

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie der  
Ludwigs-Maximilians-Universität München

Etablierung der  
DNA-Mikroarray-Transkriptom-Analyse  
für *Halobacterium salinarum* R1

Jens Twellmeyer  
aus  
Bramsche

2007

## Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. D. Oesterhelt betreut.

## Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 18.12.2006

.....  
Jens Twellmeyer

Dissertation eingereicht am 18.12.2006

1. Gutachter Herr Prof. Dr. D Oesterhelt
2. Gutachter Herr Prof. Dr. K. Förstemann

Mündliche Prüfung am 14.02.07

Für meine Eltern

## Inhaltsverzeichnis

1. Messung einer exemplarischen Genexpression mittels Mikroarrays.....	7
1.1 Einleitung.....	7
1.2 Grundlagen von DNA-Hybridisierungs-Arrays .....	8
1.3 Verschiedene Arten von DNA-Hybridisierungs-Arrays.....	10
1.3.1 Makroarrays.....	11
1.3.2 Mikroarrays.....	12
1.3.3 Hochdichte Oligonukleotid-Arrays (GeneChip®) .....	13
1.3.4 Mikroelektronik-Arrays (NanoChip®).....	14
1.4 Konstruktion und Herstellung der Sonden.....	16
1.5 Herstellung der markierten cDNA.....	17
1.6 Hybridisierung markierter cDNAs auf DNA-Mikroarrays.....	18
1.7 Signaldetektion von Mikroarrays.....	19
1.8 Datenanalyse von Mikroarrays.....	21
1.8.1 Bildanalyse von Mikroarrays.....	22
1.8.2 Vorverarbeitung und Normalisierung der Daten.....	23
1.8.3 Statistische Datenanalyse .....	26
1.8.4 Visualisierung der Daten.....	28
2. Halophile Archaea.....	30
2.1 Halobakterien.....	30
2.2 Archaeale Transkription .....	35
3. Etablierung von H. sal. R1 DNA-Mikroarrays.....	39
3.1 Herstellung der DNA-Mikroarrays.....	39
3.1.1 Sondendesign und Herstellung.....	39
3.1.1.1 Primerselektion zur Sondenherstellung.....	39
3.1.1.2 Qualitätskontrolle und Reinigung der PCR-Produkte.....	41
3.1.2 Design der DNA-Mikroarrays .....	42
3.1.3 Produktion der Mikroarrays .....	44
3.1.3.1 Produktion der Mikroarrays auf verschiedenen Trägern.....	46
3.1.3.1.1 DNA-Mikroarrays auf Poly-L-Lysin Slides.....	46
3.1.3.1.2 DNA-Mikroarrays auf CMT-GAPS II Slides.....	47
3.1.3.2 Beispiele verschiedener Mikroarrays von H. sal. R1 .....	48
3.1.4 Diskussion.....	51
3.2 Durchführung der Experimente.....	54
3.2.1 Experimentdesign.....	54
3.2.2 Experimentdurchführung/Zellzucht.....	57

3.2.2.1 Wachstum unter aeroben oder phototropen Bedingung.....	58
3.2.2.1.1 Zellzucht für Vergleich adaptierter Zellen.....	58
3.2.2.1.2 Zellzucht für Zeitreihenexperimente.....	61
3.2.2.1.3 Vergleich verschiedener Stämme S9-WT / S9-hell-Mutante.....	62
3.2.3 Diskussion.....	64
3.3 Herstellung der markierten cDNA.....	67
3.3.1 Probennahme und Isolierung der Gesamt-RNA .....	67
3.3.1.1 Qualitätskontrolle und Quantifizierung der RNA.....	68
3.3.2 Synthese der markierten cDNA.....	69
3.3.2.1 Bestimmung der Einbaueffizienz .....	69
3.3.3 Diskussion.....	70
3.4 Bestimmung der Expressionsunterschiede.....	70
3.4.1 Mikroarray-Hybridisierung.....	71
3.4.2 Bildakquisition der Mikroarray-Chips.....	71
3.4.3 Bildanalyse der Mikroarray-Chips.....	71
3.4.5 Datenanalyse.....	73
3.4.6 Diskussion.....	73
4. Statistische Analyse der Transkriptom-Daten.....	74
4.1 Vorverarbeitung der Daten.....	77
4.1.1 Hintergrundkorrektur der Daten.....	77
4.2 Normalisierung der Daten.....	85
4.2.1 Normalisierung der einzelnen Mikroarrays.....	86
3.2.2 Normalisierung der Mikroarrays untereinander.....	89
4.3 Statistische Datenanalyse / Identifizierung differentiell exprimierter Gene .....	90
4.3.1 Vervollständigung des Datensatzes: „missing value imputation“ .....	90
4.3.2 Berechnung des MAANOVA Modells.....	92
4.4 Visualisierung der Daten.....	94
4.5 Diskussion.....	95
5. Differentiell exprimierte Gene.....	100
5.1 Anbindung an die HaloLex-Datenbank.....	102
5.2 Differentiell exprimierte Gene des Zweigruppenvergleichs.....	104
5.2.1 Diskussion.....	109
5.3 Differentiell exprimierte Gene der Zeitreihenexperimente.....	114
5.3.1 Diskussion.....	122
5.4 Interpretation der Daten und weitere Experimente.....	146
5.4.2 Deletion einzelner Gene .....	147

5.4.1.1 Diskussion .....	156
5.4.2 Stoffwechselwege .....	160
5.4.2.1 Diskussion.....	165
5.4.3 Analyseansatz der S9H-Mutante.....	165
5.4.3.1 Diskussion: .....	172
6. Validierung der Daten / Ergebnisse.....	177
6.1 Quantifizierung mittels RT-PCR .....	178
6.2 Interpretation auf Proteinebene .....	188
7. Mikrobiologische Methoden .....	193
7.1 Zellzucht.....	193
7.1.1 Propagierung und Aufbewahrung von Escherichia coli.....	193
7.1.2 Propagierung und Aufbewahrung von Halobacterium salinarum.....	193
Wechsel vom aeroben auf phototrophes Wachstum .....	194
7.2 Aufnahme einer Wachstumskurve.....	194
7.3 Antibiotika .....	194
8. Molekularbiologische Methoden.....	195
8.1 Isolierung chromosomaler DNA aus H. sal. R1.....	195
8.2 Isolierung genomischer DNA aus H. sal. R1.....	196
8.3 Isolierung von Plasmid-DNA aus E. coli.....	196
8.4 Isolierung von Gesamt-RNA aus H. sal. R1 .....	196
8.5 Konzentrationsbestimmung von doppelsträngiger DNA (dsDNA).....	197
8.6 Herstellung von RbCl-kompetenten Bakterien.....	197
8.7 Ligation von DNA-Fragmenten.....	198
8.8 Transformation kompetenter Bakterien.....	198
8.9 Präparation kompetenter H. sal. R1 Zellen und Transformation.....	199
.....	200
8.10 Spaltung von DNA mit Restriktionsendonukleasen .....	200
8.11 Polymerasekettenreaktion.....	201
8.12 DNA-Sequenzierung.....	202
8.12.1 Reinigung von Sequenzierungsproben.....	203
8.13 Elektrophoretische Auftrennung von DNA-Fragmenten.....	203
8.14 Elektrophoretische Auftrennung von RNA-Fragmenten.....	203
8.15 Extraktion von DNA aus Agarosegelen .....	204
8.16 Southern-Hybridisierung mit Digoxigenin(DIG) -markierten Sonden. .	205
8.16.1 Herstellung DIG-markierter Sonden.....	205
8.16.2 Fixierung der DNA auf Membranen (Vakuum-Blot).....	205
8.16.3 Hybridisierung DIG-markierter DNA-Sonden.....	206

8.16.4 Detektion der hybridisierten DNA.....	207
8.17 Northern-Hybridisierung mit radioaktiv markierten Sonden.....	208
8.17.1 Herstellung der radioaktiv markierten Sonden.....	208
8.17.2 Fixierung der RNA auf Membranen (Vakuum-Blot).....	208
8.17.3 Hybridisierung radioaktiv markierter RNA-Sonden.....	208
8.18 Quantitative RT-PCR .....	209
9. Zellfreie Expression von Proteinen.....	210
9.1 In vitro Expression (RTS-100 System).....	210
9.1.1 Einleitung.....	210
9.2 Herstellung des linearen Expression-Templates.....	211
9.3 In vitro Protein Expression und Biotinylierungsreaktion.....	213
9.3.1 SDS-PAGE und Streptavidin Western Blot.....	214
10. Biochemische Methoden.....	215
10.1 Schnelltest für die Konzentrationsmessung von BR.....	215
10.2 Isolierung der BR-Membranfraktion.....	215
10.3 Sauerstoffmessung in H. sal. R1 Kulturen.....	216
11. Transkriptom-Analyse.....	216
11.1 Herstellung des DNA-Mikroarrays .....	216
11.1.1 Primeroptimierung zur Herstellung der Sonden .....	216
11.1.2 Herstellung der Sonden .....	218
11.1.3 Aufreinigung der Sonden-PCRs.....	220
11.1.4 Herstellung der DNA-Mikroarrays .....	221
11.1.5 Weiterverarbeitung der DNA-Mikroarrays .....	222
11.2 Zellzucht für die Expressionsexperimente .....	223
11.3 Herstellung der zu hybridisierenden DNA .....	224
11.3.1 Isolierung von Gesamt-RNA aus Halobacterium salinarum.....	224
11.3.2 Synthese der fluoreszenzmarkierten cDNA .....	225
11.4 Hybridisierung.....	227
11.4.1 Prähybridisierung der DNA-Mikroarrays.....	227
11.4.2 Hybridisierung der fluoreszenzmarkierten cDNAs.....	227
11.4.3 Waschen der DNA Mikroarrays.....	229
11.5 Bildanalyse und Rohdatenerhebung (Signal Detektion).....	230
11.5.1 Bildanalyse.....	230
11.5.2 Rohdatenerhebung.....	232
11.6 Auswertung der Hybridisierungsdaten .....	234
11.6.1 Was ist R.....	235
11.6.2 Statistische Auswertung der Mikroarray-Daten von H. sal. R1 .....	235

11.6.3 Vorverarbeitung und Normalisierung der Daten.....	236
11.6.4 Statistische Datenanalyse.....	238
11.6.5 Visualisierung der Daten.....	239
12. Material .....	241
12.1 Biologisches Material.....	241
12.1.1 Bakterienstämme.....	241
12.1.2 Antikörper.....	241
12.1.3 Enzyme und Proteine.....	241
12.1.4 Standardgrößenmarker für Proteine und Nukleinsäuren.....	242
12.1.5 Synthetische Oligonukleotide.....	242
12.2 Chemikalien.....	244
12.3 Kits.....	245
12.4 Sonstige Materialien .....	246
12.5 Geräte.....	246
12.6 Computerprogramme, Datenbanken und Internet-Seiten.....	248
13. Zusammenfassung .....	249
14. Abkürzungsverzeichnis.....	253
15. Literatur.....	255
16. Danksagung.....	273
17. Lebenslauf.....	275





# 1. Messung einer exemplarischen Genexpression mittels Mikroarrays

## 1.1 Einleitung

Proteine führen als die strukturellen Komponenten von Zellen und von Gewebe einige Schlüsselfunktionen im biologischen System aus. Die Synthese der Proteine erfolgt anhand eines Bauplans, der in den Genen festgeschrieben ist. Die Gene sind Abschnitte aus Desoxyribonukleinsäure (DNS/DNA) und sind für alle Zellen in einem Organismus gleich.

Die Genexpression umfasst auf dem Weg von der DNA zum Protein zwei Schritte, die Transkription der mRNA und die Translation der Proteine. Bei der Transkription dient ein Strang der DNA als Matrize für die Synthese der komplementären Botenribonukleinsäure (messenger RNA; mRNA), die als Informationsüberträger fungiert. Im zweiten Schritt, der Translation, wird die genetische Information der mRNA in die entsprechende Proteinsequenz übersetzt. Im Translationsprodukt, dem korrekt gefalteten Protein, wird die genetische Information funktionell wirksam (Parmigiani G, et al., 2003).

Allen Lebewesen, von Bakterien bis hin zu vielzelligen Organismen, ist gemeinsam, dass ihre Eigenschaften durch ihre Gene bestimmt werden. Dabei steuern die Gene sämtliche Lebensvorgänge in Wechselwirkung mit der Umwelt. Von entscheidender Bedeutung für die Funktion der Gene ist die Regulation ihrer Aktivität. Denn die Kontrolle darüber, zu welchem Zeitpunkt unter welchen Bedingungen und an welchem Ort diese gespeicherte Information abgerufen wird, ermöglicht in koordinierter Weise den Ablauf der höchst komplexen Lebensvorgänge. Diese präzise Kontrolle der Genaktivität, ist auf verschiedenen Ebenen realisierbar (Altmann T, 2002).

Die Analyse des Genexpressionsmusters auf mRNA-Ebene, ist zum Verständnis der Funktion eines Gens eine wichtige Voraussetzung. Denn als Ergebnis einer solchen Analyse erhält man Informationen über Ort, Zeit, Bedingung und die jeweilige Menge der vorliegenden mRNA. So lässt die Kenntnis des Expressionsprofils eines Gens weitreichende Schlussfolgerungen über die Funktion zu, die das betreffende Gen z.B. bei der Entwicklung des Organismus, seinen Stoffwechsellleistungen oder seiner Interaktion mit der Umwelt spielt. Die weiteren Ebenen der Regulation, wie die Kontrolle der Translation, des Transports, des Abbaus sowie der Aktivität des Proteins, können in erheblichem

Umfang die Wirkung das Gens modulieren und in manchen Fällen die transkriptionelle Regulation überdecken (Altmann T, et al. 2002).

## **1.2 Grundlagen von DNA-Hybridisierungs-Arrays**

Es stehen zur Genexpressionsanalyse mehrere verschiedene Methoden zur Verfügung. Zu den gebräuchlichsten Methoden, die es erlauben den Unterschied in der Genexpression zu untersuchen, gehören Northern-, Dot- bzw. Slot-Blot-Analysen, und quantitative RT-PCR (Freemann WM, et al., 2000).

Hauptnachteil all dieser Methoden ist neben den technischen Schwierigkeiten die Tatsache, dass sie ausschließlich die Analyse einiger weniger oder nur einzelner Gene pro Versuch ermöglichen.

Daher bietet die Mikroarray Technologie eine hervorragende Alternative, diese Einschränkungen zu überwinden.

DNA-Hybridisierungs-Arrays erlauben es, die Genexpressionsanalyse auf einen genomweiten Ansatz auszudehnen, denn sie gestatten es in einem Experiment simultan die Expressionsunterschiede von mehreren tausend Genen zu bestimmen (Rampal JB, 2001).

Das Funktionsprinzip eines DNA-Hybridisierungs-Arrays ähnelt herkömmlichen Hybridisierungstechniken der Molekularbiologie wie den Northern- oder Southern-Blot Analysen. Diese Verfahren nutzen die Eigenschaft der Nukleinsäuren, miteinander sequenzspezifisch zu hybridisieren. Unter Hybridisierung wird die nicht-kovalente Bindung zweier zueinander komplementärer Nukleinsäurestränge verstanden, die auf der Ausbildung von Wasserstoffbrückenbindungen zwischen den heterozyklischen Basen der Nukleinsäuremoleküle beruht. Aufgrund der hohen Spezifität dieser Watson-Crick-Basenpaarung kommt es nur bei einer perfekten Paarung der Basen (Guanin mit Cytosin und Adenin mit Thymin oder Uracil) zu einer ausreichend starken Bindung.

Bei einem DNA-Hybridisierungs-Array werden Nukleinsäuren bekannter genspezifischer Sequenz, die als Sonden bezeichnet werden, in großer Zahl und hoher Dichte mit Hilfe eines Roboters auf einen Träger orts aufgelöst aufgebracht und immobilisiert. Diese DNA-Hybridisierungs-Arrays werden daraufhin mit markierten Nukleinsäuren hybridisiert. Eine Möglichkeit zur Markierung ergibt sich durch den Einbau von radioaktiv oder Fluoreszenz markierten Nukleotiden während der reversen Transkription der mRNA in cDNA (complementary DNA; komplementäre DNA).

Eine Hybridisierung erfolgt nur zwischen exakt komplementären Nukleinsäuremolekülen. Nach Entfernung nicht gebundener cDNA-Moleküle ist die Intensität des gemessenen Signals zur Menge an hybridisierter Probe proportional. Da jede Position einer Sonde einem bestimmten Gen oder Genabschnitt entspricht, liefert die an dieser Position gemessene Signalintensität ein Maß für das relative Expressionsniveau dieses Gens. Abhängig von der Anzahl der verfügbaren Gensonden und der Dichte, mit der sie aufgebracht werden, können mit Hilfe solcher Arrays mehrere tausend oder zehntausend Gene gleichzeitig analysiert werden.

Alle DNA-Hybridisierungs-Arrays basieren auf den gleichen vier Arbeitsschritten:

- RNA-Isolierung aus den Proben,
- Synthese spezifisch markierter cDNA aus der isolierten RNA,
- Herstellung oder Kauf eines Arrays mit genspezifischen Sonden,
- Hybridisierung der markierten cDNA und Messung der Signalintensitäten der Gensonden.

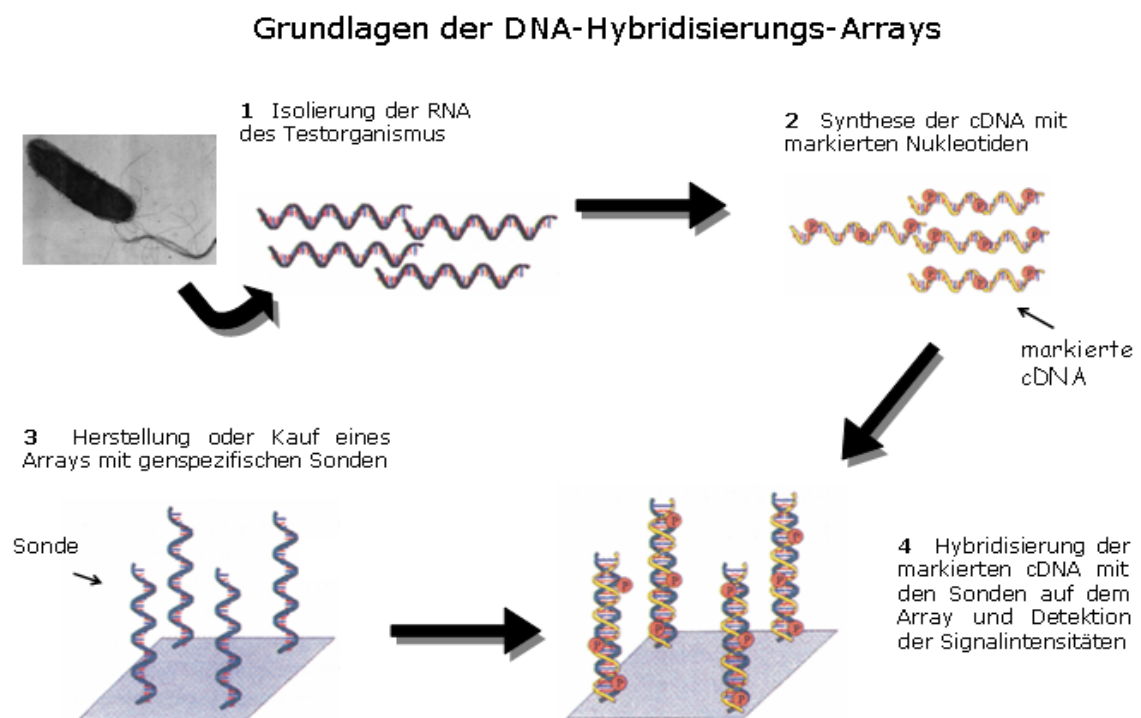


Abb. 1: Schematische Darstellung vier grundlegender Arbeitsschritte eines DNA-Hybridisierungs-Array-Experimentes.

### 1.3 Verschiedene Arten von DNA-Hybridisierungs-Arrays

Es existieren vier verschiedene Plattformen von DNA-Hybridisierungs Arrays, die sich in dem Trägermaterial, der Sonden Anzahl und Dichte, der Arraygröße und der Methode der Markierung unterscheiden. Derzeit lassen sich die verschiedenen Array Formate in vier Gruppen klassifizieren:

- Makroarrays
- Mikroarrays
- Hochdichte Oligonukleotid-Arrays (GeneChip®)
- Mikroelektronik-Arrays (NanoChip®)

	Anzahl der Gensonden	Trägermaterial	Markierungsmethode
Makroarray	200 - 500	ca. 12 x 8 cm große Nylon- Membranen 2,5 x 7,55 x 1 cm	radioaktiv / chemilumineszent
Mikroarray	≤ 2000	beschichtete Plastik- oder Glasträger	fluoreszent
Oligonukleotidarray	≤ 40000	1 x 1 cm große Halbleiterscheibe	fluoreszent
Mikroelektronikarray	64 - 1000	1 x 1 cm große Siliziumscheibe	fluoreszent

Tab. 1: Vergleich verschiedener Arten von DNA-Hybridisierungs-Arrays

Im Folgenden werden diese vier Gruppen näher beschrieben:

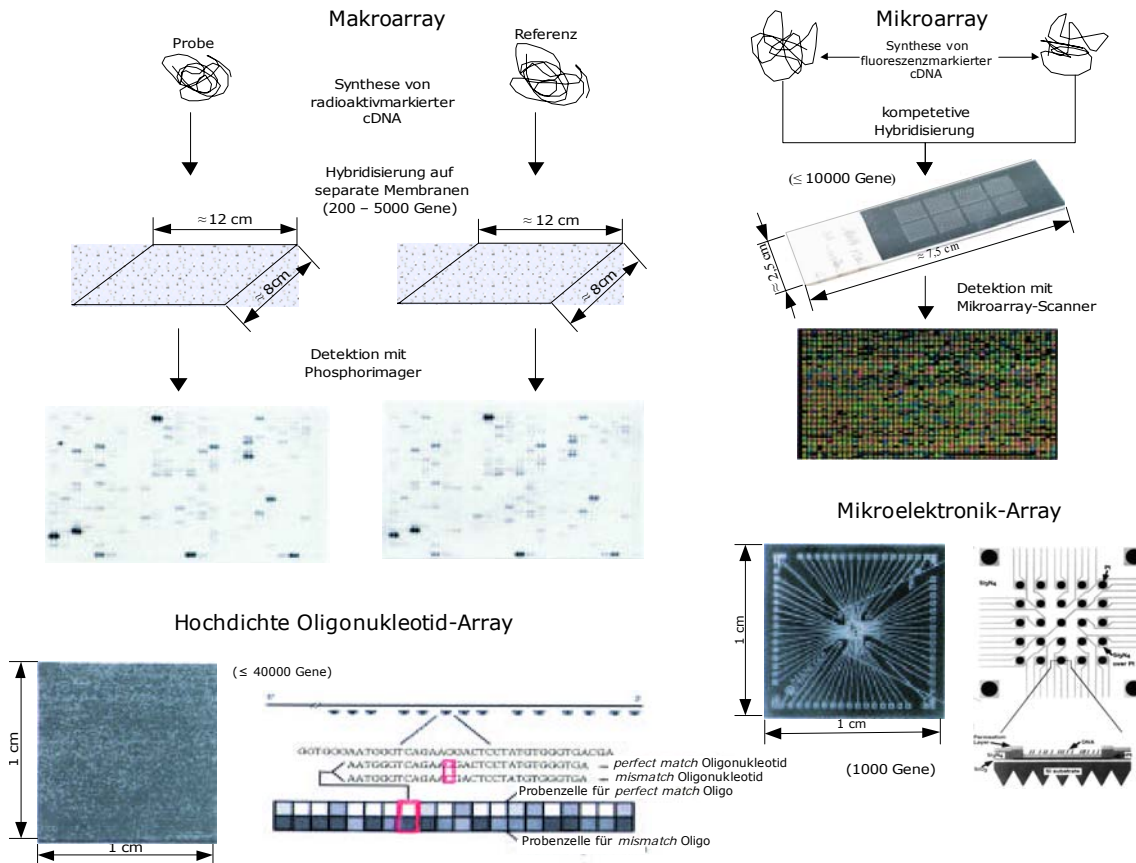


Abb. 2: Darstellung der vier Formate von DNA-Hybridisierungs-Arrays (nach Freeman WM, et al., 2000).

### 1.3.1 Makroarrays

Als Makroarrays werden diejenigen Arrays bezeichnet, bei denen die Gensonden mittels eines Roboters auf einer Membran immobilisiert werden. Der Begriff Makroarray bezieht sich auf die niedrigere Anzahl und Dichte der applizierten Sonden und somit auf den Grad der Miniaturisierung. Üblicherweise sind auf Makroarrays wenige Dutzend bis mehrere Tausend genspezifischer Sonden fixiert. Die Idee zur Herstellung von Makroarrays beruht auf ersten Hybridisierungsexperimenten mit cDNA-Bibliotheken (Gress TM, et al., 1992; Lennon GG, et al. 1991). Gegenwärtig werden DNA-Klone, PCR-Produkte oder Oligonukleotide auf Membranen aufgebracht.

Die Markierung der cDNA erfolgt, im Unterschied zu den anderen Array-Formaten, radioaktiv oder chemilumineszent. Aufgrund der radioaktiven Markierung, müssen unterschiedliche Proben auf zwei separaten Arrays hybridisiert werden.

Die Detektierung der radioaktiven Signale erfolgt auf indirektem Weg, durch Exposition von Phosphormolybdänplatten, oder weniger häufig, durch Autoradiographie eines Röntgenfilms.

### **1.3.2 Mikroarrays**

Mikroarrays unterscheiden sich in drei Punkten von Makroarrays:

- Als Trägermaterial wird Glas oder Plastik verwendet
- Die Markierung und Detektierung erfolgt über Fluoreszenz
- Sie weisen einen höheren Miniaturisierungsgrad auf als Makroarrays.

So besitzen Mikroarrays eine größere Anzahl von Gensonden bezogen auf eine kleinere Fläche. Ähnlich wie bei Makroarrays kann bei Mikroarrays ebenfalls ein breites Spektrum an Sonden, wie z.B. cDNA-Klone, PCR-Produkte oder Oligonukleotide, mit Hilfe eines Roboters auf das Trägermaterial gedruckt werden. Einer der ersten, der solche Mikroarrays produziert hat, war 1995 P.D. Brown an der Stanford Universität (Schena M, et al. 1995). Der große Vorteil von Mikroarrays gegenüber Makroarrays, liegt in ihrer Miniaturisierung. Denn nur so ist es möglich, eine Hybridisierung mit einem wesentlich geringeren Hybridisierungsvolumen durchzuführen, wodurch die Konzentration an zu hybridisierender cDNA erheblich vergrößert wird. Ebenfalls von Vorteil ist die so genannte kompetitive fluoreszente Hybridisierung, bei der zwei unterschiedlich markierte cDNAs, z.B. Probe und Referenz, die in zwei separaten Markierungsreaktionen mit verschiedenen Fluoreszenzfarbstoffen, wie Cy3 und Cy5, markiert wurden, auf einen Chip (Träger) hybridisiert werden. Die hybridisierten Arrays werden nach der Entfernung von nicht gebundenen markierten Molekülen mit Hilfe eines Fluoreszenzscanners ausgewertet. In der Regel werden beim Einsatz von zwei Fluoreszenzfarbstoffen die für jedes Gen erhaltenen Daten als Signalverhältnis der beiden Fluoreszenzfarbstoffe (Cy3/Cy5) angegeben. Zur Visualisierung werden diese Daten dann in Form von überlagerten roten (Cy5) und grünen (Cy3) Bildpunkten dargestellt. Ein Punkt, der grün erscheint, repräsentiert ein Gen, das unter einer Bedingung (z.B. in der Kontrollsituation) deutlich stärker exprimiert ist, als unter anderen Wachstumsbedingung des Testorganismus (z.B. in der Testsituation). Ein roter Punkt steht für ein Gen, welches unter der entgegengesetzten Bedingung exprimiert wird, als das Gen mit dem grünen Signal. In dem Fall, wo die

Transkriptmenge eines Gens unter den zu vergleichenden Bedingungen gleich ist, erscheint der Punkt im Verhältnisbild gelb, was der Mischfarbe aus rot und grün entspricht.

