genet* 3, 579-88.

Zinda, M. J., und Singleton, C. K. (1998). The hybrid histidine kinase dhkB regulates spore germination in *Dictyostelium discoideum*. *Dev Biol* 196, 171-83.

8 Abkürzungsverzeichnis

A	Adenin
AS	Aminosäuren
ATCC	American Type Culture Collection
bp	Basenpaar(e)
BSA	bovines Serum-Albumin
C	Cytosin
ca.	circa
cAMP	zyklische Adenosinmonophosphat
cDNA	<i>copy</i> DNA (komplementäres DNA-Fragment einer mRNA)
cfu	colony forming unit (Kolonie-bildende Einheit)
CoA	Coenzym A
Cy3	Cyanine3
Cy5	Cyanine5
d. h.	das heißt
Da	Dalton
DEPC	Diethylpyrocarbonat
DMSO	Dimethylsulfoxid
DNA	<i>Desoxyribonucleic acid</i> (Desoxyribonukleinsäure)
dNTP	Desoxynukleotidtriphosphat (dATP, dCTP, dGTP und dTTP)
dUTP	Desoxyuridintriphosphat
EDTA	<i>Ethylenediaminetetraacetic acid</i> (Ethylendiamintetraessigsäure)
ER	Endoplasmatisches Retikulum
<i>et al.</i>	<i>et alii</i> (und andere)
G	Guanin
h	Stunden
HEPES	N-2-Hydroxyethylpiperazin-N'-ethansulfonsäure
l	Liter
M	Molar [mol/l]
min	Minute
MOPS	Morpholinopropansulfonsäure
PCR	<i>Polymerase chain reaction</i> (Polymerase-Kettenreaktion)
pH	potentia Hydrogenii (Säurestärke)
ppm	<i>parts per million</i> (Teile pro Million)
qPCR	quantitative PCR
RNase	Ribonuklease
rRNA	ribosomale Ribonukleinsäure
S	Svedberg, Einheit der Sedimentationskonstante
SAM	Significance Analysis of Microarrays
SDS	<i>Sodium dodecyl sulfate</i> (Natriumdodecylsulfat)
sp.	Spezies
T	Thymin
TBE	TRIS-Borat-EDTA
TE	TRIS-EDTA
TRIS	Tris(hydroxymethyl)aminomethan
UpM	Umdrehungen pro Minute
UV	ultraviolettes Licht
V	Volt
x g	Vielfaches der Erdbeschleunigung

Chemische Elemente und Verbindungen, Präfixe im Einheitensystem sowie physikalische Einheiten werden nach den internationalen Richtlinien abgekürzt.

Fachbegriffe:

<i>accession number</i>	Zugangsnummer GenBank, National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/)
<i>Add-in</i>	Ein Programm, das als Erweiterung der <i>Office</i> -Programme von Microsoft dient
<i>alignment</i>	Darstellung der Übereinstimmung von Sequenzen bei der BLAST-Suche
<i>cluster</i>	Klassen, Gruppen, Haufen
<i>dye swap</i>	Vertauschen den Fluoreszenzfarbstoffe für Experiment und Kontrolle von zwei Microarrays
<i>in vitro</i>	im zellfreien System
<i>in vivo</i>	im lebenden Organismus
<i>score</i>	Punktzahl, hier Wert der SAM-Statistik
<i>Visual Basic</i>	Programmiersprache

Danksagungen

Frau Prof. Dr. A. A. Noegel danke ich für die Möglichkeit im Institut für Biochemie I zu promovieren.

Herrn Prof. Dr. D. Tautz möchte ich für das Korreferat, sowie das zur Verfügung stellen des Microarray-Druckers des Instituts für Genetik in der Anfangsphase des *D. discoideum*-Microarray-Projekts danken.

PD Dr. Ludwig Eichinger danke ich für die großartige Unterstützung während meiner gesamten Arbeit.

Carina Wagner und Dr. Michael Steinert danke ich für Durchführung der Infektionen und die Hilfe bei *Legionella* Fragen.

Bedanken möchte ich mich auch bei Silke Neuhauser, Thi Ha Libotte und Christiane Neuhoff. Ohne ihre tausenden PCR-Reaktionen hätte es das *D. discoideum*-Microarray nicht gegeben. Sie haben außerdem für ein gutes und lustiges Arbeitsklima gesorgt.

Der Mensa-Gruppe (Martina Munck, Thorsten Libotte, Hafida Zaim, Akis Karakesisoglou, Jessica Berthold, Ludwig Eichinger und einigen anderen) danke ich für die vielen geselligen Mahlzeiten mit all ihren Geschichten, Witzen und Gerüchten.

Bettina Lauss möchte ich für ihre zuvorkommende Hilfsbereitschaft und die „Essensreste“ danken.

Deenadayalan Bakthavatsalam danke ich für seine Unterstützung bei den Northernblots und seiner steten Hilfsbereitschaft.

Bärbel Bock und Roberto Campanella danke ich für die Bereitstellung der Puffer und des Materials.

Ich danke allen Mitgliedern des Instituts für Biochemie für die freundliche Arbeitsatmosphäre.

Anke Seesing danke ich für die Korrektur dieser Arbeit und ihren Zuspruch.

Besonderer Dank gebührt meinen Eltern und meiner Schwester die mir immer zur Seite standen.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Frau Prof. Dr. A. A. Noegel betreut worden.

Köln, im April 2004

Teile dieser Arbeit wurden veröffentlicht in:

Farbrother, P., Müller, S., Noegel, A. A., und Eichinger, L. (2002). Comparison of Probe Preparation Methods for DNA Microarrays. *BioTechniques* 33, 884-888.

Libotte, T., Farbrother, P., Tunggal, B., und Eichinger, L. (2004). Using active contours for signal extraction from microarray images. Zur Veröffentlichung eingereicht.

Lebenslauf

Patrick Farbrother
 Görresstraße 4
 40597 Düsseldorf
 02 11 / 8 79 37 47
 patrick.farbrother@uni-koeln.de

geboren am 21.04.1975 in Düsseldorf
 Staatsangehörigkeit: deutsch
 ledig

Promotion

seit 04/2001 Dissertation am Institut für Biochemie I, Medizinischen Fakultät,
 Universität zu Köln mit dem Ziel der Promotion in der Mathematisch-
 Naturwissenschaftlichen Fakultät der Universität zu Köln. Thema:
 „Untersuchung der Genexpression von *Dictyostelium discoideum* nach
 Infektion mit *Legionella* mittels DNA-Microarrays“

Studium

10/1995 – 11/2000 Heinrich-Heine-Universität, Düsseldorf
 Studium der Biologie, Diplomarbeit: „Untersuchungen zur Charakteri-
 sierung der Malat-Dehydrogenasen von *Corynebacterium glutamicum*“
 Diplomprüfung am 21.11.2000

Zivildienst

08/1994 – 10/1995 Zivildienst in der Jugendherberge Radevormwald

Schulbildung

08/1985 – 06/1994 Gymnasium Koblenzer Straße, Düsseldorf
 Abitur

08/1981 - 07/1985 Montessori Grundschule Reinhold-Schneider-Str., Düsseldorf

Sprachen

Englisch
 Lateinisch

Köln, den 8. April 2004