Mikroarrays werden in verschiedenen Formaten von unterschiedlichen kommerziellen Herstellern angeboten. Parallel dazu wird auch eine große Zahl von Mikroarrays z.B. von Forschungsinstituten speziell für einzelne Forschungsprojekte konzipiert und produziert.

### **1.3.3 Hochdichte Oligonukleotid-Arrays (GeneChip®)**

Die hochdichten Oligonukleotid-Arrays unterscheiden sich von den bereits besprochenen Formaten darin, dass die Proben (Gensonden) direkt auf der Oberfläche des Trägermaterials synthetisiert werden. Die Firma Affymetrix (Kalifornien, USA) hat ein photolithographisches Verfahren entwickelt, das eine *in-situ* kombinatorische Festphasensynthese der Oligonukleotide auf dem Trägermaterial ermöglicht (Lipshutz RJ, et al., 1999). Bei der für die Halbleitertechnik entwickelten Methode, werden auf einer Glasoberfläche, die mit einer photolabilen Schutzgruppe beschichtet wurde, mittels einer Lochmaske definierte Bereiche mit ultraviolettem Licht bestrahlt und durch Abspalten der lichtsensitiven Schutzgruppen für eine Reaktion aktiviert. Diese aktivierten Areale werden mit einem ausgewählten Nukleosid inkubiert, so dass es mit den Hydroxylgruppen der aktivierten Oberfläche reagieren kann. Jedes eingesetzte Nukleosid ist seinerseits mit einer lichtsensitiven Schutzgruppe an der 3'-Position der Riboseeinheit versehen. Wiederholt man diese Prozedur, können weitere Bereiche aktiviert und mit anderen Nukleosiden inkubiert werden. Daher kann durch wiederholtes Auflegen verschiedener Masken, anschließender Aktivierung mit Licht und Inkubation mit geschützten Nukleotiden ein dichter „Rasen“ definierter Oligonukleotide auf der Oberfläche synthetisiert werden (Fodor SP, et al., 1991; McGall, et al., 1996).

Der große Nachteil dieser Technik sind die hohen Kosten für die Lochmasken und die schlechte Ausbeute von 95% pro Reaktionszyklus, welche die Größe und Qualität der Oligonukleotide nachhaltig beeinflusst.

Für die Expressionsanalyse werden ca. 20 Oligonukleotide, die zu verschiedenen Sequenzabschnitten eines Gens exakt komplementär sind und als *perfect-match*-Oligonukleotide bezeichnet werden, auf dem Array synthetisiert. Zur



Bestimmung des Anteils der nicht spezifischen Hybridisierung, wird für jedes *perfect-match*-Oligonukleotid ein weiteres, so genanntes *mismatch*-Oligonukleotid auf dem Array aufgebracht. Der Unterschied des *mismatch*- zum *perfect-match*-Oligonukleotid besteht in einem Basenaustausch in der Mitte des *mismatch*-Oligonukleotids. Unter idealen Bedingungen findet zwischen der Probe und dem *mismatch*-Oligonukleotid keine Hybridisierung statt. Signale, die dennoch an der Stelle des *mismatch*-Oligonukleotids auftreten, müssen daher von unspezifisch hybridisierten Nucleinsäuren stammen. Durch einen Vergleich der Signalintensitäten zwischen dem *perfect-match*- und *mismatch*-Oligonukleotid kann entschieden werden, ob die Hybridisierung an die *perfect-match*-Sequenz spezifisch war oder nicht.

Bei einem anderen Verfahren der *in-situ* Synthese von Oligonukleotiden werden die Phosphoramidite mittels eines Roboters an bestimmten Stellen des Chips aufgebracht und durch schrittweise Reaktionszyklen die Oligonukleotide in der gewünschten Sequenz synthetisiert. Durch eine hydrophobe Oberfläche, die hydrophile Bereiche aufweist, wird die Produktion von gleichförmigen Punkten auf dem Chip ermöglicht (Hughes TR, et al., 2001).

### 1.3.4 Mikroelektronik-Arrays (NanoChip®)

Ein gänzlich anderes Format von DNA-Hybridisierungs-Arrays wurde von der Firma Nanogen (SanDiego, USA) entwickelt. Die Basis dieser *active programmable electronic device technology* bildet ein Siliziumchip mit einer Fläche von 1 cm<sup>2</sup>, der mit 25, 64 oder 100 Platinelektroden bestückt ist. Zur Herstellung wird auf einem Siliziumträger eine Schicht Siliziumdioxid erzeugt. Diese wird zunächst mit Aluminium und dann mit einem Photolack bedeckt. Mit Hilfe einer Maske wird der Lack an der Stelle, an denen später die Platinelektroden liegen sollen, entwickelt und anschließend entfernt. Auf den freigelegten Bereichen wird eine 20 nm dicke Grundierung aus Chrom aufgebracht, darüber die 500 nm dicke Elektrodenschicht aus Platin. Nun werden Lack und Aluminium entfernt, anschließend eine 2µm dicke dielektrische Schicht aus Trisiliziumtetranitrid (Si<sub>3</sub>N<sub>4</sub>) aufgetragen, die wiederum mit einem Photolack bedeckt wird. Genau über den Platinelektroden wird der Photolack sowie die Si<sub>3</sub>N<sub>4</sub>-Schicht entfernt, so dass die Elektroden von oben frei zugänglich und an den Seiten durch Si<sub>3</sub>N<sub>4</sub> isoliert sind. Nach der Entfernung des gesamten Lacks wird der Chip mit einer Gelpermeabilitätsschicht aus porösem Hydrogelmaterial (Agarose oder Polyacrylamid) beschichtet, die zuvor mit Streptavidin derivatisiert

wurde und die die Immobilisierung biotinylierter Moleküle auf der Oberfläche ermöglicht (Sosnowski RG, et al., 1997; Edman CF, et al., 1997; Gurtner C, et al., 2002) .

Für die Expressionsanalyse wird der Chip zum Aufbringen der Sonden mit einer Lösung aus biotinylierten Oligonukleotiden bedeckt. Durch wiederholtes Anlegen eines positiven Potentials an einzelne Elektroden können die negativ geladenen, biotinylierten Oligonukleotide gezielt an mit Streptavidin präparierte Stellen transportiert werden. Durch die hohe Bindungsaffinität der Streptavidin/Biotin-Wechselwirkungen ( $K_D=10^{-15}$  M) werden die Oligonukleotide an diesen Stellen gezielt immobilisiert. Das Umkehren des elektrischen Feldes am spezifischen Platz produziert einen Stringenzeffekt, wodurch die Hybridisierungsspezifität des Systems stark erhöht wird. (Heller MJ, et al., 2000) Das komplette System besteht aus der elektronischen Beladungskomponente, einem Fluoreszenzleser zur Detektion und einer Computer-Kontrolleinheit.

Wie unter 1.2 bereits erwähnt, besteht ein DNA-Mikroarray-Experiment aus mehreren Arbeitsschritten. Diese einzelnen Schritte im Ablauf eines DNA-Mikroarray-Experimentes werden im weiteren Verlauf näher beschrieben. Zur Verdeutlichung der Zusammenhänge der Abläufe eines solchen Experimentes, ist folgende Abbildung vorangestellt.

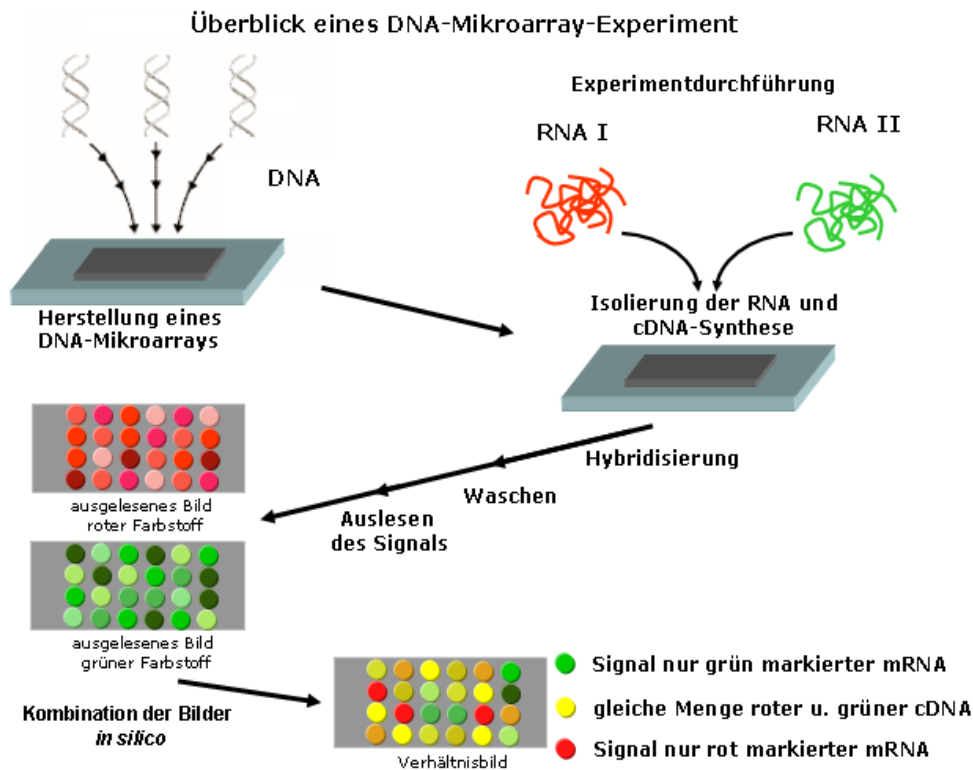


Abb. 3: Schematische Darstellung des Ablaufes eines DNA-Mikroarray-Experimentes

## 1.4 Konstruktion und Herstellung der Sonden

Die genetische Information von Mikroarrays ist in den auf dem Trägermaterial verankerten Nukleinsäuresequenzen gespeichert. Die Eindeutigkeit dieser Sequenzen bedingt welche Information man von einem Mikroarray Experiment bekommt und wie verlässlich das Ergebnis ist. Deshalb sind die Konstruktion und die Herstellung der Sonden, die auf das Trägermaterial gedruckt werden, Schlüsselschritte, die ohne die Fülle der zur Verfügung stehenden Sequenzdaten aus den verschiedensten Genomprojekten und Sequenzdatenbanken nicht möglich wären. Mit Hilfe der genetischen Information aus diesen Datenbanken ist man in der Lage die Sonden so zu konstruieren, dass sie organismus- und genspezifisch sind. Somit kommt es bei der Hybridisierung nicht zu unspezifischen Bindungen, die das Ergebnis des Experimentes in erheblichem Maße verfälschen würden. Aus diesem Grund sollten Sonden für homologe Gene so beschaffen sein, dass sie nur die Bereiche mit der geringsten Sequenzübereinstimmung umfassen.

Man unterscheidet drei verschiedene Arten von Sonden: cDNA-Klone, PCR-Produkte und Oligonukleotide.

cDNA-Sonden repräsentieren ganze Gene oder Fragmente von Genen, die aus cDNA-Bibliotheken stammen und nicht spezielle Bereiche mit niedriger Homologie

zu anderen Genen abdecken, wodurch es zu Kreuzhybridisierungen kommen kann. Ein weiterer Nachteil dieses Sondentyps ist die Heterogenität in seiner Sequenzlänge, die zwischen ein paar hundert Basen und mehreren Kilobasen variieren kann.

Sonden aus PCR-Produkten decken nicht homologe Bereiche zwischen 200-500 Basen eines Gens ab und können, ausgehend von genomischer DNA, cDNA-Bibliotheken oder RNA, anhand eines umfassenden Primer-Satzes mittels PCR hergestellt werden.

Die Oligonukleotid-Sonden können entweder *in-situ* auf der Trägerfläche synthetisiert werden, oder werden extern hergestellt und mechanisch auf die Oberfläche appliziert. Es stehen mehrere Verfahren zur *in-situ* Synthese von Oligonukleotiden zur Verfügung, mit denen Oligonukleotide von 25-80 Basen hergestellt werden können.

### **1.5 Herstellung der markierten cDNA**

Zur Expressionsanalyse werden zwei oder mehr RNA-Proben miteinander verglichen, um Unterschiede in der Transkriptmenge festzustellen. Damit man die Informationen, die in den verschiedenen Transkripten der biologischen Proben enthalten sind, auf einem Mikroarray überhaupt detektieren kann, müssen die Transkripte unterschiedlich markiert werden. Eine allgemeine Methode zur Markierung von cDNA stellt die reverse Transkription mit dem retroviralen Enzym „reverse Transkriptase“ dar, wobei z.B. ein radioaktiv- oder fluoreszenzmarkiertes Nukleotid in der Reaktionsmischung enthalten ist. Zur Einleitung der reversen Transkription benötigt man einen Primer, der entweder ein oligo(dT)-Primer, zur Markierung von eukaryontischer mRNA mit poly-(A) 3`-Enden, ein random Hexamer / Nonamer Primer, zur Markierung von Gesamt-RNA, oder ein genspezifischer Primer sein kann. Neben den molekularbiologischen Markierungsmethoden existiert auch die Möglichkeit die cDNA chemisch zu markieren.

Wenn die Menge an isolierter RNA ein limitierender Faktor ist, kann die RNA mit viralen RNA-Polymerasen amplifiziert werden. Hierbei ist darauf zu achten, dass die Kinetik der Amplifikation exakt zu reproduzieren ist, da anderenfalls gemessene Expressionsunterschiede durch Artefakte in der

Amplifikationseffizienz hervorgerufen wurden und nicht durch unterschiedliche Transkriptmengen bedingt sind.

## 1.6 Hybridisierung markierter cDNAs auf DNA-Mikroarrays

Eine Hybridisierung wird durchgeführt, um Nukleinsäuren in einer großen Probe zu identifizieren und zu quantifizieren. Dabei lagert sich ein Einzelstrang einer Nukleinsäurekette an seinen komplementären Zielstrang an. Ist der Nukleinsäureeinzelstrang markiert, so ist das detektierbare Signal zur Menge an markierter Probe proportional. Die Hybridisierung und das Waschen der DNA-Arrays sind wichtige Schritte, um hochwertige Genexpressionsdaten zu erhalten. Egal um welches DNA-Array-Format es sich handelt, muss immer die gleichmäßige Verteilung der Hybridisierungs- und der Waschlösungen gewährleistet sein, um eine optimale Wechselwirkung zwischen den komplementären Sequenzen der Probe und Sonde zu erhalten und eine Minimierung des Hintergrundsignals zu erzielen.

Bei einem DNA-Hybridisierungs-Array sind folgende Punkte zu beachten:

- Erzielen eines möglichst großen Signal-Hintergrund-Verhältnisses.
- Minimierung von Kreuzhybridisierungen der markierten cDNA und den Sonden.
- Gewährleistung der Proportionalität der Signalintensität zur Konzentration der markierten cDNA.

All diese Ziele können durch Optimierung einiger experimenteller Parameter, wie Bedingung und Zeit der Prä- und Hybridisierung, das Volumen der Hybridisierungslösung, Ionen- und Pufferbedingungen und der Menge an markierter cDNA in der Hybridisierungslösung erreicht werden. So reduzieren Puffer mit einer großen Ionenstärke die elektrostatischen Abstoßungen und begünstigen die Anlagerung von komplementären Sequenzen. Detergentien und Blockierungsreagenzien minimieren das Hintergrundsignal und Hybridisierungs- und Waschlösungen mit moderater bis hoher Stringenz unterdrücken Kreuzhybridisierungen.

Optimalerweise sollte die immobilisierte Sonde mindestens in einem zehnfachen Überschuss gegenüber der markierten cDNA vorliegen, damit während der Hybridisierung eine Kinetik pseudo-erster Ordnung für einen breiten

Konzentrationsbereich gilt. Bei limitierter Konzentration an immobilisierter Sonde, wird der dynamische Bereich des Systems verkleinert und die Bestimmung der differentiellen Expression komprimiert. Deshalb sollten die immobilisierten Sonden einen Konzentrationsbereich von 100-500 ng/ $\mu$ l abdecken. (Bowtell D und Sambrook J, 2003)

Es existieren verschiedene Methoden zur Hybridisierung von Mikroarrays, die automatisch oder von Hand durchgeführt werden können.

Eine dieser Methoden ist die so genannte Deckgläschen-Methode, bei der die Hybridisierungslösung, die die markierte cDNA enthält, auf dem Mikroarray unter einem Deckgläschen inkubiert wird. Zum Schutz vor Verdunstung der Hybridisierungslösung unter dem Deckgläschen, wird der Mikroarray mit dem Deckgläschen in eine Hybridisierungskammer gelegt und in einem Wasserbad oder einem Hybridisierungsöfen inkubiert. Die Hybridisierungskammer enthält neben einer Vertiefung für den Mikroarray auch Reservoirs für Wasser, so dass innerhalb der Kammer eine Luftfeuchtigkeit von über 95% erzielt werden kann, die so während der Hybridisierungszeit von ca. 16 Stunden konstant gehalten werden soll.

### **1.7 Signaldetektion von Mikroarrays**

Der Gebrauch von fluoreszierenden Farbstoffen zur Markierung der cDNA ist bei Mikroarrayexperimenten sehr weit verbreitet, denn sie erlauben eine hohe Sensitivität und ermöglichen die gleichzeitige Detektion von verschiedenen Proben innerhalb eines Experimentes. Die Unterscheidung verschieden markierter Fluoreszenzproben ist möglich, weil es bei Fluoreszenzfarbstoffen eine Differenz zwischen der Absorptions- und Emissionswellenlänge gibt, die so genannte Stokesche Verschiebung (Stoke`s Shift). Fluoreszenzfarbstoffe absorbieren Licht einer bestimmten Wellenlänge, wodurch Elektronen angeregt und in einen höheren Energiezustand angehoben werden. Von diesem angeregten Zustand werden Photonen einer bestimmten Wellenlänge emittiert, die für jede Verbindung spezifisch ist. Sind nun in einem Experiment zwei Proben mit verschiedenen Fluoreszenzfarbstoffen markiert, die einen gut detektierbaren Unterschied in ihrer Emissionswellenlänge besitzen, können die Signale der Farbstoffe differenziert werden.

Wenn bei der Anregung der Fluoreszenzfarbstoffe die Wellenlänge und die Intensität konstant gehalten werden, wie es z.B. unter Verwendung von Laser-Quellen der Fall ist, so bildet die Zahl der emittierten Photonen eine lineare

Funktion zu der Anzahl der vorhandenen fluorochromen Moleküle. Bei einer sehr hohen Fluorochromkonzentration ist das Signal nicht mehr linear, weil die Fluorochrommoleküle so dicht gepackt sind, dass eine Anregung nur direkt oder nahe an der Oberfläche der Probe / des Spots stattfinden kann. Des Weiteren kann es in diesem Fall eintreten, dass emittiertes Licht von einem anderen Fluoreszenzfarbstoffmolekül zur Anregung benutzt wird, man spricht dann von Selbstabsorption.

Eine Messung der emittierten Photonen, bzw. eine Messung der fluoreszenten Lichtintensität, kann mit jedem photosensitiven Gerät durchgeführt werden. Zur Bestimmung von sehr niedrigen Lichtintensitäten wird eine Photomultiplerröhre eingesetzt, die aus einer eingebauten photoelektrischen Zelle besteht, die beim Auftreffen von Licht mit ausreichender Energie auf die Photokathode Elektronen emittiert und den resultierenden Strom verstärkt. Dabei ist die Stromstärke proportional zur Intensität des detektierten Lichtes.

Die photochemischen Eigenschaften der eingesetzten Fluoreszenzfarbstoffe, wie die Absorptions- und Emissionswellenlänge, bestimmen die Spezifikationen der Mikroarrayscanner. Folgende generelle Eigenschaften eines Mikroarrayscanners zur Detektion von Fluoreszenzsignalen müssen erfüllt sein:

- Ausreichende Intensität, um eine große Fläche in einer möglichst kurzen Zeit abbilden zu können.
- Möglichkeit der Unterscheidung von mindestens zwei Fluoreszenzfarbstoffen.
- Abdeckung eines großen linearen Bereichs, zur Messung von unterschiedlich starken Fluoreszenzsignalen.

Alle oben genannten Eigenschaften eines Fluoreszenzbildverarbeitungssystems, werden durch folgende fünf Hauptbestandteile realisiert:

- Eine Anregungsquelle
- Eine optische Einheit zur Lichtleitung
- Eine optische Einheit zur Lichtsammlung
- Einen Emissionsfilter
- Eine Vorrichtung zur Detektion, Amplifikation und Digitalisierung des emittierten Lichtes

### Hauptkomponenten eines Fluoreszenzbilderkennungssystems

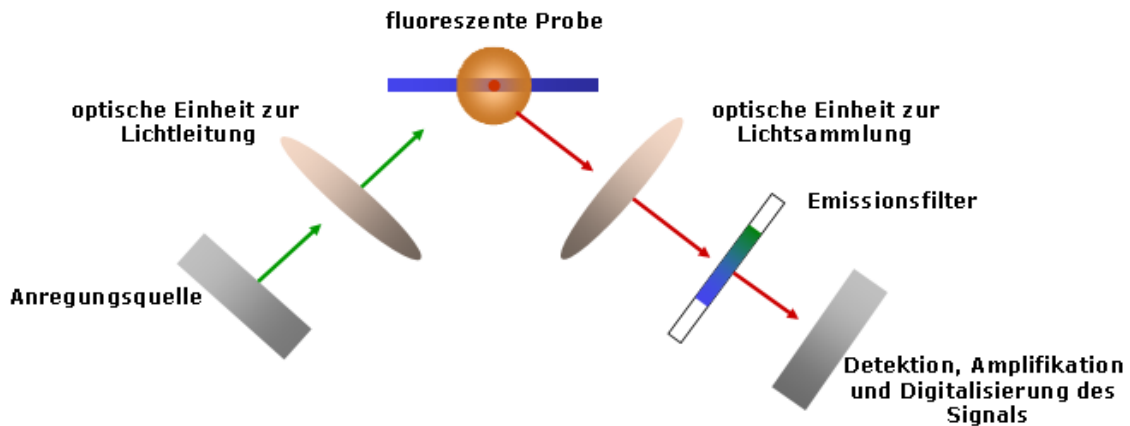


Abb. 4: Schematische Darstellung der Hauptkomponenten eines Fluoreszenzbilderkennungssystems (nach Microarray Handbook, Amersham Biosciences, 2002)

## 1.8 Datenanalyse von Mikroarrays

Die Arrayproduktion, die Hybridisierung und die Signaldetektion sind nur drei der anfänglichen Schritte einer Mikroarray-Expressionsanalyse, denn mit dem aufgenommenen Bild der verschiedenen Intensitätswerte der Gene, beginnt der nächste große Punkt, die Datenanalyse.

Methoden zur Analyse von Mikroarraydaten werden ständig weiterentwickelt und optimiert, denn bis jetzt gibt es noch keine universelle Datenanalysemethode oder ein Standard-Experimentdesign für Mikroarrays. Allerdings sind Bemühungen vorhanden eine einheitliche Benennung und Standards für Mikroarrays einzuführen, um eine Basis zu schaffen, auf der öffentliche Datenbanken für Mikroarray-Ergebnisse entstehen können (Brazma A, et al., 2000a; Brazma A, et al., 2000b; Brazma A, et al., 2001; Microarray Gene Expression Data Society <http://www.mged.org/Workgroups/MIAME/maime.html>).

Die Datenanalyse beginnt mit dem Experimentdesign, denn wenn man ein Mikroarrayexperiment plant, ist es wichtig, sich zu überlegen an welchen Stellen des Experiments Fehler auftreten könnten, die das Ergebnis verfälschen würden. So können Abweichungen von den gesammelten Proben ausgehen, wie z.B. unterschiedliche Kulturen oder verschiedene Gewebetypen. Des Weiteren treten im Ablauf des Experimentes selbst Unterschiede auf, die berücksichtigt werden müssen (Bowtell D und Sambrook J, 2003; Kerr MK, et al., 2001).



Um solche Fehler zu erkennen und zu beheben, sollten deshalb von einem Gen mehrere Sonden auf den Mikroarray aufgebracht und Wiederholungen der Hybridisierungen durchgeführt werden. Nur durch diese Wiederholungen ist es möglich, eine ausreichende Menge an Daten zu erheben, die es erlaubt statistische Verfahren anzuwenden, mit denen die experimentell bedingten Unterschiede zu kontrollieren und zu überwachen sind.

Eine weitere Möglichkeit ist die Durchführung eines Farbwechsel-Experiments (*Dye-Swap*), bei dem die beiden RNA-Proben, die miteinander verglichen werden sollen, separat mit beiden Fluoreszenzfarbstoffen Cy3 und Cy5 markiert und die zwei Farbkombinationen auf verschiedenen Mikroarrays hybridisiert werden. Bei diesem Verfahren können Unterschiede in der Einbaueffizienz der Fluoreszenzfarbstoffe ausgeglichen werden (WU H., et al., 2003).

Das letzte Stadium eines Mikroarrayexperiments ist die Durchführung der Datenanalyse. Dieser komplexe Prozess setzt sich aus folgenden vier Einzelschritten zusammen:

- Bildanalyse
- Vorverarbeitung und Normalisierung der Daten
- statistische Datenanalyse
- Visualisierung der Daten

### **1.8.1 Bildanalyse von Mikroarrays**

Die vom Mikroarrayscanner aufgenommenen Bilder des Arrays enthalten die Informationen über die Fluoreszenzintensitäten für jeden Fluoreszenzfarbstoff aller immobilisierter Sonden, inklusive der Pixel innerhalb (Vordergrundwert) und außerhalb (Hintergrundwert) des DNA-Punktes. Zur Analyse dieser Bilder werden bestimmte Computerprogramme eingesetzt, die unter anderem eine Quantifizierung der Signalintensitäten ausführen können (Bowtell D und Sambrook J, 2003)

Dazu wird im ersten Schritt eine Adressierung (gridding) durchgeführt. Bei diesem Prozess, der auch Punktfindung heißt, wird mit Hilfe des Computers ein Raster erstellt, das einem idealisierten Abbild des Arrays entspricht. Dieses Raster enthält Informationen über die Lage jeder Sonde und den jeweiligen Gennamen und wird mit dem Bild des Mikroarrays überlagert. Der Aufbau und das Layout dieses Rasters ist durch die Produktion der Mikroarray-Slides

vorbestimmt. Die dort enthaltenen Informationen werden bei der Punktfindung, die von Hand, halbautomatisch oder vollautomatisch mit Hilfe einer Bildverarbeitungssoftware durchgeführt werden kann, genutzt, um eine Zuweisung von Lagekoordinaten und Gennamen zu den einzelnen Punkten zu erhalten.

Der nächste Schritt der Bildanalyse ist die Segmentierung. Hierbei findet eine Art Zerlegung des Bildes statt, bei der verschiedenen Regionen des Bildes bestimmte Eigenschaften zu geordnet werden, wie z.B. Vordergrund- und Hintergrundsignal eines Punktes.

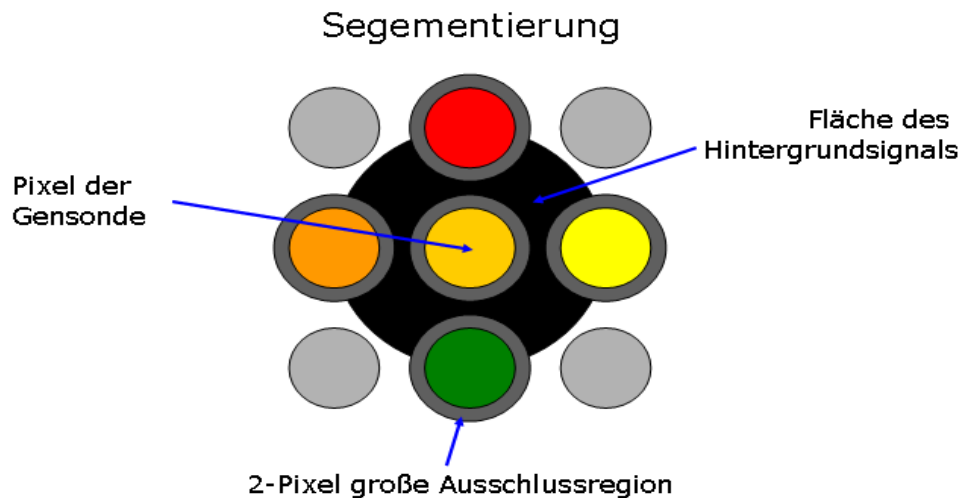


Abb. 5: Segmentierung; Grafische Darstellung der Bestimmung des lokalen Hintergrundwertes einer Gensonde (nach GenePix<sup>®</sup>Pro 6.0 User's Guide & Tutorial, 2004).

Der abschließende Schritt der Bildanalyse ist die Signalberechnung, bei der die Verhältnisse zwischen den verschiedenen Fluoreszenzintensitäten ermittelt werden. Die Signale sind als Mittelwert der Pixelintensitäten anzusehen, die nach der Segmentierung innerhalb einer Punktfläche zu finden sind.

### 1.8.2 Vorverarbeitung und Normalisierung der Daten

Es gibt bei einem Mikroarrayexperiment viele Quellen für Instabilitäten bzw. Artefakte, die alle Ebenen eines Experimentes, wie Sondenherstellung, Arrayproduktion, cDNA-Synthese, Hybridisierung und Bildanalyse, umfassen können, wie die folgende Liste verdeutlichen soll (Schuchhardt J, et al., 2000):

- RNA-Isolierung: unterschiedliche Qualität / Quantität bedingt durch verschiedene Effizienzen bei der Isolierung und oder durch Degradierung der RNA-Probe.

- Synthese der cDNA: verschiedene Einbaueffizienzen der Fluoreszenzfarbstoffe und unterschiedliche Längen der cDNAs.
- Sondenherstellung: Konzentrationsunterschiede bedingt durch PCR-Amplifikation und Reinigung der Produkte.
- Arrayproduktion: leichte Unterschiede zwischen den Drucknadeln verursachen Konzentrationsunterschiede beim Drucken der Sonden auf den Chip, die auch durch verschiedene Volumina der Sonden in den Ursprungsplatten (Quellplatten) bedingt sein können.
- Hybridisierungsparameter: Parameter der Hybridisierung wie Temperatur, Zeit, Pufferbedingungen und die Menge an eingesetztem Material sind nicht für alle Sonden optimal.
- Hybridisierung: unregelmäßige Verteilung der Hybridisierungslösung oder unterschiedliche Hybridisierungseffizienz verursachen eine Inhomogenität des Hybridisierungsprozesses und Kreuzhybridisierungen liefern falsch positive oder falsch negative Ergebnisse.
- Bildanalyse: nichtlineare Transmissions- und Sättigungseffekte, sowie unregelmäßig geformte Punkte erzeugen Schwankungen der Fluoreszenzsignale.

Deshalb sind die ersten Schritte der Datenanalyse die Hintergrundkorrektur und die Normalisierung, um die systematischen Unterschiede der Intensitätswerte auszugleichen.

Bei der Hintergrundkorrektur wird ein nicht spezifisches Hintergrundsignal von dem detektierten Signal abgezogen, damit ein Vergleich der spezifischen Signale durchgeführt werden kann. Allerdings kann durch diese Vorgehensweise eine größere Varianz bei den Daten auftreten, was besonders im unteren Intensitätsbereich kritisch ist. Als Hintergrund wird oft das Signal nicht belegter Bereiche verwendet. In den meisten Fällen treten jedoch räumliche Variationen des Hintergrundsignals auf, wodurch einige Spots stärker und andere schwächer betroffen sind. Eine mögliche Variante dieses Problem zu umgehen, ist komplett auf eine Hintergrundkorrektur zu verzichten, oder eine spezielle Datentransformation durchzuführen. So wird durch eine Logarithmustransformation der Expressionsdaten die Proportionalität zwischen dem statistischen Fehler und der Signalintensität aufgehoben und in ein additives Verhältnis überführt ( $\log(a/b) = \log(a) - \log(b)$ ). Außerdem kann für

logarithmierte Werte eine Normalverteilung angenommen werden, was für eine weitere statistische Analyse eine entscheidende Bedeutung hat.

Die Normalisierung wird am besten als ein iteratives Verfahren beschrieben, bei dem mögliche Artefakte veranschaulicht, identifiziert und wenn notwendig entfernt werden. Das Ziel einer Normalisierung ist also Schwankungen nicht biologischer Herkunft zu korrigieren, die zum einen zwischen den Proben auf einem Array auftreten und zum anderen zwischen den einzelnen Arrays eines Experimentes bestehen können (Bowtell D und Sambrook J, 2003).

Es gibt drei weit verbreitete Methoden zur Normalisierung von Expressionsdaten aus Mikroarrayexperimenten. Jede dieser Methoden basiert auf der Annahme, dass auf dem Array alle Gene (oder jedenfalls größtenteils alle), eine Untergruppe von Genen oder ein Set von exogenen Kontrollen, die vor der Markierung zu den RNAs dazugegeben wurden, nicht reguliert sind. Der resultierende Normalisierungsfaktor wird dann dazu verwendet, die Daten anzupassen, um die experimentellen Variabilitäten aufzuheben und die Fluoreszenzsignale abzugleichen (Quakenbush J, et al., 2001).

Normalisierung mittels globaler Intensitäten:

Die Gesamtintensitäts-Normalisierung beruht darauf, dass zur Markierung der zu vergleichenden Proben, dieselben Konzentrationen an RNA eingesetzt wurden. Des Weiteren wird angenommen, dass wenige Gene reguliert sind und sich im Hinblick auf die vielen tausend Gene auf dem Chip die Unterschiede herausmitteln. Ist die Gesamtmenge an RNA die gleiche so sind die Intensitäten aller Gene auf dem Array ebenfalls gleich für die Cy3- und Cy5-Farbstoffe. Unter diesen Voraussetzungen kann ein Normalisierungsfaktor berechnet werden, der zur neuen Skalierung der Intensitätswerte für jedes Gen auf dem Array verwendet werden kann.

Normalisierung mittels Regressionstechniken

Bei Expressionsanalysen von eng miteinander verwandten RNA-Proben kann angenommen werden, dass ein signifikanter Anteil der zu untersuchenden Gene auf demselben Niveau exprimiert wird. Wenn dies der Fall ist, würden die Gene in einem Scatterplot-Diagramm, bei dem die Intensitätswerte von Cy3 gegen die von Cy5 aufgetragen wurden, sich entlang der Winkelhalbierenden anhäufen und eine Gerade ergeben. Bei Markierung nur einer RNA-Probe mit zwei unterschiedlich fluoreszierenden Farbstoffen, sollte bei gleicher Markierungs- und Detektionseffizienz ebenfalls (bei einer grafischen Auftragung) eine Gerade mit

einer Steigung von eins ergeben. Eine Normalisierung solcher Daten ist äquivalent zu den Methoden, die dazu verwandt werden, eine Ausgleichsgerade mit Hilfe von Regressionsmethoden zu berechnen. In den meisten Experimenten sind die Auftragungen der Intensitäten nichtlinear, so dass lokale Regressionsmethoden, wie das LOWESS (L**O**cally **W**Eighted **S**catterplot **S**MOOTHING = lokal gewichtete Scatterplot Glättung), als viel tauglicher anzusehen sind (Cleveland WS, et al., 1979).

Normalisierung mittels Ratiostatistik:

Bei dieser Methode wird angenommen, dass innerhalb einer bestimmten Untergruppe von Genen, z.B. „*housekeeping genes*“, der Logarithmus vom Verhältnis Cy5 zu Cy3 im Durchschnitt Null ergibt. Von Chen et al. wurde ein iteratives Verfahren entwickelt, bei dem das arithmetische Mittel (*Mean*) des Expressionsverhältnisses gleich eins gesetzt und auf Grund eines berechneten Konfidenz-Intervalls bestimmt wird, welche Gene als unterschiedlich exprimiert anzusehen sind.

### 1.8.3 Statistische Datenanalyse

Das Ziel in diesem Stadium der Analyse ist es, einen verlässlichen, biologisch relevanten, Unterschied in der Expression festzustellen. Schwierigkeiten, diese Unterschiede im Expressionsverhalten zu detektieren, resultieren aus der Vermischung von systematischen Fehlereffekten und Versuchsdurchführung, was ein statistisch signifikantes Ergebnis liefert, aber keinen biologischen Ursprung besitzt. Ein weiterer Punkt, der eine Interpretation der Daten erschwert, sind zu wenige Wiederholungen, wodurch ein großer statistischer Fehler erzeugt wird. Statistische Ergebnisse müssen immer im Zusammenhang von Experimentdesign und Ziel der Studie interpretiert werden. Denn ein statistisch signifikantes Ergebnis kann eher systematische Fehler, statt eine zu Grunde liegende biologische Ursache widerspiegeln. Allerdings spiegelt ein Fehlen der statistischen Signifikanz eher eine zu geringe Sensitivität des Experimentdesigns und eine schlechte Versuchsplanung wieder, als die Abwesenheit (das Ausbleiben / das Fehlen) eines biologischen Effekts.

Traditionell werden signifikante Unterschiede bei der Expression von Genen mittels einer festen Schwellenwertmethode abgeleitet, z.B. zweifache Induktion oder Repression. Dies ist statistisch gesehen nicht relevant, da zahlreiche systematische und biologische Varianzen eines Mikroarrayexperimentes dadurch nicht berücksichtigt werden. Bedingt durch die Vernachlässigung dieser

Varianzen bei der Verwendung eines festen Schwellenwertes, vergrößert sich die Anzahl der falsch positiven und der falsch negativen Ergebnisse der Analyse. Ein besseres Bezugssystem zur Bestimmung der Signifikanz ist einmal eine Statistik, basierend auf einer ausreichenden Anzahl an Wiederholungen. Dabei wird die Reihenfolge der Genexpression nach der Wahrscheinlichkeit ihrer unterschiedlichen Expression aufgestellt und die Berechnung einer Ausschlussgrenze durchgeführt, die zur Ablehnung der Bedingungen der Nullhypothese dient, die besagt, ob ein Gen unterschiedlich exprimiert wird oder nicht (Leung YF, et al., 2003).

Wiederholungen der Mikroarrayexperimente sind zur statistischen Berechnung von Unterschieden in der Genexpression unverzichtbar. Deshalb wird vorgeschlagen jedes Mikroarrayexperiment mindestens dreimal zu wiederholen, um die Verlässlichkeit der Daten zu steigern. Dabei ist zu beachten, dass es zwei verschiedene Arten von Wiederholungen gibt, die biologische und die technische Wiederholung (Bowtell D und Sambrook J, 2003).

Als biologische Wiederholung wird die mehrmalige Analyse von unabhängigen biologischen Proben angesehen, wie z.B. individuelle Proben von einer Zelllinie oder Bakterienkultur gleicher Behandlung oder ein Gewebetyp von verschiedenen Patienten mit derselben Krankheit. Von einer technischen Wiederholung eines Mikroarrayexperimentes spricht man, wenn zur mehrmaligen Analyse ein und dieselbe isolierte RNA Probe benutzt wird. Mit Hilfe von biologischen Replikaten können in Expressionsanalysen besser die Variationen zwischen den unterschiedlichen Proben in der Expression ermittelt werden. Technische Replikate erlauben eine exakte Bestimmung der Expression, da sie die systematischen Varianzen eines Mikroarrayexperimentes entfernen, jedoch das Problem in der biologischen Varianz nicht lösen können. Aus diesen Gründen sind biologische Wiederholungen den technischen Wiederholungen vorzuziehen, vorausgesetzt, dass die Experimente sorgfältig durchgeführt wurden.

Statistische Methoden wie der Student'sche t-Test (Lönnstedt I, et al., 2002) und seine Varianten, ANOVA (ANalysis Of Variance) (Kerr MK, et al., 2000), „Bayesian“ Methoden oder der Mann-Whitney-U-Test (Long AD, et al., 2001; Wu TD, 2001) können dazu verwandt werden, Gene nach ihrem statistisch signifikanten Rang einer Analyse einzuordnen.

Die Festlegung eines Grenzwertes für die unterschiedliche Expression von Genen kann kompliziert sein, denn man muss bei diesem Vorgehen darauf achten, ein ausgewogenes Verhältnis von falsch-positiven (Fehler 1. Art) und falsch-

negativen Ergebnissen (Fehler 2. Art) zu erreichen. Außerdem ist bei der statistischen Analyse von zehntausenden von Genen zu berücksichtigen, dass das Problem des multiplen Hypothesen-Tests entsteht. Als Beispiel sei ein Mikroarrayexperiment mit zehntausend unabhängigen Genen angenommen, bei dem kein Gen differenziell exprimiert wird und das Signifikanzniveau  $\alpha$  auf 0,05 gesetzt wird. So kann man folgern, dass ( $10000 \times 0,05 = 500$ ) 500 Gene als signifikant reguliert angesehen werden, die in Wahrheit keine Unterschiede in ihrer Expression aufweisen. Zum Lösen des Problems des multiplen Testens werden konventionell konservative statistische Ansätze benutzt, die die *family-wise error rate* (FWER) kontrollieren unter der Annahme mindestens einen Fehler 1. Art zuzulassen (Dudoit S, et al., 2002). Ein klassisches Beispiel einer solchen konservativen statistischen Methode ist die Bonferroni-Korrektur. Allerdings kann die Kontrolle der FWER auch zu stringent sein, da das Signifikanzniveau  $\alpha$  nicht voll ausgeschöpft wird und dadurch die Kraft der Analyse sehr gering ist, wodurch nur wenige der tatsächlich differenziell exprimierten Gene entdeckt werden. Oftmals ist es annehmbar wenige falsch-positive Entscheidungen getroffen zu haben, wenn man dafür die Mehrheit der wirklich positiven Ergebnisse gewählt hat. Deshalb scheint es praktikabler zu sein die *false discovery rate* (FDR) zu kontrollieren, die das Verhältnis der Anzahl der Fehler 1. Art und der Anzahl der abgelehnten Hypothesen widerspiegelt (Reiner A, et al., 2003). In der Tat ist es anzuraten, die Ergebnisse einer statistischen Expressionsanalyse experimentell mittels RT-PCR, Northern-Blot oder *in-situ* Hybridisierung zu verifizieren.

### 1.8.4 Visualisierung der Daten

Als einen abschließenden Punkt der Datenanalyse kann man den Schritt der Profil- oder Mustererkennung innerhalb der Expressionsdaten ansehen. Dabei können verschiedene Methoden angewandt werden, die alle zum Ziel haben, die Gene einer Expressionsanalyse in unterschiedliche Gruppen einzuteilen, wobei die Gene jeweils einer Gruppe ein ähnliches Expressionsprofil aufweisen. Häufig eingesetzte Gruppierungsmethoden beinhalten die Hauptkomponentenanalyse (= *principal component analysis* (PCA)) oder die Singulärwertzerlegung (= *singular value decomposition* (SVD)), wodurch eine Dimensionsreduktion erzielt wird. Das hierarchische Gruppieren, das „k-means“ Gruppieren und selbstorganisierende Karten (= *self organizing maps* (SOM)) gehören ebenso zu den Methoden, die eingesetzt werden, um Daten in Segmente, Klassen oder Gruppen einzuteilen. Bei diesen Gruppen weisen deren Mitglieder untereinander eine hohe und zu

anderen Mitgliedern außerhalb dieser Gruppe eine möglichst geringe Ähnlichkeit auf. Diese Ähnlichkeiten werden durch Funktionen beschrieben, die Distanzfunktionen für Paare von Genen aufstellen, wobei gilt, je kleiner die Distanz zwischen den Genen, desto ähnlicher sind sie sich bzw., je größer die Distanz, desto unähnlicher sind sich die Gene. All diese Methoden gehören zu den nicht überwachten (*unsupervised*) Datenanalysemethoden, bei denen vorher kein Wissen über die Struktur vorhanden sein muss. Infolge dessen ist es zweckmäßig verschiedene nicht überwachte Methoden anzuwenden, die jeweils für verschiedene Parameter berechnet werden, damit man die Datenstruktur besser untersuchen kann. Des Weiteren ist zu beachten, dass die Stichhaltigkeit der Ergebnisse sowohl auf der statistischen Signifikanz, als auch auf dem biologischen Wissen bzw. der biologischen Relevanz basiert.

Als Abschluss dieses Kapitels werden die einzelnen Analyseschritte eines Mikroarray-Experimentes in Tab. 2 zusammengefasst.

Analyseschritt	Beschreibung
Experimentdesign und Durchführung	<ul style="list-style-type: none"> <li>• Eindeutig die biologische Fragestellung und Arbeits-hypothesen festlegen;</li> <li>• sorgfältige Planung des Mikroarrayexperimentes – Anzahl der Wiederholungen und Experimentdesign –</li> <li>• Vermeidung von experimentellen Fehlern</li> </ul>
Datenerhebung und Archivierung	<ul style="list-style-type: none"> <li>• Einhaltung der Standards zur Mikroarray Daten-erhebung z.B. nach MIAME (= <i>minimum information about a microarray experiment</i>)</li> </ul>
Bilderfassung	<ul style="list-style-type: none"> <li>• Vermeidung der Signalausbleichung</li> <li>• Bestmöglicher Abgleich der Cy-Farbstoffintensitäten</li> <li>• Scan des Bildes mit einer adäquaten Auflösung</li> </ul>
Bildanalyse	<ul style="list-style-type: none"> <li>• Sorgfältige Kontrolle der Rasterüberlagerung-Anpassung von Hand und Markierung schlechter Spots falls nötig-</li> </ul>
Vorverarbeitung der Daten	<ul style="list-style-type: none"> <li>• Ausschluss von Punkten mit schlechter Qualität</li> <li>• Berechnung der Ausschlussgrenze für Punkte im unteren Intensitätsbereich (keine Unterscheidung gegenüber Hintergrund möglich)</li> <li>• Logarithmustransformation der Intensitätswerte</li> </ul>
Normalisierung der Daten	<ul style="list-style-type: none"> <li>• Anwendung geeigneter grafischer Darstellung der Daten zur Datenevaluierung</li> <li>• in Erwägung ziehen LOWESS (= <i>locally weighted scatterplot smoothing</i>) oder eine Variante davon anzuwenden</li> </ul>



Analyseschritt	Beschreibung
Statistische Datenanalyse und Visualisierung	<ul style="list-style-type: none"> <li>• Keinen festen Schwellenwert zur Signifikanzabschätzung anwenden</li> <li>• Berechnung einer Statistik zur Festlegung einer Rangfolge der signifikant unterschiedlichen Gene</li> <li>• Bestimmung eines Ausschlusswertes zur Ablehnung der Nullhypothese das ein Gen differentiell exprimiert wird; Anpassung an multiples Testen</li> <li>• Anwendung verschiedener Analysemethoden zur Bestimmung des Expressionsprofils</li> <li>• Bestätigung der statistischen Ergebnisse durch nachfolgende Experimente</li> </ul>

Tab. 2: Zusammenfassung der Analyseschritte eine DNA-Mikroarray-Experimentes

## 2. Halophile Archaea

Die Zahl der Mikroorganismen, die in extremen Lebensräumen gefunden werden, steigt stetig an. Vertreter dieser Gattungen werden als Extremophile bezeichnet und gedeihen unter Umweltbedingungen, die extreme Temperaturen, Salzgehalte, pH-Werte oder Drücke aufweisen, von denen vorher angenommen wurde, dass sie lebensfeindlich seien. Ein großer Teil der Extremophilen zählt zu den Archaea, aber auch Bakterien und Eukaryonten sind in der Lage einige dieser extremen Lebensbedingungen zu tolerieren.

Halophile sind Organismen, die salzliebend sind und folglich Habitate mit außergewöhnlich hohen Salzkonzentrationen bevorzugen. Die hypersalinen Lebensräume der Halophilen sind über die ganze Welt verteilt und erstrecken sich über natürliche ausgedörrte oder küstennahe Solen und Tiefseegebiete, bis hin zu künstlichen Salinen zur Meersalzgewinnung.

Halophile können je nach ihrem Salzbedarf grob in drei Klassen eingeteilt werden, leicht, gemäßigt oder extrem halophil. Zu ihnen gehören prokaryontische und eukaryontische Mikroorganismen, die darauf spezialisiert sind, sich den hohen Salzkonzentrationen anzupassen und dem denaturierenden Effekt des Salzes zu widerstehen.

### 2.1 Halobakterien

Beispiele für gut angepasste und weit verbreitete extreme halophile Mikroorganismen, sind die archaealen *Halobacterium* Spezies. Diese extremen

Halophilen weisen ein Wachstumsoptimum bei sehr hohen Salzkonzentrationen von 3,4 – 5 mol/l Natriumchlorid auf und sind z.B. für die rote Färbung von einigen Salinen verantwortlich.

Halobakterien sind als Archaea klassifiziert, werden auch als halophile Archaea oder Haloarchaea bezeichnet und gehören zu der Familie der Halobacteriaceae. Die ersten mikrobiologischen Untersuchungen wurden mit einigen dem *Halobacterium* verwandten Stämmen ausgeführt, die am Anfang des zwanzigsten Jahrhunderts in Nordeuropa und Nordamerika von gesalzenem Fisch und Fleisch isoliert wurden. Hauptsächlich gehörten zu diesen Stämmen fakultative Aerobier, die neben Aminosäuren und anderen Wachstumsfaktoren, eine optimale Wachstumstemperatur von 38 – 40°C benötigen.

Es wird angenommen, dass Halobakterien keine organischen Osmolyte besitzen, aber eine sehr hohe intrazelluläre Kaliumsalzkonzentration von bis zu 5 mol/l enthalten. Der Kaliumgradient wird durch die Kombination eines elektrogeneren Natriumion / Protonen Antiporters und einem Kaliumuniporter aufrechterhalten (Ventosa A, et al., 1998; Oren A, 1999).

Die Proteine der Halobakterien sind beständig gegenüber hohen Salzkonzentrationen und benötigen sogar für ihre Aktivität einen gewissen Gehalt von Salz, was sich dadurch zeigt, dass die Proteine der Halophilen einen Überschuss an sauren Aminosäuren gegenüber basischen Aminosäuren besitzen. Durch die negative Oberflächenladung der Proteine wird unter hohen Salzkonzentrationen die Solvatisierung gewährleistet und verhindert, dass eine Denaturierung, Aggregation oder Präzipitation einsetzt, wie es bei nichthalophilen Proteinen der Fall wäre.

Einige Halobakterien haben Besonderheiten wie Gasvesikel, Purpurchromophore oder rotorangene Carotenoide, sowie die Fähigkeit ohne Sauerstoff wachsen zu können. Das Wachstum unter Sauerstoffausschluss erfolgt durch dissimilatorische Nitrat-Reduktion und Denitrifikation, Fermentation verschiedener Zucker, Abbau von Arginin und den Gebrauch der Lichtenergie, die mittels Retinalproteinen umgewandelt wird.

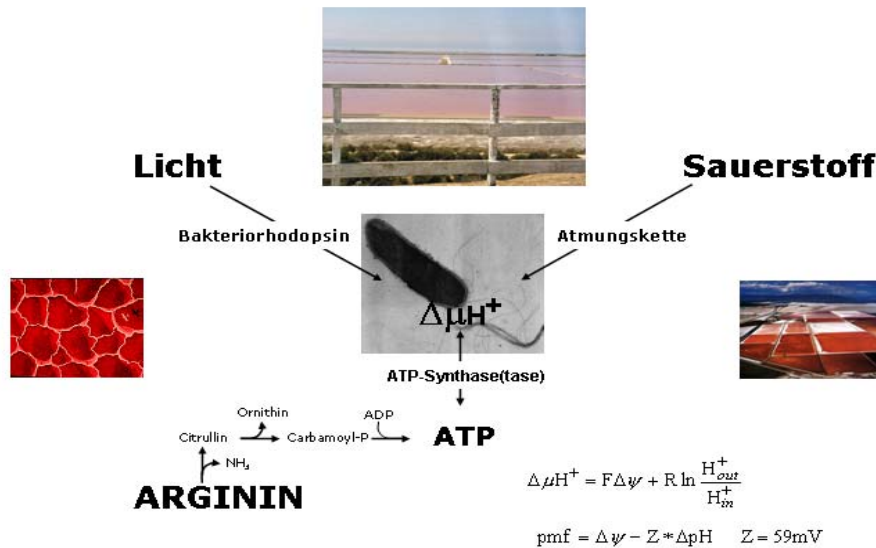


Abb. 6: Darstellung von drei möglichen Wegen zur Energieerzeugung in *Halobacterium salinarum*.

Die rotorangenen Carotenoide, die in einer hohen Konzentration vorliegen, stimulieren aktiv das Photoreparatursystem, bei dem Thymindimere repariert werden, die durch ultraviolette Strahlung entstanden sind. Die prominentesten Vertreter der Carotenoide sind die C-50 Bakterioruberine. Auch biosynthetische Intermediate wie das  $\beta$ -Carotin oder Lycopon können nachgewiesen werden.  $\beta$ -Carotin liefert durch eine Sauerstoff abhängige oxidative Spaltung das Retinal, welches in Retinalproteinen wie dem Bakteriorhodopsin (lichtgetriebene Protonenpumpe), dem Halorhodopsin (lichtgetriebene Chloridpumpe) oder den Sensorrhodopsinen (Lichtsensoren) kovalent als Chromophor gebunden ist.

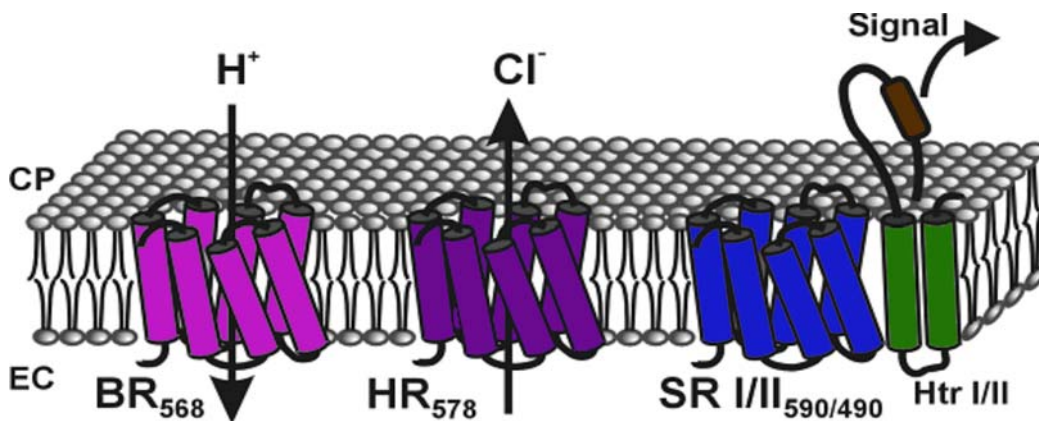


Abb. 7: Schematische Darstellung verschiedener archaealer Retinalproteine in der Membran; BR: lichtgetriebene Protonenpumpe Bakteriorhodopsin, HR: lichtgetriebene Chloridpumpe, SR I/II: Lichtsensoren Sensorrhodopsin I und II für UV und oranges Licht, im Komplex mit den jeweiligen Signalüberträgern HtrI/II (durch die Indizes werden die entsprechenden Absorptionsmaxima der Pigmente angegeben); CP: cytoplasmatisch, EC: extrazellulär (nach Oesterhelt D, 1998).

Die Purpurmembran stellt ein einmaliges Merkmal der Halobakterien dar, die als zweidimensionaler Kristall, aus Bakteriorhodopsin und Lipiden, in der Zellmembran vorhanden ist. Wie oben erwähnt, besteht das Bakteriorhodopsin aus einem Proteinanteil, dem Bakterioopsin, und einem kovalent gebundenem Chromophoren, dem Retinal und fungiert in der Zelle als eine transmembrane lichtabhängige Protonenpumpe. Bakteriorhodopsin wird durch niedrige Sauerstoffkonzentrationen und hohe Lichtintensitäten induziert und kann bis zu 50% der Zelloberfläche einnehmen (Oesterhelt D, 2004).

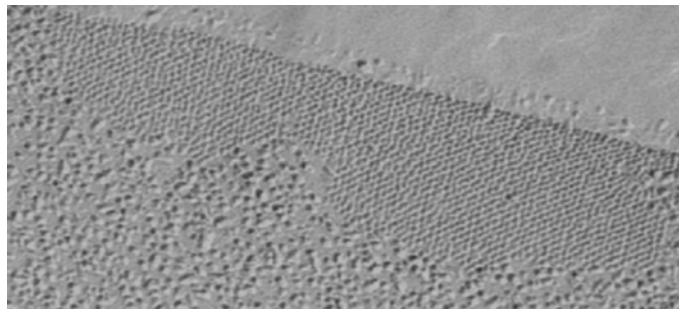


Abb. 8: Gefrierbruchaufnahme einer halobakteriellen Zelle mit deutlich sichtbaren Bereichen kristalliner Anordnung; hervorgerufen durch die zweidimensionale kristalline Struktur von Retinalproteinen (Bakteriorhodopsin oder Halorhodopsin) und Lipiden (Aufnahme von H. Besir).

Membranen weisen eine komplexe Struktur auf und bestehen aus einer Lipid-Doppelschicht (oder molekularen Schicht), in der Proteine integriert sind. Diese grundsätzlichen Eigenschaften der Membran werden durch das Model des flüssigen Mosaiks („fluidic-mosaic“) von Singer und Nicholson beschrieben (Singer SJ und Nicholson GL, 1972). Archaeale Lipide bestehen aus zwei Phytlyketten, die mittels einer Etherbindung an Glycerin oder andere Alkohole wie Nonitol gebunden sind. Im Gegensatz zu prokaryontischen oder eukaryontischen Lipiden, die mittels einer Esterbindung verknüpft sind.

Cytoplasmatische Membranen spielen eine entscheidende Rolle bei der Generierung von Stoffwechselenergie, die während des Katabolismus der Substratkettenphosphorylierung oder bei Energieübertragungsprozessen entsteht. Spezielle Pumpen transportieren Protonen oder Natriumionen von der cytoplasmatischen auf die äußere Seite der Zelle durch die Membran und erzeugen so einen elektrochemischen Gradienten (der Protonen oder Natriumionen). Wenn die Protonen ausgeschleust werden, übt der elektrochemische Gradient eine Kraft auf die Protonen aus, die „proton motive

force“ (PMF) protonen-motorische Kraft genannt wird. Diese PMF ist aus zwei Komponenten zusammengesetzt, einmal aus dem chemischen Protonen- oder pH-Gradienten, dem  $\Delta pH$  und zweitens aus dem Membranpotential, dem  $\Delta\Psi$ , welches durch den Transport von elektrischer Ladung erzeugt wird. Die PMF [mV] kann durch folgende Gleichung beschrieben werden:

$$PMF = \Delta\Psi - 2,3 \left( \frac{RT}{F} \right) \Delta pH$$

Mit der Gaskonstanten R, der absoluten Temperatur T [K] und der Faraday Konstanten F. Der Unterschied von einer pH-Einheit bewirkt eine Änderung von

$2,3 \left( \frac{RT}{F} \right)$ , was bei 25°C 59 mV entspricht. Die protonen-motorische Kraft kann

dazu genutzt werden Adenosindiphosphat (ADP) und Phosphat mittels der membranständigen ATPase in Adenosintriphosphat (ATP) umzuwandeln. Des Weiteren wird die PMF zum Transmembrantransport von Substraten, zur Rotation der Flagellen und zur Aufrechterhaltung des intrazellulären pH-Wertes gebraucht.

Halophile, wie *Halobacterium salinarum*, generieren den elektrochemischen Protonen Gradienten über die Membran durch die oxidative Phosphorylierung (Atmung) oder mittels der lichtgetriebenen Protonenpumpe Bakteriorhodopsin (Michel H und Oesterhelt D, 1980).

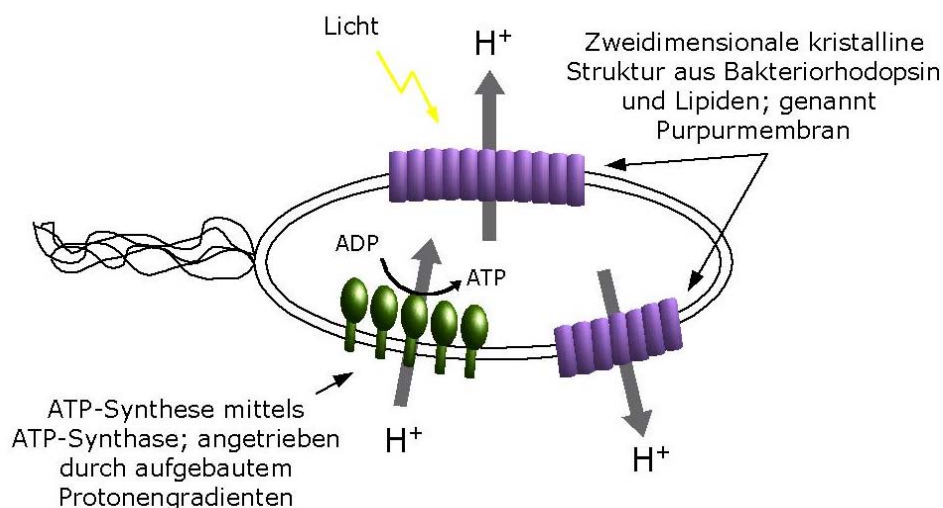


Abb. 9: Schematische Darstellung der halophilen archaealen Photosynthese (nach Oesterhelt D, 2004).

## 2.2 Archaeale Transkription

Halobakterien sind phylogenetisch unterschiedlich sowohl zu den Bakterien, als auch zu den Eukaryonten. Als Archaea haben sie spezielle Eigenschaften, wie einen eukaryontischen Transkriptions- und Translationsapparat, Ether verknüpfte Lipide und ähnlich den Bakterien eine Zellwand mit einem so genannten S-Layer, einer parakristallinen Oberflächenschicht aus Glykoproteinen.

Im Zuge der Charakterisierung der ersten archaealen DNA abhängigen RNA-Polymerase (RNAP) zeigte sich, dass die Hauptkomponenten der archaealen Transkriptionsmaschinerie der eukaryontischen RNA-Polymerase II ähneln (Zillig W, et al., 1979). Nach den ersten vollendeten archaealen Genomsequenzen bestätigen sich die Ergebnisse der ersten Versuche, und es konnten auch einige Transkriptionsregulatoren des bakteriellen Typs identifiziert werden, wodurch der chimäre Charakter des archaealen Transkriptionsapparates unterstrichen wurde (Aravind L und Koonin EV, 1999).

Ähnlich den eukaryontischen RNAPs ist die archaeale RNAP allein nicht in der Lage Promotorsequenzen zuerkennen. Es sind grundsätzlich zwei weitere Faktoren notwendig, um eine Transkription einleiten zu können.

Ein TATA-Box bindendes Protein (TBP), welches eine AT-reiche Initiatorsequenz (ca.8 Basenpaare (bp) lang) ca. 25 (bp) vor dem Transkriptionsstart erkennt und ein Transkriptionsfaktor B (TFB), der verwandt ist mit dem eukaryontischen Transkriptionsfaktor II B (Bell SD, et al., 2001; Soppa J, 1999).

### Schematische Darstellung der generellen Transkriptionsfaktoren in Archaea

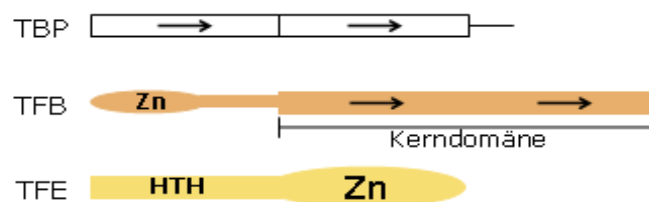


Abb. 9: Schematische Darstellung der generellen Transkriptionsfaktoren in Archaea. TBP (TATA-Box-bindendes Protein) besitzt eine unvollständige direkte Wiederholung. Einige Spezies weisen ein kurzes, stark saures C-terminales Ende auf. TFB (Transkriptionsfaktor B) ist aus zwei Domänen aufgebaut, einer N-terminalen „zinc ribbon“-Domäne und einer C-terminalen Kerndomäne, die aus einer unvollständigem Wiederholung eines Cyclin-Faltungsmotivs besteht. Beide Domänen sind durch einen hochkonservierten Verknüpfungsbereich voneinander getrennt. TFE (Transkriptionsfaktor E) besitzt N-terminal einen Leucinreichen Bereich, der ein „helix-turn-helix“-Motiv ausbildet und C-terminal eine „zinc ribbon“-Domäne (Bell SD, et al., 2001).

Dieser bindet an den TBP-DNA-Komplex, wodurch die spezifische Aktivierung (Initiation) der Transkription der RNA-Polymerase (RNAP) eingeleitet wird.

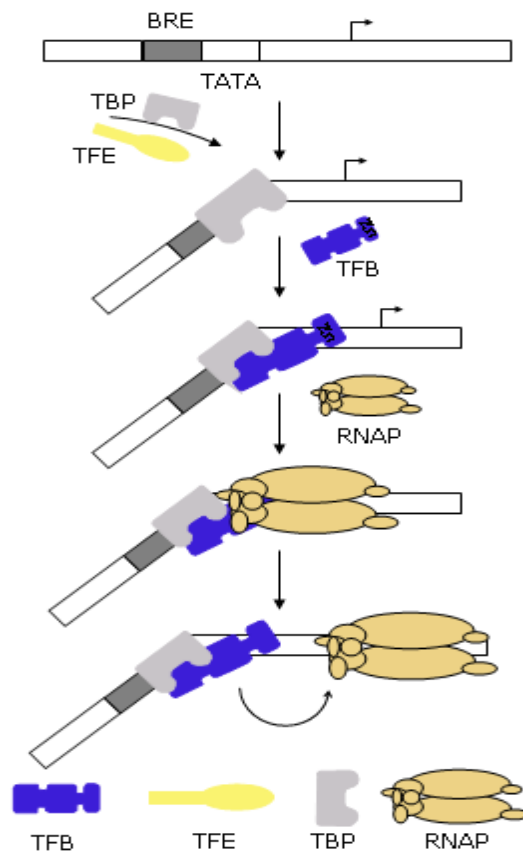


Abb. 10: Diagramm der archaalen Transkriptionsinitiation. Beim schrittweisen Aufbau des archaalen Initiationskomplexes wird die Bindung von TBP an die TATA-Box durch TFE unterstützt. Anschließend findet eine Bindung von TFB an den gebildeten TBP-DNA-Komplex statt, wodurch der sequenzspezifische Kontakt mit dem Transkriptionsfaktor B Bindungselement (BRE) hergestellt wird. Die N-terminale „zinc ribbon“-Domäne von TFB interagiert mit der RNA-Polymerase (RNAP), was zur Ausbildung des Pre-Initiationskomplexes führt. Die Freisetzung der RNAP von der Promotorsequenz wird durch die hochkonservierte Verbindungsregion zwischen der „zinc ribbon“- und der Kerndomäne von TFB unterstützt (siehe Pfeil in der letzten Zeile) (nach Bell SD, et al., 2001).

Eine Feinabstimmung bzw. Regulation des kontrollierten Aufbau des TBP-DNA-Komplexes wird durch eine ebenfalls der TATA-Box benachbarte purinreiche Transkriptionsfaktor-Erkennungs-Sequenz („transcription factor B recognition element“ (BRE)) ausgeführt (Bell SD, et al., 2001; Ouhammouch M, 2004).

Im nun folgenden Teil werden die Ergebnisse dieser Arbeit vorgestellt. Zur Veranschaulichung wird kurz beschrieben wie das Kapitel gegliedert ist.

Ausgangspunkt für dieses Projekt ist die Etablierung der DNA-Mikroarray-Technologie für das halophile Archaeon *H. sal.* R1. Wobei diese Methode die Basis für Transkriptomuntersuchungen darstellen soll, um die differentielle Genexpression nach einem Wechsel von aerobem auf phototrophes Wachstum zu analysieren.

Dies beinhaltet, die Identifizierung der regulierten Gene, das Erkennen von regulatorischen Mechanismen und die weitere Charakterisierung ausgewählter Gene durch unabhängige Methoden, wie z.B. RT-PCR und oder Proteomanalysen.

Als eines der ersten Teilziele soll die Herstellung der DNA-Mikroarrays realisiert werden, gefolgt von der Etablierung und Optimierung der Experimentdurchführung sowie der Synthese der zu hybridisierenden markierten cDNA.

Die statistische Datenanalyse, die Identifizierung der differentiellen Genexpression und die Validierung der erhaltenen Ergebnisse, repräsentieren dabei die späteren großen Teilziele dieses Projektes.

Ein Transkriptom-Experiment findet auf mehreren Ebenen statt und jede dieser Ebenen beinhaltet mehrere Arbeitsschritte. Zur Verdeutlichung der Komplexität des Experiments ist im Folgenden ein Flussdiagramm gezeigt, in dem die Organisation der notwendigen Arbeitsabläufe eines Transkriptom-Experiments mit *H. sal.* R1 dargestellt ist (Abb.11).



Organisation eines Transkriptom-Experimentes

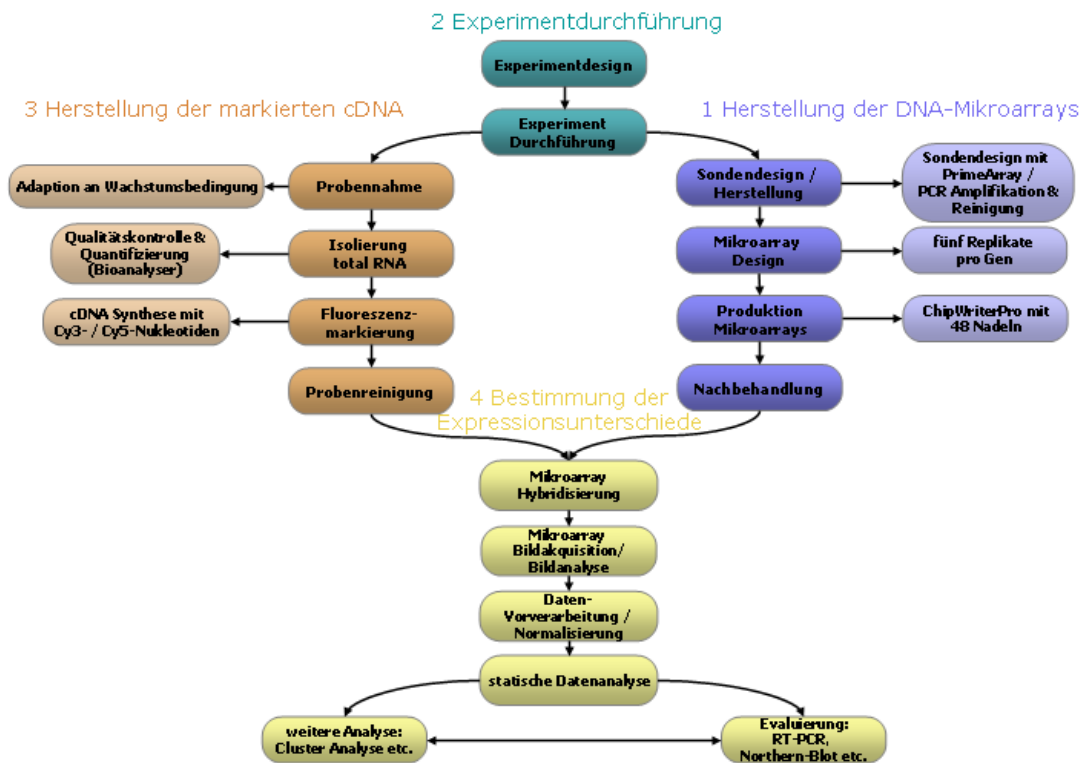


Abb. 11: Flussdiagramm der Organisation der Arbeitsabläufe eines Transkriptom-Experimentes von *H. sal.* R1.

Das Flussdiagramm der Transkriptom-Analyse ist in vier Teilbereiche unterteilt, die jeweils in verschiedenen Farben dargestellt sind. Begonnen wird mit der Herstellung der *H. sal.* R1 DNA-Mikroarrays, gefolgt von der Beschreibung der Experimentdurchführung und der Synthese der zu hybridisierenden markierten cDNA. Abgeschlossen wird das Kapitel mit der Identifizierung der Expressionsunterschiede, wobei die statistische Datenanalyse und die Validierung der Ergebnisse in separaten Kapiteln abgehandelt wird.

Aufgrund der aufgezeigten Komplexität der Arbeitsschritte, erfolgt die Diskussion der Ergebnisse direkt im Anschluß der einzelnen Teilabschnitte.

### **3. Etablierung von *H. sal.* R1 DNA-Mikroarrays**

Dieser Teil der Ergebnisse beschreibt in den nächsten Abschnitten die im Organisationsschema aufgezeigten Arbeitsschritte der Etablierung einer DNA-Mikroarray-Technologie für *H. sal.* R1.

#### **3.1 Herstellung der DNA-Mikroarrays**

Das erste Unterkapitel wird die Realisierung und den Weg zur Herstellung der DNA-Mikroarrays beschrieben.

##### **3.1.1 Sondendesign und Herstellung**

###### **3.1.1.1 Primerselektion zur Sondenherstellung**

Die Genomsequenz von *Halobacterium salinarum* R1 und dessen Annotation der Gene ist Ausgangspunkt zur Herstellung eines *H. sal.* R1 spezifischen DNA-Mikroarrays. Dieser DNA-Mikroarray basiert auf DNA-Fragmenten, die mittels spezifischer PCR-Primer amplifiziert werden. Es gibt zum Design und zur Optimierung von Primerpaaren für die PCR einige Computerprogramme, die mit Hilfe von ausgefeilten Algorithmen verhindern, dass z.B. Primer-Primer Wechselwirkungen auftreten. Keines dieser Programme war darauf ausgelegt, Primerpaare im Gesamtgenom-Maßstab zu berechnen.

Zu Beginn dieses Projektes bot jedoch das neu entwickelte Computerprogramm PrimeArray einen Ausweg (Raddatz G, et al., 2001). Dieses Programm erlaubt die automatische Extraktion von proteinkodierenden Sequenzen aus verschiedenen Eingabedateiformaten und ermöglicht so die schnelle und zuverlässige genomweite Primerselektion. Ziel der Primerauswahl und Optimierung ist, alle Primer genspezifisch mit möglichst ähnlicher Schmelztemperatur ( $T_m$ ), dem gleichen GC-Gehalt und minimalen Primer-Primer Wechselwirkungen zu erhalten.

Das Sondendesign wurde parallel zum Stand der Genomannotation von *H. sal.* R1 durchgeführt. Aus diesem Grund ergaben sich verschiedene Stadien der Sondenoptimierung und Herstellung, die im Folgenden näher beschrieben werden.

In Vorarbeiten zu diesem Projekt wurden für einen Test-Mikroarray 200 Sonden, von bis dahin bekannten Gensequenzen, mittels des Programms Primer3 hergestellt ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)). Zur

Optimierung dieser Sonden wurden die im experimentellen Teil erwähnten Zielvorgaben für T<sub>m</sub>, GC-Gehalt usw. verwandt.

Für die nächste Gruppe von 2000 „open reading frames“ (ORFs), basierend auf den Gensequenzen des Chromosoms von *H. sal.* R1, wurde das Sondendesign zum ersten Mal mit Hilfe des Programms PrimeArray durchgeführt. Die auf diese Weise abgeleiteten Sonden, dienen zur Produktion eines so genannten Chromosom-Mikroarrays, der zusammen mit den Sonden der ersten Gruppe, insgesamt 2200 genspezifische Sonden umfasst.

Nach Abschluss der Annotation des Genoms wurden für die noch fehlenden Gensequenzen die computergestützte Auswahl und Optimierung der Primerpaare durchgeführt. Die genspezifischen Sonden wurden mittels PCR im 96-iger Format amplifiziert, wobei als Matrize genomische DNA von *H. sal.* R1 eingesetzt wurde. Auf Grund einer sehr guten Effizienz der PCR-Amplifikation wurde eine Abdeckung des Genoms von 97,30% erreicht.

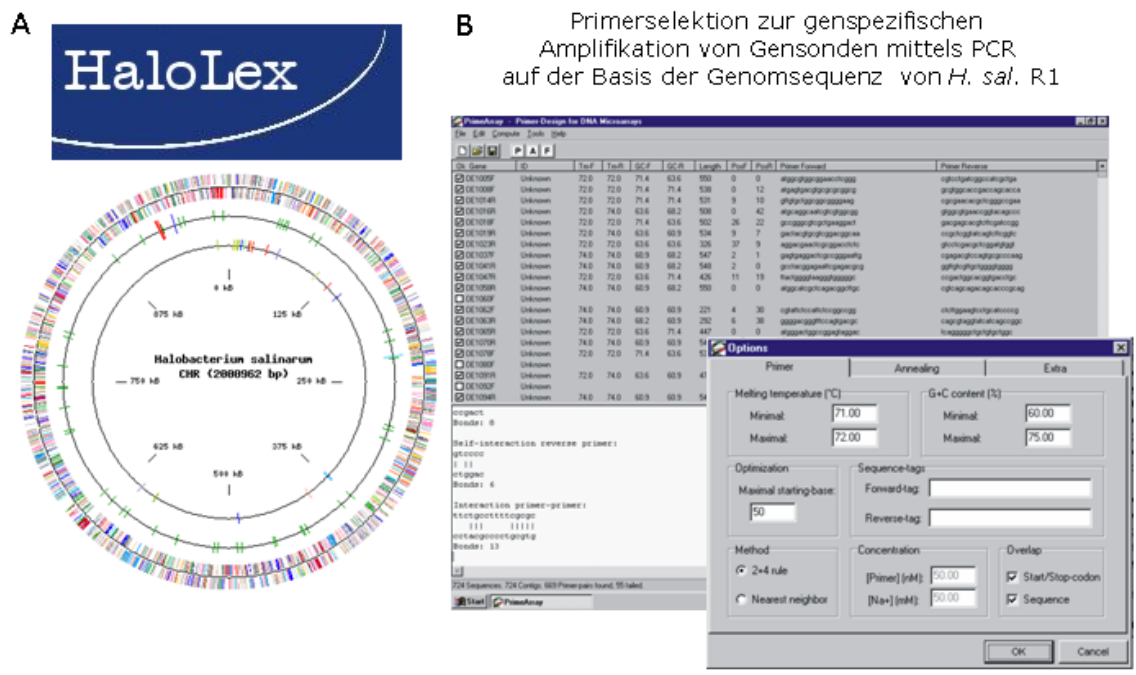


Abb. 12: Abbildung zum Sondendesign. A) Schematische Darstellung der Genomkarte von *H. sal.* R1 (<http://www.halolex.mpg.de>). B) Programmoberfläche (Screenshot) des Computerprogramms PrimeArray zur Primerselktion; mit Einblendung des Optionsfensters, zur Einstellung der Parameter für die Primeroptimierung.

Eine Abdeckung von über 97% entspricht einer Summe von 2709 genspezifischen Sonden für den Gesamtgenom-Mikroarray von *H. sal.* R1, wenn man eine Gesamtzahl von 2784 Genen für das Genom zugrunde legt.

Die erzielte Abdeckung und die Effizienz der PCR-Amplifikation belegen eine optimale Primerselktion und eine hohe Ausbeute des PCR-Produktes, wodurch nur eine geringe Rate von Nachamplifikationen einzelner Sonden notwendig wurde.

### 3.1.1.2 Qualitätskontrolle und Reinigung der PCR-Produkte

Zur Kontrolle der Effizienz der PCR-Reaktionen wurden sämtliche Ansätze mittels einer Agarosegelelektrophorese überprüft. Anhand dieser Ergebnisse konnte gezeigt werden, dass unter einheitlichen Bedingungen fast alle PCR-Produkte amplifiziert werden und zum Großteil nur ein spezifisches PCR-Produkt entstanden ist.

Die Agarosegelelektrophorese kann die Einheitlichkeit der PCR-Amplifikation nachweisen, jedoch zur Kontrolle der Eindeutigkeit bzw. der Spezifität des entstanden PCR-Produktes ist eine Sequenzierung notwendig. Deshalb wurden stichprobenartig Sequenzierungen der amplifizierten Genfragmente durchgeführt, wodurch bei allen untersuchten PCR-Produkten die Spezifität per BLAST-Suche (**B**asis **L**ogical **A**lignment **S**earch **T**ool) gegen die *H. sal.* R1 Genomsequenz bestätigt werden konnte.

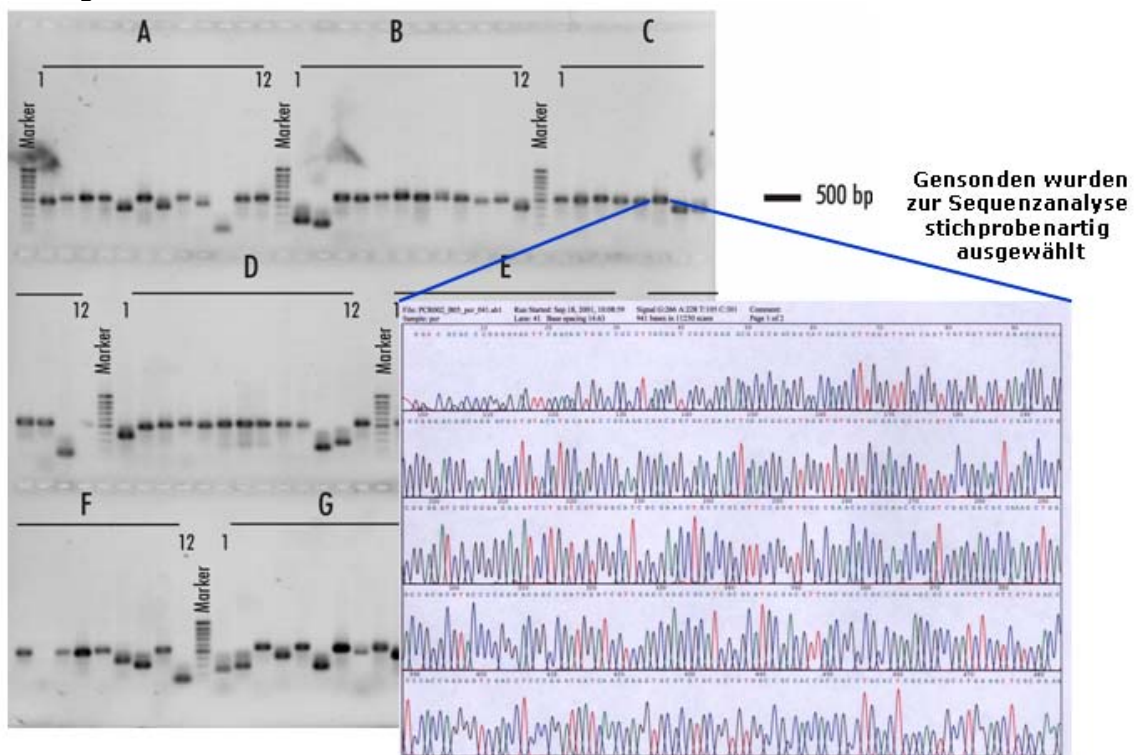


Abb. 13: Ergebnis einer Agarosegelelektrophorese von PCR-Produkten zur Herstellung des Gesamt-Genom-Mikroarrays und einer stichprobenartigen Sequenzanalyse einer Gensonde. Die durchschnittliche Bandengröße der PCR-Produkte beträgt 500 bp. Die Buchstaben und Zahlen, oberhalb der Banden,

ermöglichen eine eindeutige Zuordnung der Banden zu den Positionen auf der 96-iger PCR-Platte (Ergebnis der BLAST-Suche ist nicht gezeigt).

Vor dem Aufbringen der PCR-Produkte/Gensonden auf den Mikroarray wurden sämtliche PCR-Produkte mittels einer vakuumgetriebenen Filtration durch eine Größenausschlussmembran gereinigt. Dieser Arbeitsschritt der PCR-Produkt-Reinigung dient dazu, Bestandteile der PCR-Reaktion, wie Enzyme, Primer und Salze zu entfernen. Dadurch können Verstopfungen der Drucknadeln, Kreuzkontaminationen und Beeinflussung der Hybridisierungseffizienz verhindert bzw. vorgebeugt werden.

### 3.1.2 Design der DNA-Mikroarrays

Bei der Herstellung von DNA-Mikroarrays, werden die Gensonden, mit Hilfe von Drucknadeln („pins“) auf die Träger („slides“) appliziert. Diese Drucknadeln befinden sich bei einem „Arrayer“ (Mikroarray-Druckmaschine) in einer Halterung, die als Druckkopf bezeichnet wird und in der folgenden Abbildung zu sehen ist. (Die Begriffe DNA-Mikroarray, DNA-Chip und Slide werden als Synonyme verwandt.) Alle Mikroarray-Chips, die für Transkriptom-Experimente eingesetzt wurden, sind mit einem Portalroboter der Firma Virtek hergestellt worden, der mit einem Druckkopf und 48 Drucknadeln ausgestattet ist.

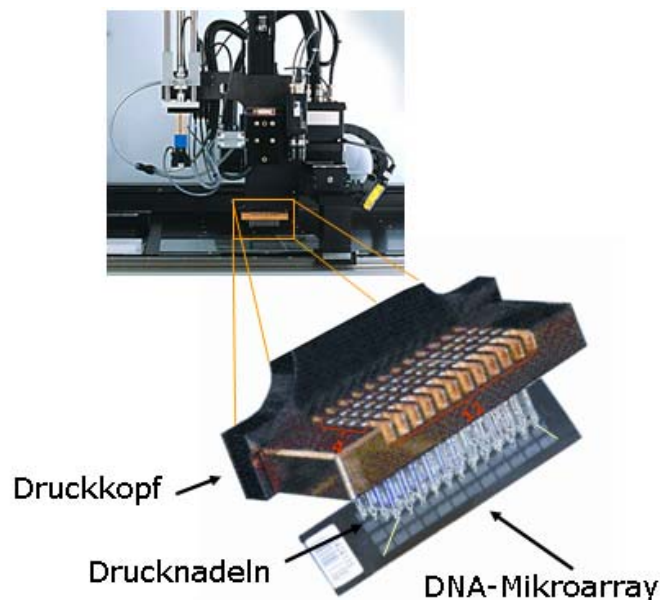


Abb. 14: Vergrößerte Darstellung eines Druckkopfes von einem Virtek ChipWriterPro mit 4 x 12 (48) Drucknadeln („pins“).

Entsprechend der Anordnung der Nadeln in dieser Halterung, werden auf dem zu bedruckenden DNA-Mikroarray sogenannte Blöcke erzeugt. Dies sind die Flächen

auf dem DNA-Mikroarray, auf denen die Gensonden gedruckt und verankert werden.

Die Anordnung der Blöcke auf den DNA-Chips, spiegelt die Geometrie der Besetzung der Drucknadeln in dem Druckkopf wider. Der in Abbildung 14 gezeigte Druckkopf des Arrayers ChipWriterPro ist mit 48 Nadeln ausgestattet und erzeugt auf dem Objektträger (Slide) entsprechend eine Summe von 48 Blöcken, in einem Muster von 4 x 12.

Beim Design der Mikroarrays muss neben der Anzahl der Nadeln auch berücksichtigt werden, wie viele Gensonden auf den Chip gedruckt werden sollen und wie groß der maximale Abstand zwischen den Punkten („spots“) gewählt werden kann. Diese Spots auf dem DNA-Chip entsprechen somit kleinen Aliquots der DNA-Lösung der jeweiligen genspezifischen Sonde und sind nach einer Hybridisierung als farbige Punkte des ausgelesenen Bildes zu sehen.

Für den Fall der Herstellung des Gesamtgenom-Mikroarrays von *H. sal.* R1 standen 2709 Gensonden zur Verfügung. Auf Grund dieser Anzahl von Sonden und unter Verwendung des in (Abb. 15) gezeigten Designs des Gesamtgenom-Mikroarrays von *H. sal.* R1 war es zu realisieren, dass fünf spots pro Sonde auf den DNA-Mikroarray gedruckt werden konnten. Daraus ergab sich, dass ein Block mit 320 spots (16 x 20) 64 Gene repräsentiert. Der Gesamtgenom-Mikroarray-Chip von *H. sal.* R1 mit diesem Design, weist eine Gesamtzahl von 15360 spots auf und ist damit der bisher am dichtesten bedruckte DNA-Chip eines Archaeons.



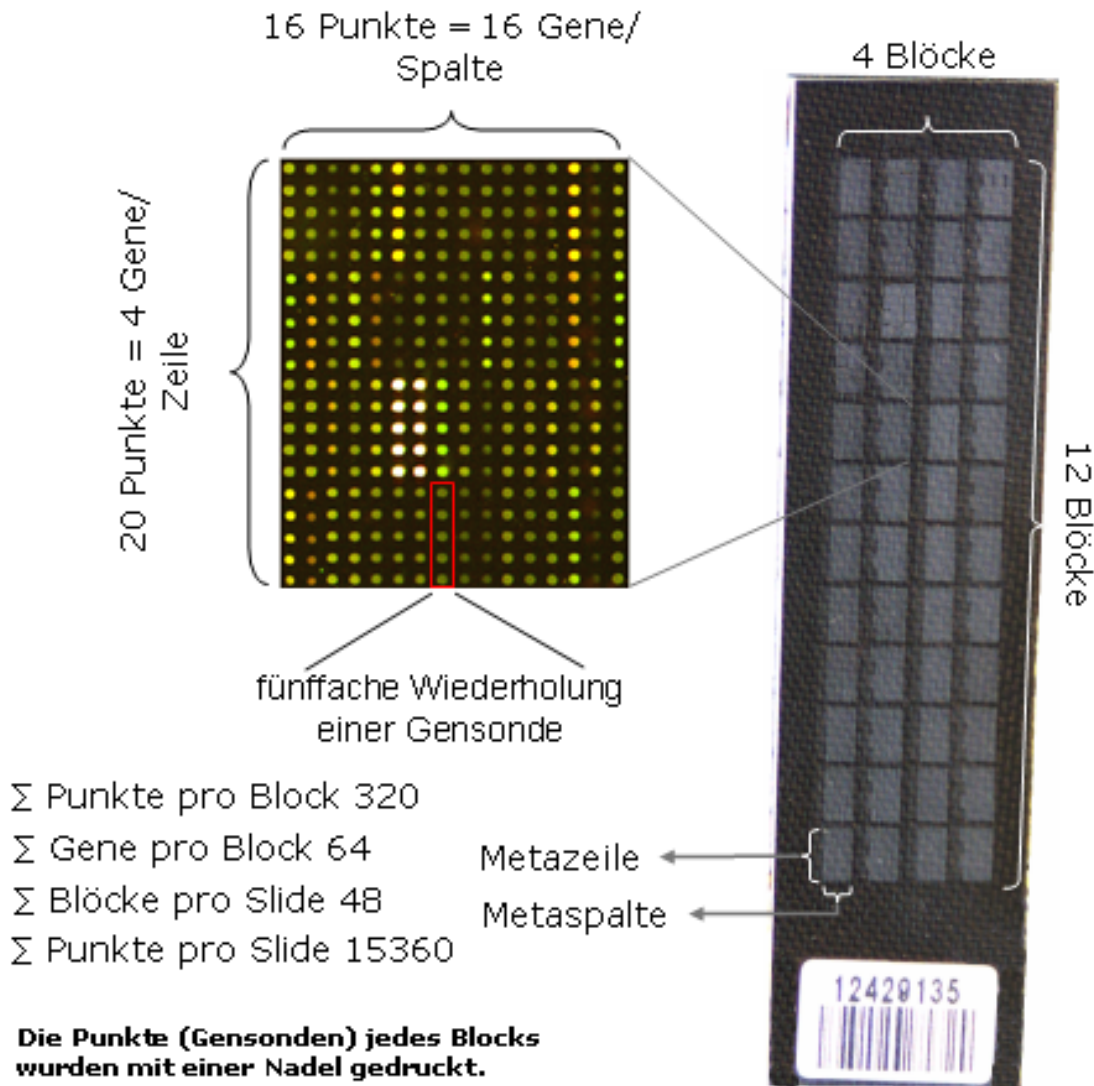


Abb. 15: Darstellung des Layouts des Gesamtgenom-Mikroarray von *H. sal.* R1. Jede Gensonde (Punkt, „spot“) wurde in fünffacher Wiederholung auf den Objektträger gedruckt. Durch die Angaben des Blocks, der Metazelle, Metaspalte, Zeile und Reihe, ist die Position einer Gensonde auf dem DNA-Mikroarray eindeutig festgelegt.

### 3.1.3 Produktion der Mikroarrays

Die Gesamtgenom-Mikroarrays von *H. sal.* R1 konnten in gleich bleibender Qualität und großer Stückzahl nach dem oben beschriebenen Layout, produziert werden. Die so hergestellten Mikroarrays wurden nicht nur für dieses Projekt verwandt, sondern auch in weiteren Transkriptom-Projekten eingesetzt, wie z.B. in der Promotion von Herrn Andy Wende und Herrn Christoph Schwarz.

Die präzise und zuverlässige Produktion von Mikroarrays ist abhängig von verschiedenen Faktoren. Eine Schlüsselfunktion besitzen die Drucknadeln, denn ungleichmäßige oder beschädigte Nadeln transportieren während des

Druckvorganges unterschiedliche Volumina an DNA-Lösung und beeinflussen somit die spätere akkurate Bildanalyse der Mikroarrays. Genauso hat die Qualität der Objektträger selbst einen entscheidenden Einfluss auf das Ergebnis und die Aussagekraft der Mikroarray-Experimente.

Zur Immobilisierung von DNA-Sonden bieten Glas-Objektträger hervorragende Eigenschaften. Sie weisen eine geringe spezifische Eigenfluoreszenz auf, besitzen eine hohe optische Transparenz und besitzen einfache chemische Modifikationsmöglichkeiten ihrer Oberfläche. Für die Produktion von *H. sal.* R1 Mikroarrays wurden konventionelle Mikroskop- und speziell gefertigte Objektträger verwandt. Die Oberflächen dieser Träger können mit unterschiedlichen Arten von Beschichtungen versehen sein, die fest mit dem Glasuntergrund verbunden sind und gleichzeitig auch eine optimale Verankerung der DNA auf der Oberfläche gewährleisten.

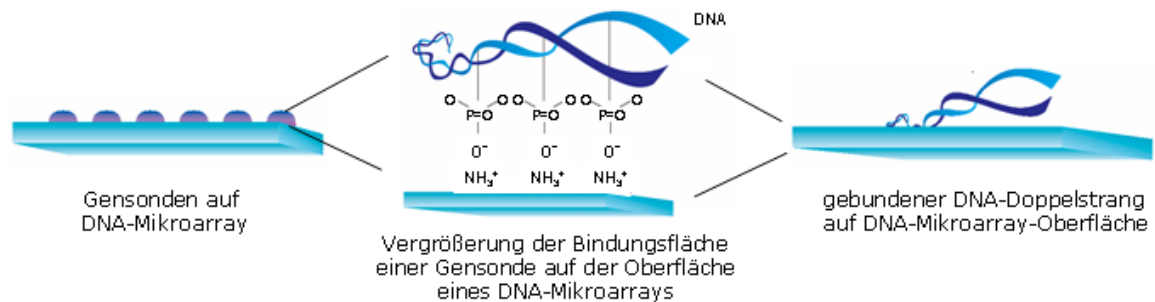


Abb. 16: Schematische Darstellung der Bindung des Phosphatrückgrates einer Gensonde an die beschichtete Oberfläche eines DNA-Mikroarray-Slides. Als Beispiel für eine Beschichtung ist hier die Darstellung von Amin-Derivaten gewählt worden (nach Amersham Biosciences, Handbuch).

Zur Produktion der *H. sal.* R1 Mikroarrays wurden zwei unterschiedliche Beschichtungen verwandt, wie aus der folgenden Tabelle zu entnehmen ist.

Mikroarray-Objektträger	Beschichtung	Bindung der DNA
Poly-Prep™ Objektträger	Poly-L-Lysin (adsorptiv)	elektrostatische Wechselwirkungen; in kovalente Bindung zu überführen durch Hitze- oder UV-Behandlung
CMT-GAPS-II Objektträger	γ-Aminopropylsilan (kovalent)	elektrostatische Wechselwirkungen; in kovalente Bindung zu überführen durch UV-Behandlung

Tab.3: Zusammenfassung der verwendeten Mikroarray-Objektträgern und deren Oberflächenbeschichtung.



Die Unterschiede zwischen den beiden Beschichtungen liegen darin, dass eine Poly-L-Lysin Beschichtung nicht länger als vier Monate haltbar ist, während des Druckvorganges beschädigt werden kann und dazu neigt, bei höheren Temperaturen, wie z.B. bei dem Prozess der Denaturierung der Mikroarrays, sich vom Glasuntergrund abzulösen. All diese Punkte können somit die Qualität der Mikroarrays und damit auch die der Hybridisierungsergebnisse negativ beeinflussen (Bowtell D und Sambrook J, 2003).

### 3.1.3.1 Produktion der Mikroarrays auf verschiedenen Trägern

Im Rahmen dieser Arbeit wurden in Zusammenarbeit mit zwei Gruppen verschiedene *H. sal.* R1 DNA-Mikroarrays unter Benutzung unterschiedlicher Roboter und Qualitäten an Objektträgern hergestellt.

Kooperation	Roboter	Objektträger
Universität Frankfurt (Arbeitsgruppe J. Soppa)	<i>MicroGrid II</i> (Biorobotics)	Poly-L-Lysin Objektträger (Sigma)
MPI für Psychiatrie (M. Panhuysen)	<i>ChipWriterPro</i> (Virtek)	CMT-GAPS II Objekttrüge (Corning)

Tab. 4: Tabelle mit Angaben über die Kooperation, die verwendeten Roboter und die eingesetzten Objektträger.

#### 3.1.3.1.1 DNA-Mikroarrays auf Poly-L-Lysin Slides

Zur Herstellung der Mikroarrays im Rahmen der Kooperation während der Doktorarbeit von Herrn A. Zaigler in der Arbeitsgruppe von Herrn J. Soppa (Universität Frankfurt), wurden Poly-L-Lysin beschichtete Objektträger der Firma Sigma und der Roboter *MicroGrid II* von BioRobotics eingesetzt.

Hierbei konnte in Vorversuchen gezeigt werden, dass der sehr häufig verwendete 3x SSC-Puffer (Eisen und Brown), bei Verwendung von Poly-L-Lysin beschichteten Objektträgern, keine ausreichende Homogenität der gedruckten Sonden ergab. Diese konnte jedoch durch Verwendung eines 150 mM Phosphatpuffers mit 0,01% SDS erreicht werden weshalb die PCR-Produkte zur Mikroarray-Herstellung auch für dieses Projekt alle in dem SDS-Phosphatpuffer resuspendiert und gelagert wurden (Dr-Arbeit A. Zaigler).

Zum Drucken der Arrays wurde der Druckkopf des *MicroGrid II* im Fall des Test-Mikroarrays mit 4 Nadeln versehen und für den Gesamt-Genom-Mikroarray mit 8 Nadeln ausgestattet. Durch Drucken von Spotting-Puffer auf Test-Slides, wurde vor jeder Produktion die einwandfreie Funktionsfähigkeit der eingesetzten Nadeln

überprüft. Zur Produktion von 110 Mikroarrays mit 2880 Sonden, unter Verwendung von 8 Nadeln, benötigte der Roboter eine Laufzeit von ca. 2,5 Tagen. Die entstandenen Test- und Chromosom-Mikroarrays werden in Abschnitt 3.1.3.2 näher beschrieben.

### **3.1.3.1.2 DNA-Mikroarrays auf CMT-GAPS II Slides**

Im Rahmen der Kooperation mit Herrn M. Panhuysen wurde für die Mikroarrayproduktion der Roboter *ChipWriterPro* von Virtek benutzt. Gedruckt wurden die Mikroarrays auf CMT-GAPS II Slides unter Verwendung von 48 Nadeln.

Routinemäßig wurde das gesamte System des Roboters an Hand einer am MPI für Psychiatrie erarbeiteten Checkliste vor jeder Arrayproduktion auf seine einwandfreie Funktionsfähigkeit hin überprüft. Durch die Konfiguration dieses Roboters war es erstmals möglich, 5 Replikate jeder genspezifischen Sonde auf einen Mikroarray zu drucken. So sind Arrays entstanden, die es auf Grund dieser Anzahl an Wiederholungen erlauben, die gemessenen Signalintensitäten jeder Sonde zu mitteln, wodurch ein statistisch abgesichertes Ergebnis erhalten wird. Mit den Standarddruckbedingungen benötigte der Roboter zur Produktion von 50 Slides mit 15360 Spots eine Laufzeit von 7 Stunden.

Die eingesetzten CMT-GAPS II Slides besitzen eine  $\gamma$ -Aminopropylsilan Beschichtung. Für diese Beschichtung erwies sich ein 3x SSC-Puffer mit 1,5 M Betain als optimal (Diehl F, et. al., 2001). Neben dem positiven Einfluss des Betains auf die Homogenität der gedruckten Sonden, erhöht es die Viskosität der DNA-Lösung, wodurch die Verdunstung während der Herstellung der Mikroarrays verringert werden kann.

Da am Max-Planck-Institut für Psychiatrie mit diesem Puffer ebenfalls sehr gute Resultate bei der Mikroarray-Produktion erzielt wurden, sind die 2784 Sonden in einer zweiten Runde erneut mittels PCR amplifiziert, gereinigt und in dem Betain haltigen 3x SSC-Puffer resuspendiert worden.

Beispiele für die Realisierung der gedruckten Mikroarrays von *H. sal.* R1 werden im folgenden Abschnitt 3.1.3.2 beschrieben.

### 3.1.3.2 Beispiele verschiedener Mikroarrays von *H. sal.* R1

Wie unter 3.1 bereits erwähnt, wurde je nach Status der Genomannotation von *H. sal.* R1 das Sondendesign durchgeführt. Dadurch ergeben sich die Tabelle 5 gezeigten drei verschiedenen Gruppen von Arrays unterschiedlicher GröÙer:

Mikroarray-Art	Anzahl der genspez. Sonden	Gesamtzahl der Sonden auf Array
Test-Mikroarray	207	621 (dreifach)
Chromosom-Mikroarray	2197	4394 (zweifach)
Gesamt-Genom-Mikroarray	2709	13545 (fünffach)

Tab. 5: Tabelle mit verschiedenen Mikroarray-Arten von *H. sal.* R1.

Wie der Tabelle zu entnehmen ist, wurden die Gensonden für die verschiedenen *H. sal.* R1 Mikroarrays in zwei- bis fünffacher Wiederholung auf die Träger gedruckt. Da während des Produktionsvorganges eines DNA-Mikroarrays mittels Roboters der Fall eintreten kann, dass Gensonden nicht auf dem Träger appliziert werden, ist das Drucken zumindest eines Duplikats jeder Sonde zu empfehlen.

Abgesehen von der Tatsache dass man durch mehrfaches Aufbringen einer Sonde einen möglichen Informationsverlust minimiert, erlaubt das mehrmalige Drucken beim späteren Prozess der Datenauswertung einen Mittelwert der Fluoreszenzsignale zu bilden und damit auftretende Schwankungen zu beurteilen.

Für *H. sal.* R1 wurden so Transkriptom-Daten gesammelt, die durch die Kombination des technischen Replikats der Sonde auf dem DNA-Chip und des biologischen Replikats der wiederholten Hybridisierungen, es aufgrund ihrer Vielzahl an Datenpunkten erlaubten, eine statistisch abgesicherte Auswertung der Hybridisierungs-Experimente durchzuführen.

Bilder der in Tabelle 5 aufgeführten DNA-Mikroarrays sind in der folgenden Abbildung zu sehen.

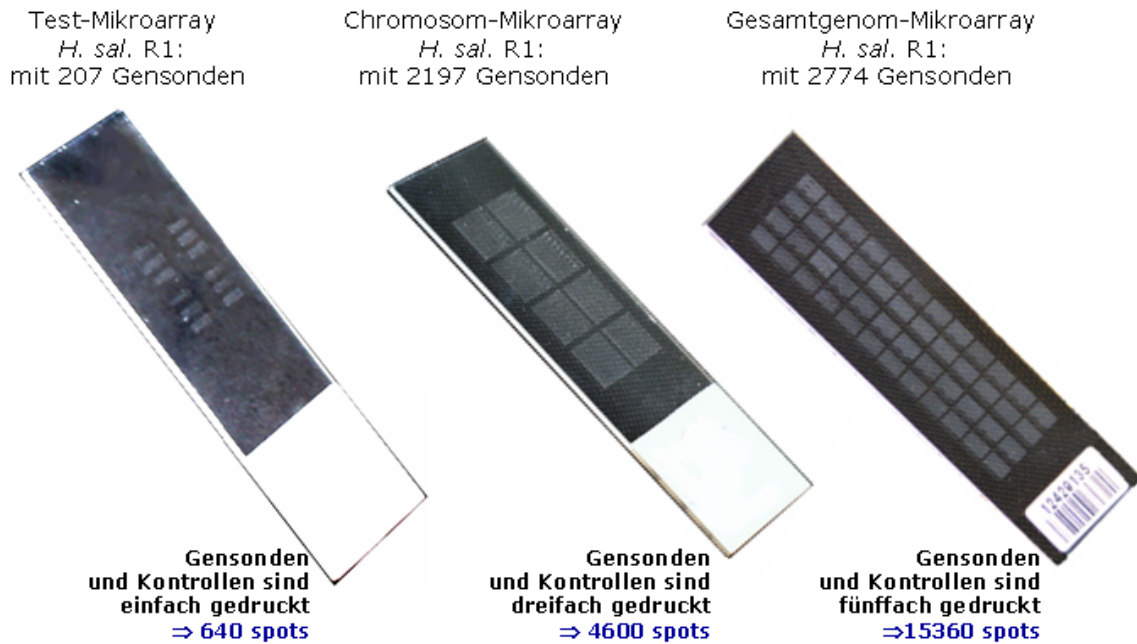


Abb. 17: Bilder der verschiedenen Realisierungen der DNA-Mikroarrays für *H. sal.* R1, mit den Angaben zur Anzahl der Gensonden und der Gesamtanzahl der applizierten Sonden auf den Objektträgern.

Anhand von zwei Beispielhybridisierungen soll in diesem Abschnitt aufgezeigt werden, wie unterschiedlich die Ergebnisse einer DNA-Mikroarray Produktion und damit auch einer Mikroarray-Hybridisierung ausfallen können und wie wichtig es ist, ein optimal aufeinander abgestimmtes System zu haben. Denn nur so kann ein optimales Signal/Rausch-Verhältnis erzielt werden.

Als Beispiele werden sogenannte Verhältnisbilder von einer Hybridisierung eines Chromosom-Mikroarrays, gedruckt auf einem Poly-L-Lysin-Slide und eines Gesamtgenom-Mikroarrays, gedruckt auf einem CMT-GAPS II Slide, gezeigt. Bei einem Verhältnisbild handelt sich um ein computergeneriertes Bild des Mikroarrayscanners. Von der Bildsoftware des Scanners werden die Intensitätssignale der beiden Fluoreszenzfarbstoffe kombiniert und in einem Bild zusammen dargestellt. Zur Unterscheidung der Intensitätssignale im Bild werden sie durch die Farben rot und grün repräsentiert. Weist eine Sonde die gleiche Intensität für die beiden Cy-Farbstoffe auf, werden diese Gensonden durch einen gelben Punkt dargestellt. Nähere Informationen zum Prozess der Bildakquisition ist in Abschnitt 3.4 zu finden.

Verhältnissbilder nach der Detektion des Fluoreszenzsignals

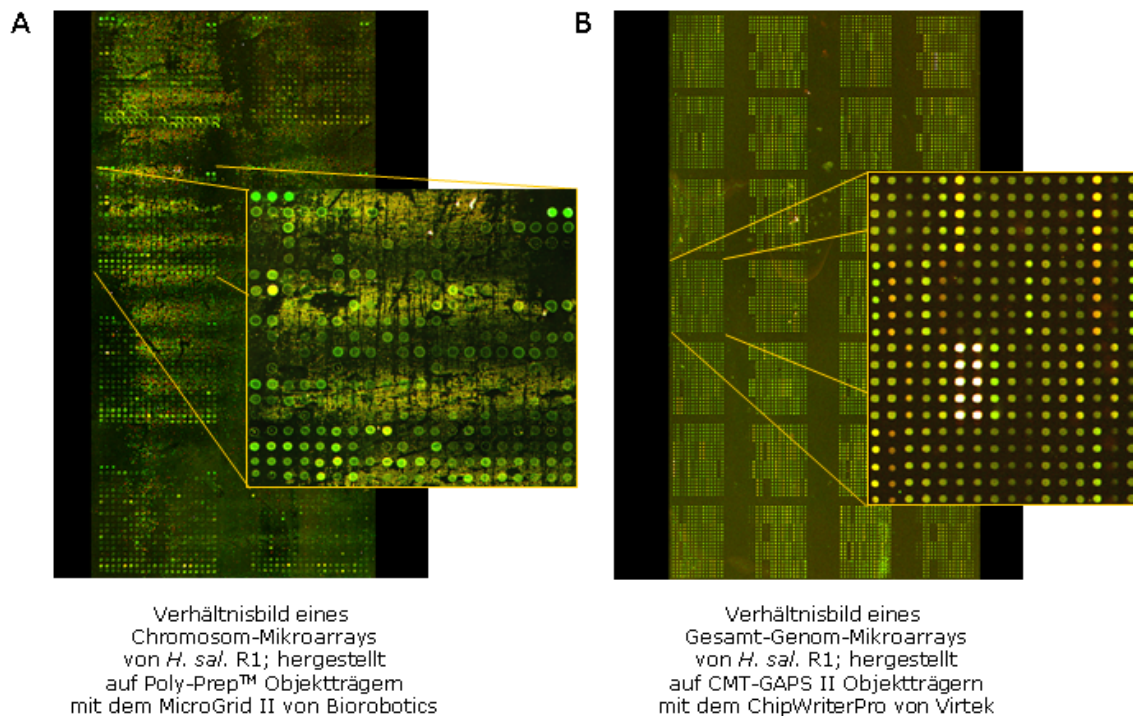


Abb. 18: Verhältnissbilder zweier *H. sal.* R1 DNA-Mikroarrays. A) Bild nach einer Hybridisierung eines Chromosom-Mikroarrays von *H. sal.* R1 gedruckt auf Poly-Prep™-Slides. Die Vergrößerung eines Blocks zeigt deutlich, dass das Signal/Rausch-Verhältnis, auf Grund eines sehr hohen unspezifischen Hintergrundsignals, nicht gut ist. B) Bild eines Gesamt-Genom-Mikroarrays von *H. sal.* R1 gedruckt auf CMT-GAPS II-Slides. An Hand des vergrößerten Ausschnitts, ist eindeutig ein optimales Signal/Rausch-Verhältnis zu sehen.

Die Abbildung 18 zeigt zwei Beispiele von hybridisierten und ausgelesenen *H. sal.* R1 Mikroarrays. Bei der Verwendung von Poly-Prep™ Slides zeigte sich, trotz einer Blockierung mit BSA, ein sehr hohes unspezifisches Hintergrundsignal. An einigen Stellen der Mikroarrays war der Hintergrundwert so hoch, dass keine Unterscheidung zwischen den Signalen der Gensonden und des Hintergrundes möglich und somit die automatische Bildanalyse nicht anzuwenden war. Die Bildanalysesoftware würde eine manuelle Bearbeitung des Prozesses der Punktfindung („spot finding“) ermöglichen, doch durch einen solchen Schritt wären die Ergebnisse der Mikroarray-Experimente nicht mehr zu reproduzieren, da je nach Anwender die Größe und Qualität der Spots willkürlich beurteilt würde. Deshalb ist von einer Nachbearbeitung der automatischen „spot finding“ im großen Umfang abzuraten. Das Problem des starken Hintergrundes führte zu einem erheblichen Verlust an Ergebnissen. Da dieser Effekt auf allen der hybridisierten Chromosom-Mikroarrays von *H. sal.* R1 aufgetreten ist, waren die Ergebnisse dieser Hybridisierungen nicht zu verwenden. Aus diesem Grund

wurden für alle weiteren Hybridisierungen nur noch CMT-GAPS-II-Slides von der Firma Corning eingesetzt. Diese Slides werden speziell für die Herstellung von DNA-Mikroarrays produziert und besitzen einen sehr hohen Qualitätsstandard. Wie in (Abb. 18 B) zu sehen ist, sind keine unspezifischen Bindungen auf den nicht bedruckten Flächen zu beobachten, was auf die hervorragende Qualität dieser Slides zurückzuführen ist. Des Weiteren ist durch dieses optimale Signal/Hintergrund-Verhältnis der automatische Prozess der Punktfindung ohne Probleme durchführbar.

Die gezeigten erheblichen Qualitätsunterschiede der Hybridisierungen, sind allein auf die Qualität der verwendeten Slides zurückzuführen und nicht durch die zwei verschiedenen „spotting“-Roboter verursacht worden, die zur Produktion der Slides eingesetzt wurden (siehe 3.1.3.1).

Alle für dieses Projekt relevanten Hybridisierungsergebnisse beziehen sich auf den Einsatz der Gesamt-Genom *H. sal.* R1 DNA-Mikroarrays, die auf CMT-GAPS II-Slides gedruckt wurden.

### **3.1.4 Diskussion**

Für dieses Projekt wurde zur Realisierung der DNA-Mikroarrays die Variante gewählt genspezifische PCR-Fragmente auf Glasobjektträger aufzubringen. Wie bei allen anderen möglichen DNA-Mikroarray-Varianten, stellt die Selektion der zu druckenden Genfragmente einer der ersten Herausforderungen dar.

Auch bei einer kleinen Gesamtgenomgröße von 2 Mb für *H. sal.* R1, im Vergleich z.B. zu eukaryontischen Genomen, ist eine bioinformatisch gestützte Primerselektion eine wichtige Voraussetzung. Denn so kann gewährleistet werden, dass nicht nur die gewählten Primer einer Optimierung unterliegen, sondern auch die Qualität der Hybridisierung. Da bereits *in silico* durch Sequenzvergleich der generierten Amplifikate mögliche Kreuzhybridisierungen bereits am Computer stark minimiert werden. Außerdem können ebenfalls durch das automatische Einlesen der Gensequenzen Fehler bei der Zuordnung der Primerpaare zu den entsprechenden Gensequenzen ausgeschlossen werden. Ein weiterer Vorteil einer computergestützten Primerauswahl ist die vollständige Dokumentation dieses Prozesses, was in der heutigen Zeit eine immer wichtiger werdende Notwendigkeit darstellt.

Die Wahl des Computerprogramms PrimeArray zur Primerselektion, zeigte sich als eine sehr gute Lösungsvariante. Wie unter 3.1.1.1 beschrieben, konnte durch

den Einsatz dieses Programms der Grundstein für die Hochdurchsatzproduktion der genspezifischen PCR-Fragmente gelegt werden. Ein gewisser Grad an Automation erlaubte es, die benötigten Genfragmente in einem vertretbaren Zeitrahmen herzustellen.

Neben einer schnellen und verlässlichen Primerselektion, war auch die Etablierung eines effizienten Standardprotokolls zur PCR-Amplifizierung, Aufreinigung und Kontrolle, eine wesentliche Voraussetzung zum Erreichen einer Abdeckung von 97,30% des Genoms von *H. sal.* R1. So konnte sichergestellt werden, dass die spezifischen Genfragmente immer in gleich bleibender Qualität produziert werden können. Dieser Punkt wird dadurch bestätigt, dass die im Rahmen dieser Arbeit erstellten genspezifischen Sonden und entwickelten Arbeitsabläufe der Transkriptom-Analyse auch in anderen Projekten zum Einsatz kommen und gekommen sind.

Bei der Primerselektion wurde auf die Einführung einer Modifikation der Oligonukleotide verzichtet, weil für dieses Projekt ausschließlich Glasobjektträger eingesetzt wurden, die eine aminomodifizierte Oberfläche besitzen und die Bindung von DNA-Fragmenten an solche Oberflächen hauptsächlich durch Wechselwirkung des negativ geladenen Phosphatrückgrats der DNA-Stränge erfolgt (Southern E, 1999).

Die Qualität von DNA-Mikroarrays hängt neben den zu druckenden Gensonden auch entscheidend ab von der Qualität des Roboters, den eingesetzten Drucknadeln und den verwendeten Glasobjektträgern.

Unterschiede zwischen den verfügbaren Mikroarray-Robotern bestehen z.B. im Design des Druckkopfes sowie der Geschwindigkeit und weniger in der Präzision ihres Dreiachsen mechanischen Roboterarms.

Wie ab (3.1.3) beschrieben wurden die verwendeten DNA-Mikroarrays in Kooperation mit verschiedenen Arbeitsgruppen hergestellt, wodurch die Optimierung des Druckvorgangs für dieses Projekt entfiel und auf die jeweils etablierten Protokolle und Erfahrungen zurückgegriffen werden konnte.

Im Rahmen der Zusammenarbeit mit der Universität Frankfurt sind die ersten DNA-Mikroarrays mit Genfragmenten von *H. sal.* R1 entstanden. Darunter waren die 200 Gensonden umfassenden so genannten Test-Mikroarrays (siehe Abb. 17).

Diese genspezifischen DNA-Mikroarrays wurden eingesetzt um die Funktionalität der Arrays zu überprüfen. Der Beweis dafür wurde durch die Charakterisierung der Transkriptionsänderung nach dem Wechsel von aerobem zu anaerobem

Wachstum mit Arginin als Energiequelle erbracht (Doktorarbeit A. Zaigler). Denn durch ein Zeitreihenexperiment konnte gezeigt werden, dass die drei Strukturgene des Arginin-Deiminase-Weges (die Arginin-Deiminase (ADI), die katabole Ornithin-Transcarbamoylase (cOTCase) und die Carbamat-Kinase (CK)) deutlich induziert werden. Dabei wurden Proben nach 15 und 90 Minuten nach Setzen des Stimulus mit aeroben Proben vor dem Wechsel mit einander verglichen. Des Weiteren konnte gezeigt werden, dass Gene des Pyruvat-Dehydrogenasekomplexes und des Citratzykluses reprimiert werden. Damit stehen diese Ergebnisse im Einklang mit früheren Experimenten (Soppa J, et al., 1998) und legen den Grundstein zur Etablierung der Mikroarray-Technologie für *H. sal.* R1.

Während der Optimierungsphase des Projektes bildete sich immer deutlicher heraus, wie wichtig die Abstimmung nicht nur der Arbeitsmethoden, sondern auch der einzelnen eingesetzten Komponenten ist. Wie in Abschnitt (3.1.3.2) beschrieben und in Abb. 18 dargestellt, wiesen die verwendeten Poly-L-Lysin Slides einen hohen unspezifischen Hintergrund auf, was zu einem ungünstigen Signal/Rausch-Verhältnis führte und eine automatische Bildanalyse der ausgelesenen Fluoreszenzsignale der Arrays verhinderte.

Dieses Ergebnis führte dazu, dass für die weiteren DNA-Mikroarray-Produktionen die CMT-GAPS II Slides (Corning) verwendet wurden, welche speziell für den Einsatz in der Mikroarray-Technologie konzipiert sind. Bei dieser Umstellung konnte auf die Expertise und Ressourcen für der Herstellung von hochdichten DNA-Mikroarrays am Max-Planck-Institut für Psychiatrie zurückgegriffen werden.

Im Laufe dieser Kooperation ist auch das endgültige Layout des Gesamtgenom-Mikroarrays von *H. sal.* R1 entwickelt worden (siehe Abb. 15). Die Vorgabe der Konfiguration des Druckkopfes von 12x4 Drucknadeln des ChipWriter Pro (Virtek) und die Aufgabe 2709 Gensonden auf einen Array zu applizieren, erlaubte es jedes Gen von *H. sal.* R1 in fünffacher Wiederholung auf die Slides zu drucken. Damit weist der Gesamtgenom-Mikroarray von *H. sal.* R1 eine Summe von 15360 Spots auf und repräsentiert den zurzeit am dichtesten bedruckten archaealen DNA-Mikroarray (siehe 3.1.2). Der Vorteil einer fünffachen Wiederholung besteht darin, dass zum einen das Signal eines Gens aus fünf gemessenen Werten gemittelt werden kann und zum anderen wird dadurch die Wahrscheinlichkeit erhöht, überhaupt ein Signal für das Gen zu bekommen. Jedoch ist zu beachten, dass eine zu hohe Anzahl an Wiederholungen die Gefahr beinhaltet, eine Art Verdünnung des Signals zu verursachen.



Eine zukünftige Veränderung beim Design des Gesamtgenom-DNA-Mikroarrays von *H. sal.* R1, könnte neben der zufälligen Anordnung der Gene auch eine zufällige Verteilung der fünf Wiederholungen der Gensonden auf dem Slide darstellen. So könnte durch die Mittelung der gemessenen Intensitätssignale noch besser auf mögliche räumliche Effekte bei der Hybridisierung eingegangen werden. Diese Phänomene bei der Mikroarrayhybridisierung werden in den Abschnitten der Diskussion zur Datenauswertung genauer beschrieben.

### 3.2 Durchführung der Experimente

Wie in der Einleitung dieses Kapitels erwähnt, folgt die Gliederung des Flussdiagramms aus Abb. 11. Im weiteren wird die Durchführung und das Design der Experimente näher beschreiben.

#### 3.2.1 Experimentdesign

Das den Zweifarben-Mikroarrays zu Grunde liegende Prinzip ist die kompetitive Hybridisierung unterschiedlich fluoreszenzmarkierter cDNAs. Die wichtigsten Fragen beim Design eines DNA-Mikroarray-Experiments sind: 1. Welche RNA-Probe mit welchem Fluoreszenzfarbstoff markiert und 2. Welche Kombination von markierten cDNAs zusammen auf einem Mikroarray-Chip hybridisiert werden soll (Churchill GA, et al., 2002; Yang H und Speed T, 2002)? Dabei gibt es die Möglichkeit des direkten und indirekten Vergleichs.

In der Literatur ist zurzeit keine eindeutige Antwort auf die Frage des „richtigen“ Experimentdesigns zu finden. Ein optimales („richtiges“) Mikroarray-Experimentdesign berücksichtigt, dass Farbstoff bedingte Unterschiede mathematisch zu korrigieren sind und alle interessierenden Vergleiche mit einem Minimum an Mikroarrays und einer möglichst geringen Fehlerrate durchgeführt werden können. Mit dem Farbstoffeffekt ist gemeint, dass die Fluoreszenzfarbstoffe Cy3 und Cy5 bei einer direkten Markierung der cDNA vom Enzym mit unterschiedlicher Effizienz in den Einzelstrang eingebaut werden. Diese Unterschiede gilt es, durch ein optimales Experimentdesign zu minimieren.

Für den Hauptanteil der vorliegenden Arbeit wurde als Experimentdesign der Transkriptom-Experimente, das sogenannte einheitliche Referenzdesign („*common reference design*“) gewählt, welches in der Abbildung 19 C) dargestellt wird.

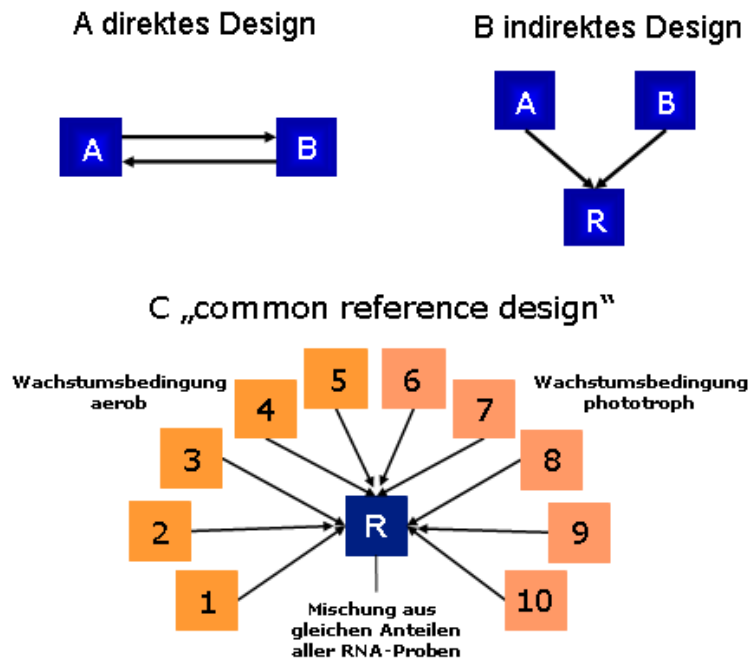


Abb. 19: Schematische Darstellung des direkten und indirekten Transkriptom-Experimentdesigns. In dieser Darstellung repräsentieren die Kästchen die RNA-Proben und die Pfeile die Hybridisierungen. Es gibt zwei mögliche Realisierungen eines Experimentdesigns zum Vergleich zweier verschiedener Proben A und B. A) Der Konvention zu folge, ist die Probe am Pfeilende mit dem grünen Farbstoff (Cy3) und die Probe an der Pfeilspitze mit dem roten Farbstoff (Cy5) markiert worden. Beim direkten Vergleich des Expressionsniveaus zweier Proben (A u. B), werden die unterschiedlich markierten Proben gemischt und auf einem DNA-Mikroarray hybridisiert. Dieses Beispiel zeigt ein so genanntes Farbstoffwechsel- („dye swap“)-Design, bei dem jede Farbstoffkombination auf einem DNA-Mikroarray hybridisiert wird. B) Beim indirekten Vergleich der Proben (A u. B), werden die Expressionsunterschiede von A u. B auf verschiedenen DNA-Mikroarrays durch den Vergleich mit einer Referenz (R) bestimmt. C) Das „common reference design“ ist eine spezielle Realisierung des indirekten Vergleichs. Als Referenz dient hier eine Mischung zu gleichen Teilen aller RNA-Proben eines Transkriptom-Experimentes. Dargestellt ist das „common reference design“ des Transkriptom-Experimentes zum Vergleich aerober und phototropher Wachstumsbedingung von *H. sal.* R1 (nach Yang YH und Speed T, 2002).

Bei dieser Methode des indirekten Vergleichs, wird aus allen für ein Experiment zur Verfügung stehenden RNA-Proben eine Mischung zu gleichen Teilen hergestellt, die dann für jede durchzuführende kompetitive Hybridisierung als Referenz gilt und immer mit dem gleichen Fluoreszenzfarbstoff markiert wird.

Dieses Design des Experiments eliminiert den erwähnten Farbstoffeffekt und gewährleistet, dass die Vergleichbarkeit zwischen allen Arrays erhalten bleibt. Diese Tatsache ist für eine spätere statistische Auswertung der Transkriptom-Daten von entscheidender Bedeutung, weil nur auf dieser Basis der Experimentdurchführung unabhängige Daten erzeugt werden, die eine wichtige Voraussetzung zur Anwendung der Statistik darstellt.

In der folgenden Abbildung wird das „common reference design“ schematisch für das erste Experiment, dem Vergleich adaptierter aerober bzw. phototropher Zellen, dargestellt.

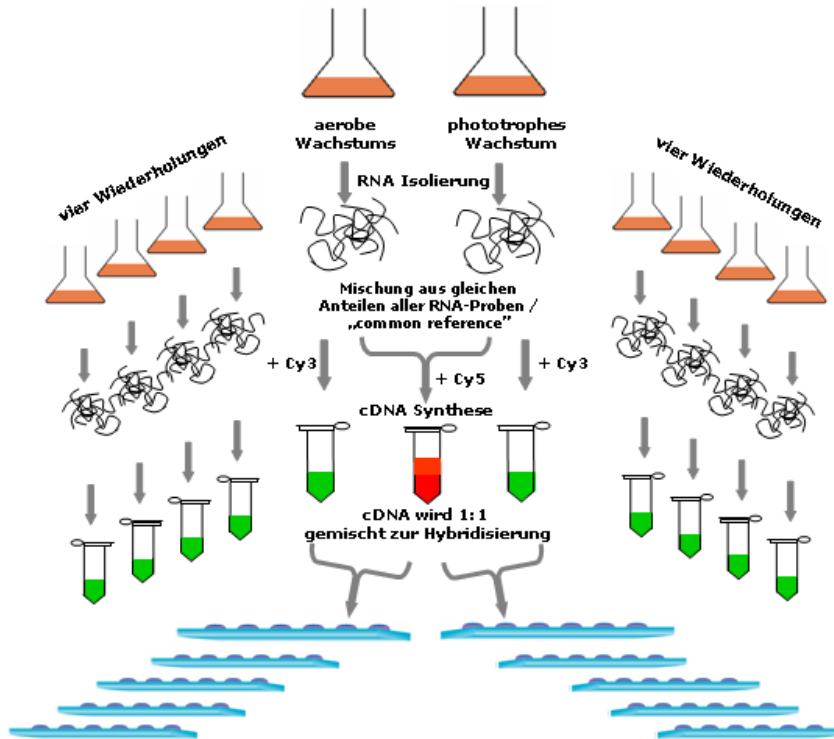


Abb. 20: Schematische Darstellung des Transkriptom-Experimentdesigns des indirekten Vergleichs nach dem „common reference design“ am Beispiel des 1. Transkriptom-Experimentes; Vergleich adaptierter aerober und phototropher Zellen.

Eine Zusammenfassung der verwendeten RNA-Referenzen ist in der folgenden Tabelle gegeben.

Experiment	„common reference“
1.Experiment: Vergleich aerob/phototroph	Je Wachstumsbedingung 5 Wiederholungen : $\Sigma$ Mischung aus 10 RNAs
2.Experiment: Zeitreihenexperiment: Wechsel aerob auf phototroph mit 4 Zeitpunkten	Je Zeitpunkt 4 Wiederholungen : $\Sigma$ Mischung aus 16 RNAs

3. Experiment :	
Zeitreihenexperiment: Wechsel	Je Zeitpunkt 5 Wiederholungen :
aerob auf phototroph mit 5	Σ Mischung aus 25 RNAs
Zeitpunkten	

Tab. 6: Zusammenstellung der verwendeten Referenzen für die durchgeführten Transkriptom-Experimente. Erläuterung des Begriffs „common reference“ siehe Text.

Des Weiteren muss in der Planung eines DNA-Mikroarray-Experiments eine ausreichende Anzahl an Wiederholungen vorgesehen sein, damit überhaupt eine statistische Auswertung der Hybridisierungsergebnisse durchgeführt werden kann (Churchill GA, et al., 2002; Yang YH und Speed T, 2002).

Diesem Aspekt wurde Rechnung getragen, indem alle Zellzucht-Experimente der Transkriptomanalysen als biologische Replikate in unabhängigen Versuchen vier- bis fünfmal wiederholt worden sind. Als technische Replikate, sind die fünf spots pro Gensonde auf den einzelnen Mikroarray-Slides anzusehen. Aus der Tabelle 5 ist nicht nur die Zusammensetzung der common reference der Transkriptom-Experimente zu entnehmen, sondern auch indirekt die Anzahl der durchgeführten Hybridisierungen, die den Grundstock der weiteren Auswertung der Experimente bilden. So wurde eine Gesamtzahl von 51 Gesamt-Genom-Mikroarrays von *H. sal.* R1 mit jeweils 15360 spots hybridisiert und statistisch ausgewertet.

Für diese 51 und alle anderen DNA-Mikroarray-Hybridisierungen wurden zur RNA Isolierung die Zellen von *H. sal.* R1 in ausreichend großen Kulturen kultiviert und die benötigten DNA-Mikroarray-Chips, für die verschiedenen Transkriptom-Experimente, in gleich bleibender Qualität, nach dem oben beschriebenen Layout, in ausreichender Stückzahl produziert.

### 3.2.2 Experimentdurchführung/Zellzucht

Eine wesentliche Voraussetzung von Transkriptomanalysen, ist die Vergleichbarkeit der verwendeten Kulturen. Deshalb wurde zur Zellzucht ein Standardprotokoll erstellt. Dieses sieht vor, das alle Experimente von einer sogenannten Stammkultur ausgehen und nach einem bestimmten Schema kultiviert werden (siehe experimenteller Teil).

In den folgenden Abschnitten wird die Zellzucht für verschiedene Wachstumsbedingungen und Stämme beschrieben.

### **3.2.2.1 Wachstum unter aeroben oder phototrophen Bedingung**

Zur Untersuchung der beiden Wachstumsbedingungen werden die Kulturen wie im experimentellen Teil der Arbeit beschrieben (3.1.2) inkubiert. Die aeroben Kulturen werden im Dunkeln in Sauerstoffatmosphäre und die phototrophen Kulturen wurden mit Licht und ohne Sauerstoffzufuhr kultiviert. Zur Unterbrechung der Sauerstoffversorgung bei der Kultivierung unter phototrophen Wachstumsbedingungen, wurden die Klettkolben direkt nach dem Animpfen mit einem Silikonseptum luftdicht verschlossen.

Für die Expressionsanalyse der Wachstumsbedingungen aerob/phototroph wurden insgesamt drei Transkriptomversuche durchgeführt, die im folgenden näher beschrieben werden.

1. Experiment: Vergleich adaptierter Zellen aerob/phototroph
2. Experiment: Zeitreihenexperiment Wechsel von aerob auf phototroph (Zeitabstände der Probennahme: 3 min, 15 min, 45 min und 120 min nach Stimulus)
3. Experiment: Zeitreihenexperiment Wechsel aerob auf phototroph (Zeitabstände der Probennahme: 45 min, 90 min, 180 min, 360 min und 1440 min nach Stimulus)

#### **3.2.2.1.1 Zellzucht für Vergleich adaptierter Zellen**

Als erstes Experiment wurde ein Vergleich von adaptierten aeroben und phototrophen Zellen durchgeführt.

Dazu wurden die Zellen, in jeweils fünf unabhängigen Versuchen nach dem, in Abbildung 21 dargestellten, Standardprotokoll kultiviert.

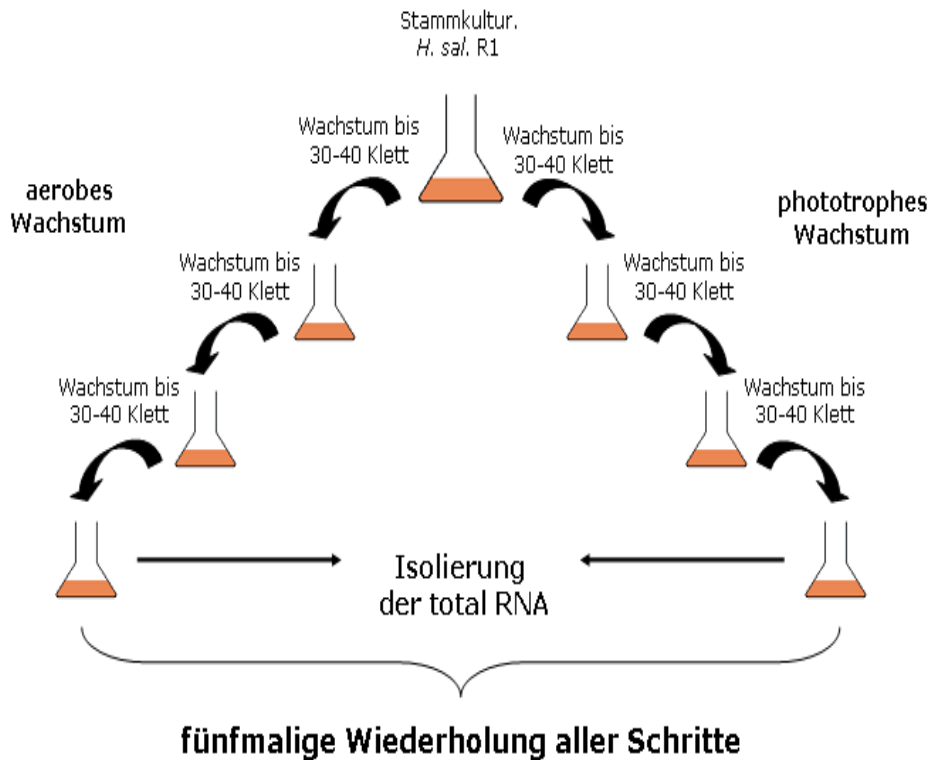


Abb. 21: Schematische Darstellung des standardisierten Ablaufs der Zellkultur zur RNA-Isolierung für das Transkriptom-Experiment 1 von *H. sal.* R1; dem Vergleich adaptierter aerober und phototropher Zellen (siehe experimenteller Teil).

Wie aus der Abb. 21 zu entnehmen ist, erfolgt die Probennahme zur RNA-Isolierung erst nach dreimaligen Kultivieren unter den zu untersuchenden Wachstumsbedingungen, um zu gewährleisten, dass sich die Zellen von *H. sal.* R1 an die jeweiligen äußeren Bedingungen adaptiert haben. Diese Abfolge des mehrmaligen Kultivierens unter den zu untersuchenden Wachstumsbedingungen ist ein wichtiger Bestandteil des für diese Arbeit entwickelten Standardprotokolls und ist bei allen Transkriptom-Experimenten durchgeführt worden.

Zum Beweis der phototropen Wachstumsbedingung der *H. sal.* R1 Zellen in den mit einem Silikonseptum luftdichtverschlossenen Kulturkolben wurde in Vorversuchen der Sauerstoffgehalt in diesen Kolben mittels einer Sauerstoffelektrode gemessen (siehe experimenteller Teil). Durch diese Vorversuche konnte eindeutig bewiesen werden, dass mit Beginn des Wachstums von *H. sal.* R1 der verbliebene Sauerstoff in den Kolben sehr schnell verbraucht wurde und sich somit bis zum Zeitpunkt der Probennahme phototrophe Wachstumsbedingungen für die Kultur eingestellt haben. Da bis zur RNA-

Isolierung die Zellen dreimal unter diesen Bedingungen kultiviert wurden, kann man von einer Adaptierung an das phototrophe Wachstum sprechen.

### O<sub>2</sub>-Verbrauch und Wachstum von *H.sal.* R1

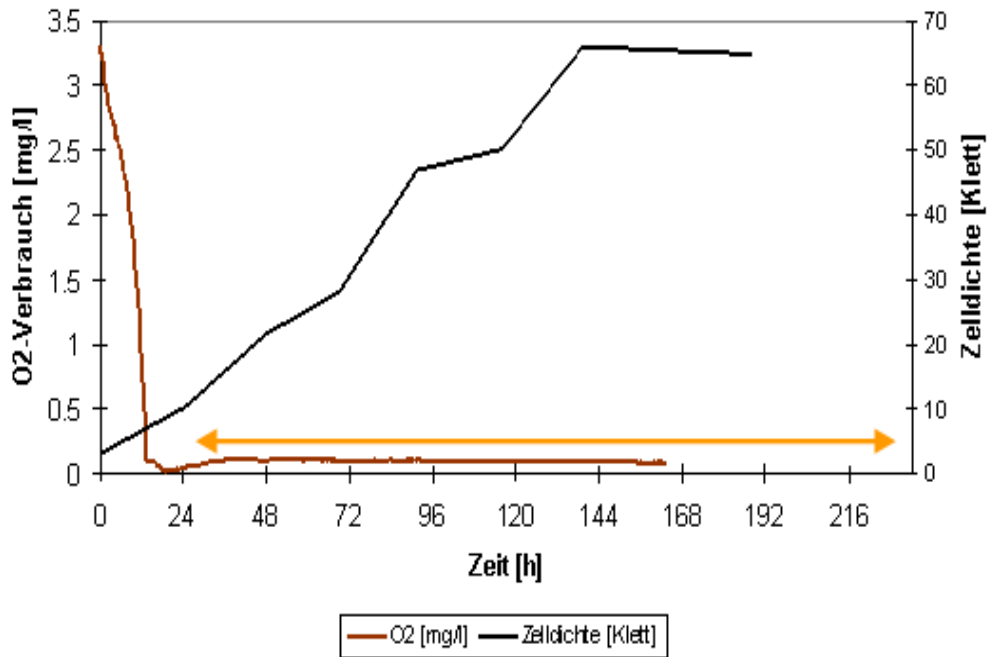


Abb. 22: Wachstumskurve unter phototrophen Bedingungen für *H. sal.* R1 (schwarz Linie). Parallel zum Wachstum ist in orange gezeigt, wie sich der Gehalt an Sauerstoff in dem Kulturkolben verändert. Deutlich zu sehen ist, dass bei Beginn, des beobachteten Wachstum von *H. sal.* R1 (zwischen 10 u. 15 Klett-Einheiten), der gesamte Gehalt an Sauerstoff verbraucht worden ist (dargestellt durch hellorangenen Doppelpfeil) und die Zellen bei einer Zelldichte von 30-40 Klett-Einheiten, dem Zeitpunkt der RNA-Isolierung, an das phototrophe Wachstum adaptiert sind.

Als Beispiel zum Ablauf der Zellzucht und des mehrmaligen Überimpfens sind für den Transkriptomversuch des Vergleichs aerob und phototroph, in Abb. 23, Wachstumskurven von den verschiedenen Kulturen der fünften Wiederholung des Experiments aufgezeichnet worden.

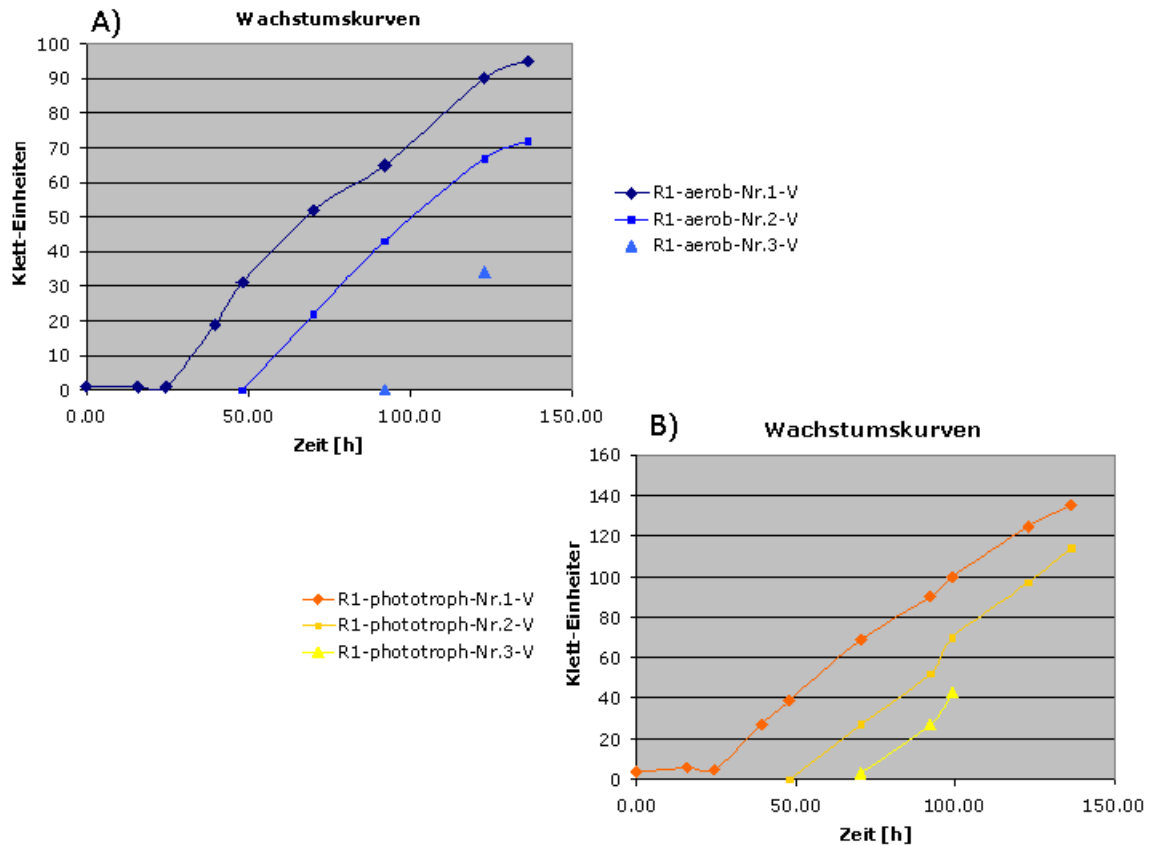


Abb. 23: Wachstumskurven von *H. sal.* R1 Zellen, die analog des Standardprotokolls (Abb. 11) für ein Transkriptom-Experiment aerob und phototroph kultiviert worden sind. In A) repräsentieren die blauen Kurven das aerobe Wachstum von *H. sal.* R1. Wie aus der Abb. 11 zu entnehmen ist, werden die Zellen, zur optimalen Adaption an die jeweilige Wachstumsbedingung, dreimal überimpft. Die Kurve mit den meisten Messpunkten steht für die Kultur des ersten Animpfens (dunkelblau) und die mit den wenigsten Messpunkten steht für die Kultur des dritten Animpfens (hellblau). Diese Kultur ist zur RNA Isolierung verwendet worden. Deutlich zu sehen ist, dass die Kulturen zwischen 30-40 Klett-Einheiten überimpft worden sind. Die unter B) dargestellten Wachstumskurven, zeigen das gleiche wie unter A), mit dem Unterschied, dass die Kulturen unter phototrophen Bedingungen kultiviert worden sind.

### 3.2.2.1.2 Zellzucht für Zeitreihenexperimente

Die folgenden Experimente waren so genannte „switch“-Experimente oder Wechselexperimente. Ziel dieser Experimente ist es, Zeit aufgelöst die differentiell exprimierten Gene bei einem Wechsel von aerobem auf phototrophes Wachstum zu bestimmen. In Abb. 24 ist das angewandte Standardprotokoll für die Zeitreihenexperimente dargestellt.



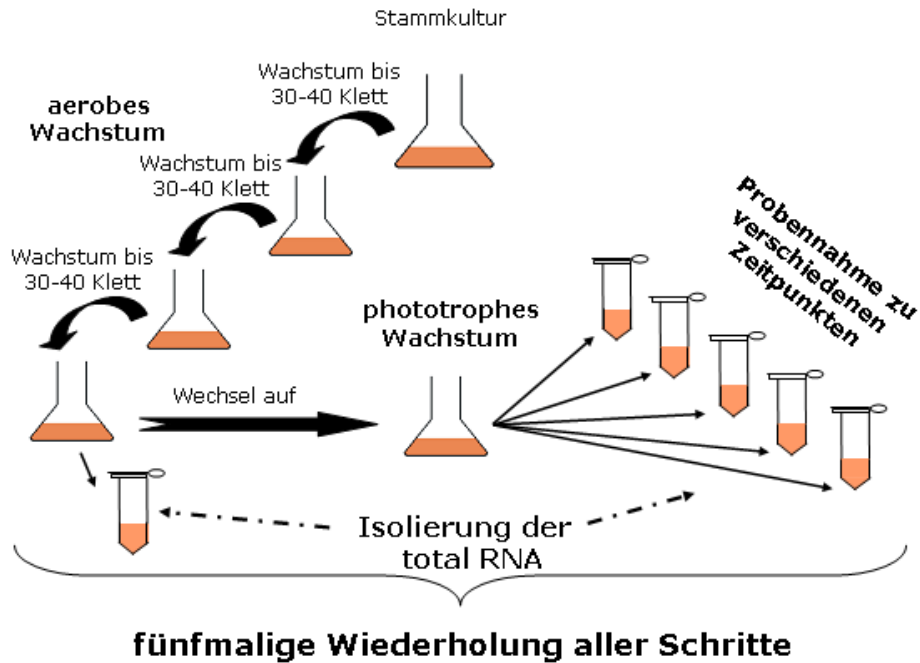


Abb. 24: Schematische Darstellung des standardisierten Ablaufs der Zellzucht zur RNA-Isolierung von *H. sal.* R1 für die Transkriptom-Experimente 2 und 3, den Zeitreihenexperimenten zum Wechsel von aerobem auf phototrophes Wachstum.

Es wurden zwei verschiedene Zeitreihenexperimente durchgeführt, mit einem Zeitintervall von 120 min (3 min, 15 min, 45 min und 120 min) und mit einem längeren von 24 h (45 min, 90 min, 180 min, 360 min und 1440 min). Diese zwei Zeitreihenexperimente wurden deshalb durchgeführt, weil man in dem ersten Experiment so Gene identifizieren kann, die schnell auf den Wechsel der Wachstumsbedingung reagieren, während man mittels des zweiten Experimentes eher differenziell regulierte Gene findet, die zu einem späteren Zeitpunkt in der Adaption wichtig sind bzw. den phototrophen Phänotyp bestimmen.

### 3.2.2.1.3 Vergleich verschiedener Stämme S9-WT / S9-hell-Mutante

Mit Hilfe der Mikroarray-Technologie können nicht nur Kulturen unterschiedlicher Wachstumsbedingungen miteinander verglichen, sondern auch Unterschiede verschiedener Stämme bestimmt werden. Als eine wichtige Voraussetzung für einen solchen Vergleich muss gewährleistet sein, dass die zu vergleichenden Stämme in ihrer Genomsequenz sich nicht zu stark unterscheiden, da sonst die Spezifität der Sonden auf den Mikroarrays nicht mehr gewährleistet ist.

Die hier zu vergleichenden Stämme sind der S9-Stamm, ein konstitutiver Überproduzent von Bakteriorhodopsin (Gropp F und Betlach M, 1994; 1995) und

eine Mutante dieses Stammes die farblos ist. Diese Mutante aufweist nicht mehr die typische Färbung des S9 Mutterstamms durch das Bakteriorhodopsin auf. Deshalb wird diese Mutante im folgenden als S9-hell (S9H) bezeichnet und der S9-Mutterstamm als S9-dunkel (S9D).



Abb. 25: Stammkulturen von der S9H-Mutante und des Wildtyps S9D.

Zur Untersuchung der Expressionsunterschiede, wurden diese Stämme aerob kultiviert und Gesamt-RNA isoliert. Im Unterschied zu den bereits beschriebenen Transkriptom-Experimenten, wurden hier die Hybridisierungen in drei unabhängigen „dye-swap“-Experimenten untersucht. Dieser direkte Vergleich der Expressionsniveaus in einem dye-swap Experiment ist in (Abb. 19) schematisch dargestellt. Dabei werden die zu untersuchenden RNA-Proben mit den beiden Fluoreszenzfarbstoffen Cy3 und Cy5 markiert und in jeder Farbstoffkombination auf einem DNA-Mikroarray hybridisiert. Zur Verdeutlichung ist das im Rahmen der Arbeit dreimal durchgeführte Experiment in Abb. 26 schematisch dargestellt worden.

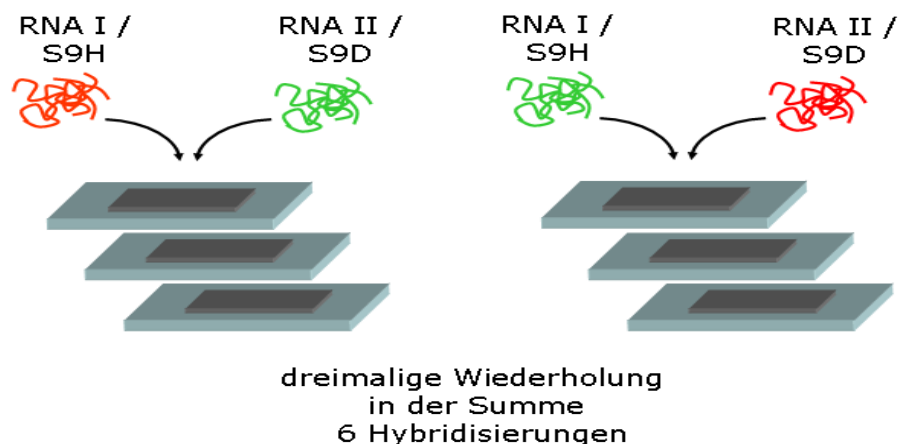


Abb. 26: Schematische Darstellung des „dye-swap“-Experimentdesigns für die Expressionsanalyse der S9H-Mutante und des S9D-WT.

### **3.2.3 Diskussion**

Nach der Beschreibung der Experimente und Vorstellung der verschiedenen Designvarianten der Transkriptomanalyse, schließt sich die Diskussion dieser zwei entscheidenden Teilaspekte eines DNA-Mikroarray-Experimentes an.

Die Untersuchung der differentiellen Genexpression mittels der Mikroarraytechnologie birgt einige Herausforderungen auf Grund der in den Daten enthaltenen Variabilitäten. Diese Variabilitäten sind systembedingt und treten auf verschiedenen Ebenen und zu unterschiedlichen Zeitpunkten des Experiments auf. Es bildet sich so ein komplexes Netzwerk von sich beeinflussenden Faktoren aus, welches auf zwei Hauptkomponenten reduziert werden kann. Zum einen auf die biologisch bedingten und zum anderen auf die experimentellen Faktoren.

Bei den biologisch bedingten Faktoren ist zu berücksichtigen, dass Prokaryoten sich rasch an Veränderungen ihrer Umgebung anpassen können. Aus diesem Grund sollte große Sorgfalt darauf verwendet werden, dass man die zahlreichen Faktoren, welche für eine Anpassung verantwortlich sind, zu kontrollieren. Dadurch kann bei der Transkriptomuntersuchung sichergestellt werden, die realen regulatorischen Netzwerke zu untersuchen und nicht unkontrollierte Variabilitäten interpretiert zu haben.

Diese Variabilitäten resultieren durch Veränderungen der Mikroumgebung der Zellen, wie Gradienten der Temperatur oder der Nährstoffe, durch Unterschiede in der Wachstumsphase der Zellen und Kulturen sowie z. B. durch den RNA-Metabolismus. Denn die RNA-Stabilität hat einen entscheidenden Einfluss auf die Transkriptmenge und die Physiologie der Zelle.

Zu den experimentell bedingten Faktoren zählen z.B. die Methode der RNA-Isolierung, die Größe der Probe, die Effizienz der Markierung der cDNA und die Hybridisierungsbedingungen.

Da all diese erwähnten Faktoren nicht einfach zu kontrollieren sind, wurde für diese Arbeit zur Extraktion der Gesamt RNA ein schnelles Protokoll entwickelt, welches durch eine rasche Inaktivierung der RNAsen-Aktivität gewährleistet, den RNA-Gehalt einer bestimmten Zellpopulation zu gewinnen. (Die genaue Beschreibung der Zellzucht für die Transkriptomexperimente ist den Abb. 21 und 24, sowie dem Text zu entnehmen.)

Das Standardprotokoll zur Zellzucht und RNA-Isolierung beinhaltet eine zeitnahe Probennahme und eine sehr schnelle Denaturierung der RNasen mittels der eingesetzten Guanidiniumthiocyanat-Methode (siehe 3.3.1). Zur Standardisierung der cDNA-Synthese wurde ein konventionell erhältliches Kit von Amersham Biosciences eingesetzt.

Die gleiche Sorgfalt zur Etablierung eines Standardprotokolls, für die Zellzucht, die Probennahme und der RNA-Isolierung, muss zum Design und zur Planung der Durchführung eines Mikroarray-Experiments aufgewandt werden.

Denn ein sorgfältiges Design des Mikroarray-Experiments bildet die Basis zur Generierung von qualitativ hochwertigen Daten und maximiert die Effizienz der statistischen Datenauswertung.

Das Experimentdesign für die durchgeführten Transkriptom-Experimente beinhaltet folgende vier Punkte:

1. Eine klar definierte biologische Frage bzw. Hypothese.
2. Die Vermeidung bzw. Minimierung von systematischen Fehlern bei der Behandlung und Handhabung des biologischen Materials und der praktischen Durchführung des Mikroarray-Experiments.
3. Eine simple und empfindliche statistische Datenanalysemethode, die den größtmöglichen Gehalt an Informationen aus dem Experiment herausholt.
4. Eine Gewährleistung der Einhaltung von Standards zur Ausführung von Mikroarray-Experimenten.

All diese Punkte sind recht allgemein gehalten, um zu zeigen, dass es bei der Durchführung eines Mikroarrayexperiments für den Wissenschaftler eine große Bandbreite von frei zu wählenden Parametern gibt.

Wie unter 3.2.1 beschrieben, gibt es zurzeit in der Literatur keine Hinweise auf das richtige Experimentdesign. Deshalb wurde für dieses Projekt stark darauf geachtet ein Design für die Experimente zu wählen, welches möglichst viele Empfehlungen berücksichtigt und die Durchführung sehr gut dokumentiert. Denn nur ein eindeutiges und flexibles Design eröffnet die Chance bei neuen Erkenntnissen, z.B. in der statistischen Analyse, die Daten erneut auswerten zu können. Außerdem wird durch eine Einhaltung von gewissen minimalen Standardanforderungen an ein Mikroarrayexperiment die Einbindung der Daten in

öffentliche Datenbanken ermöglicht, wodurch die Vergleichbarkeit und Transparenz der erhobenen Daten ersichtlich wird.

Mit der Gewissheit kein perfektes Mikroarrayexperiment-Design wählen zu können, wurde sich für die im englischen "common reference design" genannte Durchführung entschieden (Landgrebe J, et al., 2002, Yang YH und Speed T, 2002).

Bei diesem indirekten Vergleich werden alle Hybridisierungen mit einer Referenz aus einem Pool aller RNA-Proben und einer bestimmten Farbstoffkombination durchgeführt. Ein Vorteil dieses Designs besteht darin, dass der Pfad für alle Hybridisierungen nicht länger ist als zwei Schritte (siehe Abb. 19). Von Nachteil bei einem Referenzdesign ist der Punkt, dass die Hälfte der Messungen für die Referenzproben gebraucht wird, welche nur ein kleines bzw. geringes Interesse aufweisen. Es steht also nur die Hälfte an Informationen zur Verfügung, die man mit alternativen Designs, wie einem so genannten "loop design", hätte erzielen können.

Neben der Frage welche Referenz eingesetzt wird, ist die Frage der durchzuführenden Anzahl der Wiederholungen ebenso wichtig. So wurden alle Transkriptomexperimente vier- bis fünfmal in unabhängigen Experimenten wiederholt. Die Hybridisierungen wurden mit RNA-Proben gemacht, die aus jeweils unterschiedlichen Experimenten und verschiedenen Markierungen stammten. So sind die Wiederholungen als unabhängige biologische Replikate anzusehen. Diese größtmögliche Unabhängigkeit der Wiederholungen ist andererseits auch eine notwendige Voraussetzung eine verlässliche statistische Datenauswertung anwenden zu können (Churchill GA, 2002) (siehe 4.). Die genaue Anzahl der erfolgten Wiederholungen bzw. Hybridisierungen ist in Tab.6 zusammengefasst. Dabei wird deutlich, dass je komplexer ein Versuch aufgebaut ist, desto mehr Wiederholungen benötigt werden.

Es gibt zwei Hauptansätze, um den signifikanten Unterschied in der mRNA Menge zu bestimmen. Da wäre erstens der Ansatz, dass eine Änderung der mRNA Konzentration als Antwort auf regulatorische Signale zu sehen ist. Der zweite Ansatz geht davon aus, dass die mRNA Menge dynamisch durch ihren Abbau und Neusynthese kontrolliert wird. Diese Überlegungen führen wiederum zu zwei grundlegenden Experimentdesigns. Beim Typ eins bestimmt man die

Expressionsunterschiede zweier Bedingungen und Typ zwei beschreibt schrittweise die zellulären Prozesse als Antwort auf die Veränderung (Conway T und Schoolnik GK, 2003).

Die bestimmten Expressionsunterschiede wurden für beide grundlegenden Versuchsvarianten durchgeführt. So wurden in den ersten Transkriptomexperimenten die Unterschiede zwischen aerobem und phototrophen Wachstum ermittelt, um durch die zweiten Transkriptomexperimente die differentielle Expression mittels zweier Zeitreihenexperimente zu beschreiben. Das genaue Design der jeweiligen Experimente ist in Abb.21 und Abb.24 dargestellt und ausgeführt. In Abb.26 ist eine Variante des ersten Typs gezeigt. Bei diesem Versuch wurde der Wildtyp S9D mit der S9H Mutante verglichen. Diese Hybridisierungen fallen etwas aus dem Rahmen der vorher beschriebenen Transkriptomexperimente, da sie auf dem Farbwechseldesign beruhen und in dreifacher Wiederholung durchgeführt wurden.

### **3.3 Herstellung der markierten cDNA**

Der Gliederung des Flussdiagramms Abb. 11 folgend, gehen die weiteren Abschnitte auf verschiedene Punkte dieses Prozesses auf dem Weg der Herstellung der markierten cDNA ein.

Dabei ist zu beachten, dass es neben dem Farbstoffeffekt, durch Abweichungen in der Experimentdurchführung zu Unterschieden kommen kann. Diese Abweichungen können auf das Ergebnis des Hybridisierungsexperiments einen Einfluss haben. Aus diesem Grund ist es wichtig, einen standardisierten Ablauf des Protokolls zu etablieren und anzuwenden.

#### **3.3.1 Probennahme und Isolierung der Gesamt-RNA**

Für alle Mikroarray-Experimente wurde die Isolierung der Gesamt-RNA aus *H. sal.* R1 in Anlehnung an die Guanidiniumisothiocyanat-Methode von Chomczynski und Sacchi (1987) durchgeführt.

Die RNA-Isolierung nach dieser Methode hat den großen Vorteil, dass nach Beginn der Isolierung die Gefahr eines RNA-Abbaus durch RNasen erheblich herabgesetzt ist, da Guanidiniumsalze zur Denaturierung von Proteinen und damit Inaktivierung von RNasen sehr gut geeignet sind. Weil bei der Untersuchung der Expressionsunterschiede die Vergleichbarkeit der RNA-Proben

eine entscheidende Voraussetzung ist, wurde diese Methode der RNA-Isolierung für die Transkriptom-Experimente gewählt.

### 3.3.1.1 Qualitätskontrolle und Quantifizierung der RNA

Die Überprüfung der RNA-Qualität ist ein wichtiger Punkt in der Expressionsanalyse, denn degradierte oder teildegradierte RNA liefert kleine markierte cDNA-Fragmente, die das Hintergrundsignal signifikant erhöhen können.

Im Rahmen dieser Arbeit wurde die Qualität der RNA mittels des Bioanalyzers 2100 der Firma Agilent überprüft. Der Vorteil dieser Kapillarahochspannungselektrophorese ist, dass zur Analyse nur Mikrolitervolumina der RNA-Proben benötigt werden. So können in einem Arbeitsschritt, die Qualität und die Konzentration der RNA-Probe bestimmt werden.

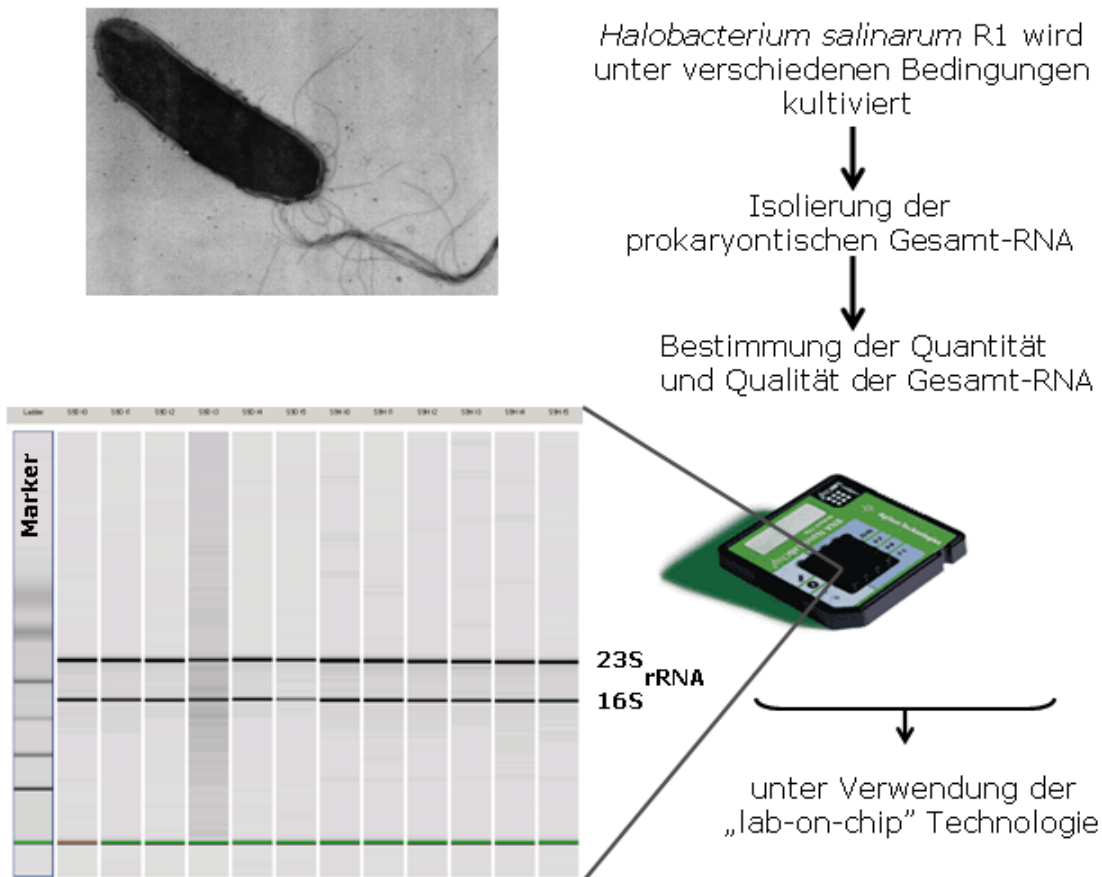


Abb. 27: Darstellung der Qualitätskontrolle und Konzentrationsbestimmung der Gesamt-RNA für die Transkriptom-Analysen von *H. sal.* R1 mittels RNA 6000 Nano Chips und dem Bioanalyser von Agilent. (oben links. Elektronenmikroskopische Aufnahme von *H. sal.* R1; Mitte: Bild eines RNA 6000 Nano Chips; unten links: Screenshot der Analyse-Software der Kapillarahochspannungselektrophorese.

Analog einer Agarosegelelektrophorese von prokaryontischer Gesamt-RNA mit sehr guter Qualität, werden die Banden der 23S und 16S ribosomalen RNA (rRNA) detektiert. Dabei sollte die Bande der 23S rRNA gegenüber der 16S rRNA-Bande die doppelte Signalintensität aufweisen. Die Methode der Qualitätskontrolle mit Hilfe der RNA 6000 Nano Chips ermöglicht die Detektion selbst der geringsten Anzeichen von Degradation der RNA.

Routinemäßig schließt sich nach jeder RNA-Isolierung ein DNase-I-Verdau an, um mögliche DNA-Kontaminationen zu entfernen, die sonst die Ergebnisse der Mikroarrayexperimente verfälschen könnten.

Erst nach Durchführung dieser Schritte werden die RNA-Proben für die Transkriptom-Experimente eingesetzt.

### **3.3.2 Synthese der markierten cDNA**

Zur Synthese der cDNA wurde für diese Arbeit das „CyScribe First-Strand cDNA Labeling Kit“ und Cy3-dUTP / Cy5-dUTP von Amersham Biosciences verwendet. Hierbei wird die Reaktion durch die Reverse-Transkriptase CyScript katalysiert, die, nach Herstellerangaben, gegenüber anderen Reversen-Transkriptasen eine verbesserte Einbaurate von farbstoffmarkierten dNTPs aufweist. Dies ist wichtig, denn generell werden fluoreszenzmarkierte dNTPs mit einer geringeren Effizienz in den cDNA-Strang eingebaut. Außerdem ist bei der direkten Markierung während der cDNA-Synthese zu beachten, dass die Farbstoffe Cy5 und Cy3 mit einer unterschiedlichen Effizienz eingebaut werden. Zur Vermeidung dieser Farbstoff bedingten Unterschiede wurde im Rahmen dieser Arbeit, so genannte Farbwechsel- („dye swap“)- oder „common reference“-Experimente durchgeführt, da sie durch ihren Aufbau es erlauben diese Effekte zu vermindern (siehe 3.2.1).

#### **3.3.2.1 Bestimmung der Einbaueffizienz**

Die Einbaurate der Fluoreszenzfarbstoffe wird indirekt durch die Menge an den nicht eingebauten fluoreszenzmarkierten Nukleotiden Cy3 (550 nm) und Cy5 (650 nm) bestimmt (siehe experimentellen Teil). Dieser Schritt wird nach jeder cDNA-Synthese und vor jeder Mikroarrayhybridisierung durchgeführt und ist notwendig zur Entscheidung, ob eine Hybridisierung mit diesen cDNAs überhaupt durchgeführt werden soll oder nicht.



$$\left(\frac{A}{E}\right) * (z \mu l) * \text{Verdünnungsfaktor} * 10^{12} \quad \text{mit}$$

A= Absorption Cy3 bei 550nm; Cy5 bei 650nm

E= Extinktionskoeffizient Cy3 (150000 l·mol<sup>-1</sup><sub>(bei 550nm)</sub>); Cy5 (250000 l·mol<sup>-1</sup><sub>(650nm)</sub>)

z μl= Volumen der cDNA nach der Reinigung

### 3.3.3 Diskussion

Die Basis für die Durchführung aller Transkriptomexperimente, egal welchen Designs, bildet die Qualität der gewonnenen Gesamt-RNA von *H. sal.* R1. Es musste sichergestellt sein, dass die RNA-Proben keine Anzeichen einer Degradierung aufweisen. Denn durch degradierte RNA-Proben, entstünden beim Prozess der Fluoreszenzmarkierung kleinere cDNA-Fragmente. Die ihrerseits maßgeblich für einen Anstieg der Hintergrundsignale verantwortlich (siehe 3.3.1.1) wären.

Die Qualität und gleichzeitig auch die genaue Konzentration der isolierten Gesamt-RNA wurde mit Hilfe des Bioanalyzers und der RNA 6000 Nano Chips der Firma Agilent untersucht. Der erhebliche Vorteil dieser Qualitätskontrolle besteht in dem geringen Probenvolumen und der parallelen Bestimmung der Quantität und Qualität der RNA. Ein Beispiel einer solchen Messung mit dem Bioanalyzer ist in Abb.27 gezeigt. Alle eingesetzten RNA-Proben haben diese Qualitätskontrolle durchlaufen und bestanden.

Eine sehr hohe Qualität der RNA ist zwar eine notwendige Voraussetzung für gute Ergebnisse, jedoch wurde auch die fluoreszenzmarkierte cDNA vor jeder Hybridisierung einer Qualitätskontrolle unterzogen. Die Effizienz der Markierung wurde indirekt durch die photometrische Messung des noch freien Fluoreszenzfarbstoffs im Reaktionsansatz kontrolliert (siehe 3.2.1).

## 3.4 Bestimmung der Expressionsunterschiede

Nach den Teilabschnitten der Herstellung der DNA-Mikroarrays, der Experimentdurchführung und der Synthese der markierten cDNA, schließt sich die Beschreibung des vierten Punktes an, die Detektion der Expressionsunterschiede.

### **3.4.1 Mikroarray-Hybridisierung**

Zur Optimierung des Signal/Hintergrund-Verhältnisses wird vor jeder Hybridisierung eine Vorhybridisierung durchgeführt. Während der Vorhybridisierung werden die nicht bedruckten Flächen des Mikroarrays mit BSA („bovine serum albumin“ / Rinderserum Albumin) blockiert, damit es nicht zu unspezifischen Wechselwirkungen zwischen den markierten cDNAs und der Oberfläche kommt. Es ist allerdings zu beachten, dass es beim Prozess der Vorhybridisierung auch zu Kontaminationen der Arrays durch das verwandte BSA kommen kann. Für dieses Projekt zeigten die Vorhybridisierungen jedoch auf die Hybridisierungsergebnisse einen optimalen Erfolg: Es konnte gezeigt werden, dass es auf den nicht bedruckten Flächen des Mikroarray-Chips zu keinen unspezifischen Bindungen kommt, die das Signal/Hintergrund-Verhältnis beeinflussen.

Eine längere oder intensive Exposition der hybridisierten Mikroarray-Chips mit Tageslicht, wurde vermieden, weil die Fluoreszenzfarbstoffe ausbleichen können was zu einem Signalverlust führen würde.

### **3.4.2 Bildakquisition der Mikroarray-Chips**

Die Signalintensitäten wurden nach der Hybridisierung der DNA-Mikroarrays mit Hilfe des GenePix™ 4000B Mikroarray Scanners (Axon Instruments) ausgelesen. Dabei werden nach Laser-Anregung (532 nm (Cy3); 635 nm (Cy5)) die emittierten Photonen simultan für die beiden Fluoreszenzfarbstoffe Cy3 und Cy5 detektiert und die Intensitätswerte als 16-bit Bilddateien gespeichert. Diese so erhaltenen Dateien stellen die Rohdaten jedes Mikroarrayexperimentes da.

Nach erfolgter Optimierung der Quantifizierung der Fluoreszenzsignale werden alle Mikroarrays eines Transkriptom-Experimentes mit diesen Einstellungen ausgelesen (siehe experimenteller Teil). So soll vermieden werden, dass aufgrund unterschiedlicher Lasereinstellungen in den späteren Ergebnissen Artefakte entstehen.

### **3.4.3 Bildanalyse der Mikroarray-Chips**

Der Ablauf der Bildanalyse ist ebenfalls festgelegt und genau im experimentellen Teil beschrieben. An dieser Stelle wird deshalb nur kurz auf die Methode der Bildanalysesoftware zur Hintergrundbestimmung eingegangen.

Die Klassifizierung der Pixelwerte einer Gensonde („spot“) in Vordergrund- oder Hintergrundwerte erfolgt durch Festlegung einer Kreisfläche, deren Mittelpunkt im Zentrum des „spots“ der entsprechenden Gensonde liegt. Der Durchmesser dieser Kreisfläche ist dreimal so groß, wie der Durchmesser des „spots“ selbst. Alle Pixelwerte dieser Region werden zur Bestimmung des lokalen Hintergrundwertes benutzt, mit Ausnahme von:

- den Pixelwerten der benachbarten „spots“
- den Pixelwerten einer Zwei-Pixel-Ausschluss-Region
- den Pixelwerten innerhalb des „spots“ des entsprechenden Gens selbst.

Bestimmung des lokalen Hintergrundsignals:

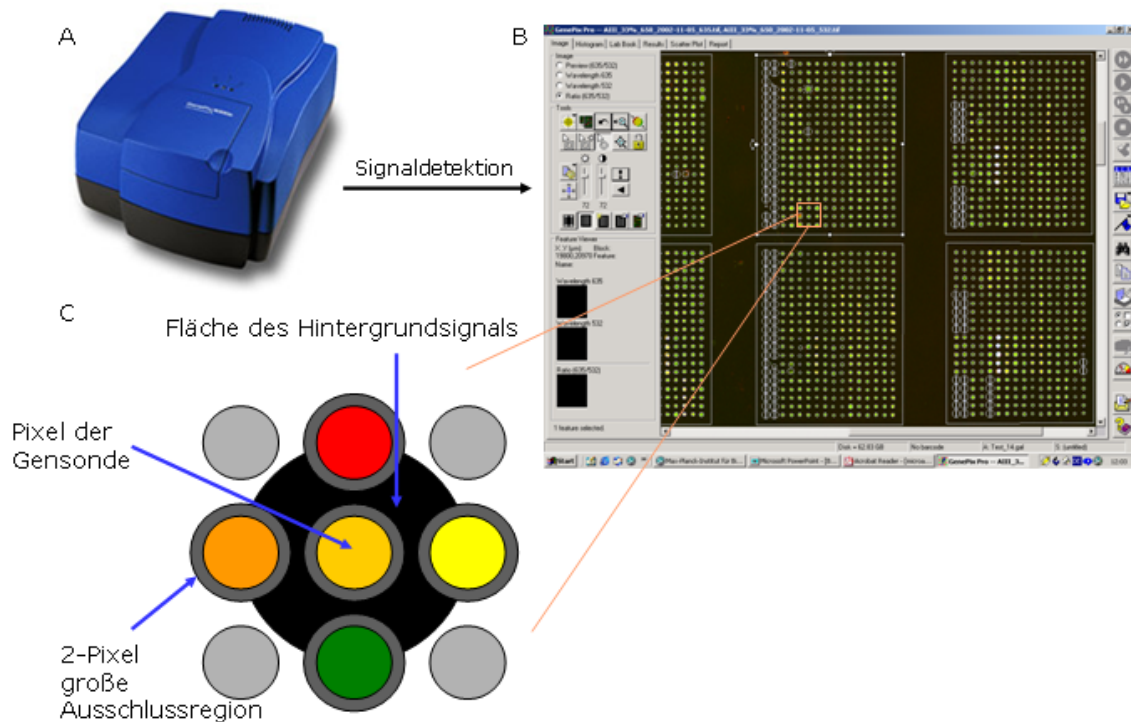


Abb. 28: Darstellung zur Bestimmung des lokalen Hintergrundsignals mit Hilfe der Software GenePix™ Pro 4.0. A) Bild des GenePix™ 4000B Mikroarray Scanners. B) Screenshot der Bildanalyse-Software mit einem berechneten Verhältnisbild. C) Schematische Darstellung einer Vergrößerung eines Bereichs von 9 Gensonden zur grafischen Beschreibung der Festlegung des lokalen Hintergrundwertes einer Gensonde. Farbige Kreise stellen die Gensonden dar. Ein dunkelgrauer Kreis um Gensonde ist die Ausschlussregion und die schwarze Fläche entspricht dem Hintergrund der jeweiligen Gensonde.

Die Software GenePix™ Pro 4.0 bestimmt den Durchmesser zur Festlegung des Bildpunktdurchmessers einzeln für jeden Punkt auf dem Mikroarray nach der Methode der „adaptive circle segmentation“. Das bedeutet, dass die Software automatisch den Durchmesser eines „spots“ variiert und somit zwischen großen und kleinen „spots“ auf dem Mikroarray unterscheiden kann. Der anzupassende

Durchmesser wird durch das Layout des Mikroarrays festgelegt und damit auch durch die Array-Liste (GAL-Datei). Nachteil dieser „adaptive circle segmentation“ ist, dass sie für exakt kreisrunde „spots“ auf dem Mikroarray-Slide optimale Ergebnisse liefert, während z.B. oval geformte Punkte nicht einwandfrei berücksichtigt werden können. Aufgrund der hohen Qualität, der für dieses Projekt hergestellten Mikroarray-Slides ist der überwiegende Anteil von 90-92% der Punkte kreisförmig, weshalb der Algorithmus des „spot finding“ Prozesses zu verlässlichen Ergebnissen führt.

Die Ergebnisse der automatischen Bildanalyse werden in einer Ergebnistabelle festgehalten und stellen zusammen mit den 16-bit Bilddateien der beiden Fluoreszenzfarbstoffe (Cy3/Cy5) die Rohdaten der Mikroarray-Experimente dar. Zur weiteren Datenanalyse bzw. zur statistischen Auswertung in der Programmumgebung R werden die erzeugten Tabellen als Textdateien exportiert.

Der Ablauf und die einzelnen Schritte der Datenanalyse werden in dem Abschnitt 4. näher erklärt.

### 3.4.5 Datenanalyse

Die statistische Auswertung der Hybridisierungsdaten der Transkriptom-Analyse von *H. sal.* R1 wurde in Kooperation mit Dr. Gerhard Welzl (GSF-Forschungszentrum für Umwelt und Gesundheit, GmbH) und Dr. Jobst Landgrebe (Universität Göttingen) durchgeführt. Dabei wurde, neben neuen R-Funktionen, im Wesentlichen das R/MAANOVA-Paket benutzt. R/MAANOVA ist eine erweiterbare, interaktive R Umgebung, die speziell für Mikroarray-Analysen entwickelt wurde. Die Abkürzung MAANOVA steht für „**MicroArray Analysis Of VAriance**“([www.jax.org/staff/churchill/labsite/software/anova/rmaanov/](http://www.jax.org/staff/churchill/labsite/software/anova/rmaanov/)).

Eine beispielhafte Beschreibung des Ablaufs der statistischen Auswertung, erfolgt in Abschnitt 4. Der präzise Ablauf ist im experimentellen Teil zu finden.

### 3.4.6 Diskussion

Die Prozesse der Hybridisierung, der Bildakquisition und Analyse dieser Arbeit basieren auf etablierten Abläufen bzw. Methoden. Deshalb wird im folgendenem Abschnitt etwas zur computergestützten Bildakquisition angemerkt.

Die Detektion der Fluoreszenzsignale erfolgte mittels des GenePix 4000B Mikroarray Scanners, der durch das Programm GenePix Pro 4.0 gesteuert wird

(siehe 3.4). Wichtig ist dass die Software einen so genannten "adaptive circle segmentation" Algorithmus verwendet (siehe 3.4.3). Damit ist die Software im Stande automatisch die Größe des auszuwertenden "spots" anzupassen. Diese Anpassung an die vorkommenden unterschiedlichen Größen der "spots" ist bei einem Transkriptomexperiment ein wichtiger Schritt und sollte deshalb nur in Ausnahmefällen manuell korrigiert werden. Denn alleine eine automatisch ablaufende Detektion der Signale, garantiert reproduzier- und nachvollziehbare Ergebnisse. Zum Ende dieser Arbeit stand eine überarbeitete Version der Bildanalysesoftware zur Verfügung, die neben dem "adaptive circle" auch den "adaptive shape" Algorithmus zur Auswahl hat. Mit Hilfe des "adaptive shape" Algorithmus kann Pixel genau die Form des "spots" ermittelt werden, wodurch die Bildanalyse verfeinert werden kann. Die Rohdaten für alle in dieser Arbeit beschriebenen Transkriptomexperimente, sind mit der Software Version 4.0 erhoben worden. Für einen Vergleich der beiden Bildanalysemethoden, müssten die Mikroarray-Bilddaten ein erneutes Mal bearbeitet werden. Dieser Punkt fällt damit in den Bereich der möglichen Schritte zur Prozessoptimierung (Bowtell D und Sambrook J, 2003).

Im vorliegendem dritten Abschnitt wurden alle Abläufe eines DNA-Mikroarray-Experimentes beschreiben, die im Labor ablaufen.

Der folgende vierte Abschnitt befasst sich mit den Schritten, die computergestützt ablaufen und enthält die Beschreibung der Auswertung der erhaltenen Rohdaten.

#### **4. Statistische Analyse der Transkriptom-Daten**

DNA-Mikroarray-Experimente haben sich zu einer hervorragenden Methode entwickelt in einem sehr großen Maßstab die Genexpression zu analysieren. Nach Bendigung der experimentellen Schritte und der Datenerhebung ist es das Ziel durch Anwendung geeigneter statistischer Datenanalysemethoden herauszufinden, welche der auftretenden Unterschiede in der Genexpression signifikant sind und welche nicht.

Ungeachtet der benutzten Mikroarray-Plattform steigt die Zahl der Methoden und Programme zur Analyse der Transkriptom-Daten stetig an. All diesen Methoden ist gemeinsam, dass sie mit der Datenvorverarbeitung und Normalisierung

beginnen, bevor mit einer weiterführenden statistischen Auswertung fortgefahren wird.

Für die statistische Datenanalyse im Rahmen diese Projektes ist die Kommandozeilensoftware R eingesetzt worden. Einer der größten Vorteile bei der Verwendung von R, ist die Möglichkeit jeden Schritt beim Prozess der statistischen Analyse auswählen und kontrollieren zu können. Dieser Punkt ist deshalb von entscheidender Bedeutung, da bis zum jetzigen Zeitpunkt noch keine universell anwendbare Analysemethode zur Verfügung steht und es wichtig ist jeden Schritt der Analyse nachvollziehen und dokumentieren zu können. Aus diesem Grund wurde als Grundlage der statistischen Analyse der *H. sal.* R1 Transkriptom-Daten das R/MAANOVA-Paket eingesetzt, welches für dieses Projekt durch selbst implementierte Funktionen optimiert wurde. Das folgende Flussdiagramm zeigt den modularen Aufbau der verwendeten statistischen Analysestrategie.

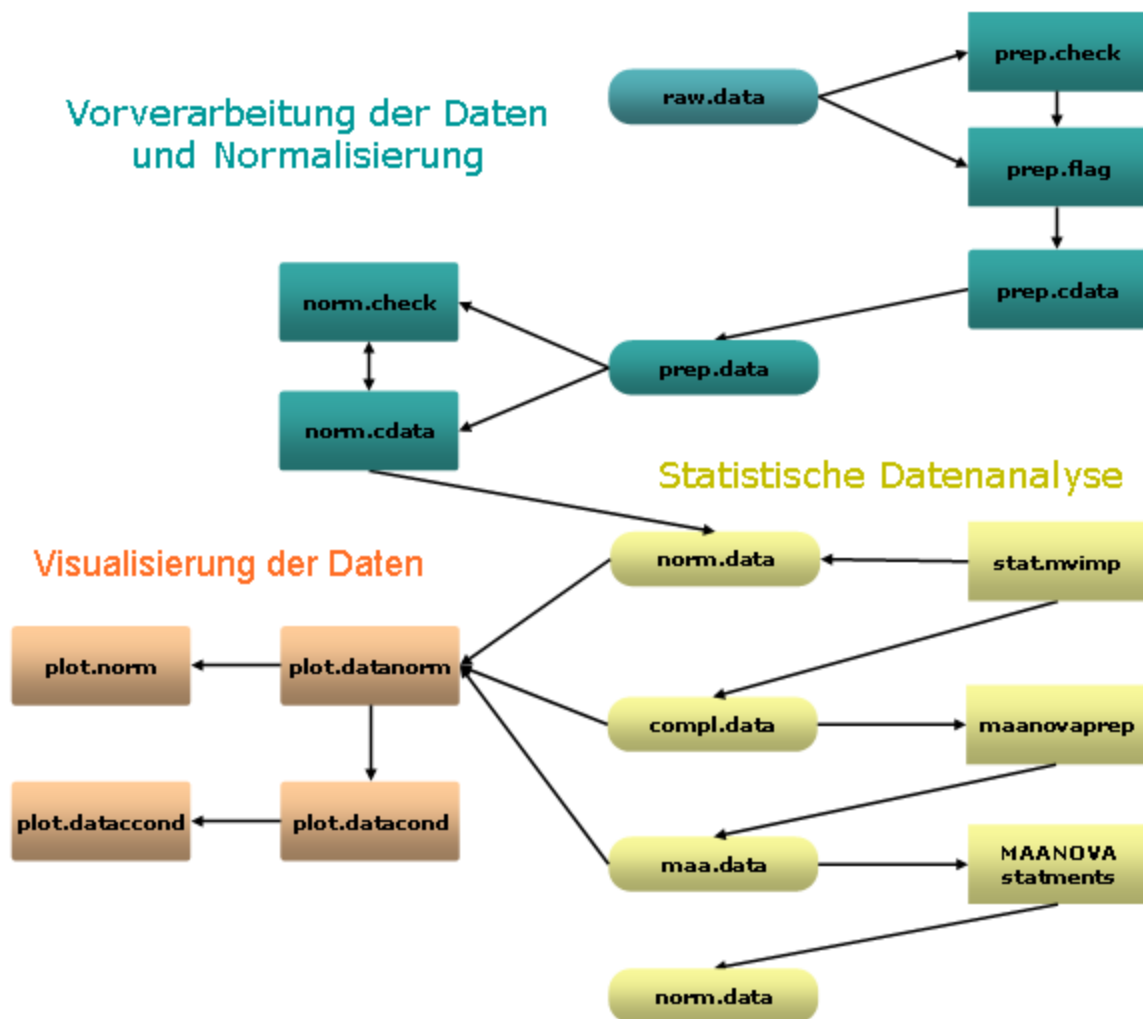


Abb. 29: Flussdiagramm mit den Arbeitsschritten der Transkriptom-Datenanalyse. Durch die Kästchen werden Funktionen repräsentiert, die auf Dateien, durch abgerundete Kästchen dargestellt, angewandt werden.

Die statistische Analysestrategie der Transkriptom-Daten basiert auf vier großen Bereichen, die im Flußdiagramm unterschiedlich farbig dargestellt sind:

- Vorverarbeitung der Daten („Data Preprocessing“)
- Normalisierung der Daten („Data Normalization“)
- statistische Datenanalyse („Statistical Data Analysis“)
- Visualisierung der Daten („Visualization“)

Bei der grafischen Darstellung der einzelnen Module repräsentieren die abgerundeten Kästchen die jeweiligen Dateien und die Kästchen enthalten die Funktionen, die je nach Stand der Analyse auf die Dateien angewandt worden sind.

Vor Beginn der statistischen Analyse wurden die Fluoreszenzintensitäten unter Gebrauch des GenePix™ 4000B Mikroarray Scanners detektiert, die Rohdaten mittels der Bildanalysesoftware GenePix™ Pro 4.0 erhoben und als Textdateien exportiert.

Diese exportierten Textdateien der Bildanalyse, die einen bestimmten Aufbau aufweisen müssen, werden in R eingelesen und dort so zusammengefasst, dass alle Hybridisierungsdaten, die zu einem Transkriptom-Experiment gehören, eine R-Datei bilden (siehe 11.6.3). Diese Datei wurde mit dem Namen raw.data bezeichnet und stellt den Ausgangspunkt der weiteren Berechnung dar.

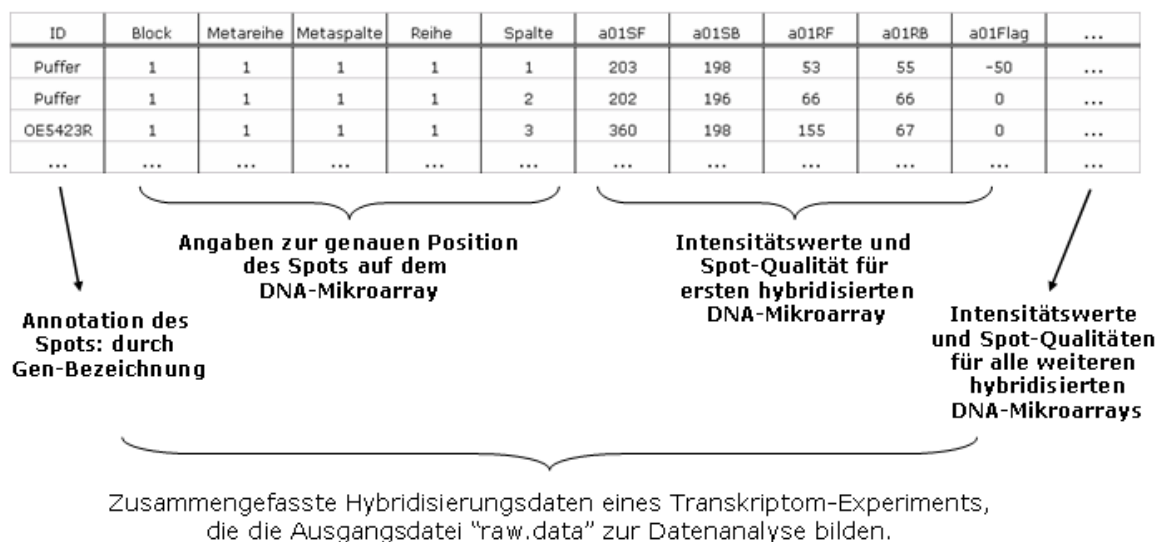


Abb. 30: Aufbau einer TAB-Getrennten Textdatei raw.data, die in R importiert wurde; Die Abkürzungen stehen für: ID = Kennung der Gene, a = aerob, 01 = Zahl der Wiederholung, SF = „sample foreground“ Vordergrundsignal der Probe, SB = „sample background“ Hintergrundwert der Probe, RF = „common reference foreground“ Referenzvordergrundwert, RB = „common reference background“ Referenzhintergrundwert. Durch ... soll auf weitere Daten in den Dateien hingewiesen werden.

In den folgenden Abschnitten werden die oben erwähnten vier Bereiche der Analysestrategie näher beschrieben.

### **4.1 Vorverarbeitung der Daten**

Die Mikroarraytechnologie basiert auf der Annahme, dass die Fluoreszenzsignalintensität eines bestimmten Gens auf dem Mikroarray mit der Menge der mRNA des entsprechenden Genes korreliert. Jedoch kann dieses gemessene Signal durch systematische Abweichungen nicht-biologischen Ursprungs beeinflusst werden. Diese systematischen Abweichungen müssen vor einer weiterführenden Datenanalyse entfernt werden. Das geschieht in den Schritten der Vorverarbeitung der Daten, die im Folgenden beschrieben wird.

Die genaue Vorgehensweise der Datenvorverarbeitung und Normalisierung ist im Experimentellen Teil unter 11.6.3 zu finden. Im Folgenden ist eine Zusammenfassung der Schritte gegeben.

Beim Vorverarbeiten der Daten werden verschiedene Funktionen auf die Datei raw.data angewandt, mit dem Ziel die Qualität der Daten zu überprüfen und die Daten dahingehend zu filtern, dass Punkte mit zu niedrigen oder zu hohen Intensitätswerten von der Analyse ausgeschlossen werden. Dieser Prozess wird „censoring“ genannt, wobei ein Datenfenster zwischen einer oberen und einer unteren Grenze der Intensitätswerte definiert wird. Die obere Grenze ist durch die Sättigung des Fluoreszenzsignales gegeben, die bei  $2^{16}=65536$  Fluoreszenzeinheiten liegt, und die untere Ausschlussgrenze des Signales wird dadurch bestimmt, dass das Signal nicht mehr vom Hintergrundsignal zu unterscheiden ist. Diese untere Grenze wird nicht durch einen willkürlich gewählten Schwellenwert festgelegt, sondern für jeden Block auf den einzelnen Mikroarrays berechnet. Die Vorgehensweise zur Berechnung dieser unteren Grenze erfolgt mittels einer Methode, die im späteren Verlauf beschrieben wird.

Im Zuge der Bildanalyse haben „spots“, die nicht den Qualitätsansprüchen genügen, eine Markierung, einen so genannten „flag“, bekommen. Die so markierten Daten müssen vor einer weiteren Analyse ebenfalls ausgeschlossen werden.

#### **4.1.1 Hintergrundkorrektur der Daten**

Ein besonders entscheidender Schritt bei der Datenanalyse ist die Hintergrundkorrektur. Hierbei muss bedacht werden, dass durch eine einfache



Subtraktion des gemessenen lokalen Hintergrundwertes eines Spots, eine größere, nicht-biologisch begründete Varianz in den Daten erzeugt werden kann, als diejenige die man versucht hat durch die Korrektur zu entfernen. Eine mögliche Lösung dieses Problems ist es, eine spezifische Datentransformation anzuwenden. Dabei muss beachtet werden, dass eine Transformation in logarithmische Werte die Gefahr beinhaltet, im unteren Intensitätsbereich die Unterschiede künstlich zu vergrößern, weil hier der additive Anteil des Fehlers dominiert. Ein Ausweg stellt eine monotone Transformation dar, die für den hohen Intensitätsbereich eine logarithmische, für den niedrigen Intensitätsbereich jedoch eine lineare Transformation durchführt. Eine solche monotone Transformation kann unter MAANOVA mit der so genannten „linlog“ Funktion ausgeführt werden.

Bevor mit dem Verfahren der Hintergrundkorrektur fortgefahren und mit der weiterführenden Datenanalyse begonnen wurde, sind die Daten der Vorder- und Hintergrundwerte beider Kanäle (Proben (grün, G, Cy-3) bzw. Referenz (rot, R, Cy-5)), mit Hilfe von grafischen Darstellungen, genau angesehen und analysiert worden. Dabei war es dienlich Box- oder MA-Grafiken zu erstellen, die der Anordnung der Gene auf dem Array Rechnung tragen, um Artefakte zu erkennen, die durch das Drucken, Hybridisieren und oder das Auslesen (Scannen) der Arrays entstanden sind.

Eine Box-Grafik kann als Zusammenfassung der Verteilung einer Variablen angesehen werden, die den Median (50%), das obere (75%) und untere (25%) Quartil und eventuelle individuelle Extremwerte darstellt. Dabei sind Extremwerte durch eine Abweichung von mehr als das 1,5 fache des Interquartilabstandes definiert und werden als Einzelwerte in das Diagramm eingezeichnet.

MA-Grafiken repräsentieren ein Diagramm der Intensitätswerte, bei dem die Punktwolke der Messwerte um 45° rotiert und die x-Achse neu skaliert wird. In einem MA-Diagramm wird der Logarithmus des Intensitätsquotienten  $M = \log_2 R/G = \log_2 R - \log_2 G$  gegen den Logarithmus des arithmetischen Mittels der Intensitäten  $A = \log_2 \sqrt{RG} = (\log_2 R + \log_2 G)/2$  aufgetragen (siehe Abb. 38).

Von Parmigiani et al. (2003) wurden folgende Faustregeln aufgestellt, nach denen man bei der Datenanalyse entscheiden kann, ob eine Korrekturmethode angewendet werden kann.

1. Sind bei der Darstellung der Hintergrundwerte räumliche Effekte zu beobachten, die nur bei einer Farbe bzw. einem Kanal auftreten, dann ist eine Hintergrundsubtraktion kritisch.
2. Ist bei der Darstellung des Probe/Referenz Diagramms ein so genannter „Hockeyschläger-Effekt“ bei niedrigen Intensitätswerten zu beobachten, dann kann eine Hintergrundkorrektur auf Genebene nicht zuverlässig durchgeführt werden. Dasselbe gilt, wenn in einem MA-Diagramm bei niedrigen Intensitätswerten ein so genannter „Fischschwanz-Effekt“ auftritt.
3. Es ist bei der Hintergrundkorrektur darauf zu achten, dass keine negativen oder sehr kleinen Werte entstehen, denn diese erzeugen bei der Quotientenbildung extreme Verhältnisse und damit unzuverlässige Werte.

Zur Überprüfung der Faustregeln 1 und 3, sind für alle Mikroarrays der unterschiedlichen Transkriptom-Experimente verschiedene Grafiken erzeugt worden, an Hand derer es möglich ist räumliche Effekte in den Daten zu erkennen. Exemplarisch sind in Abb. 31 Box-Grafiken des Mikroarray-Chips 22 des zweiten Zeitreihenexperiments dargestellt.

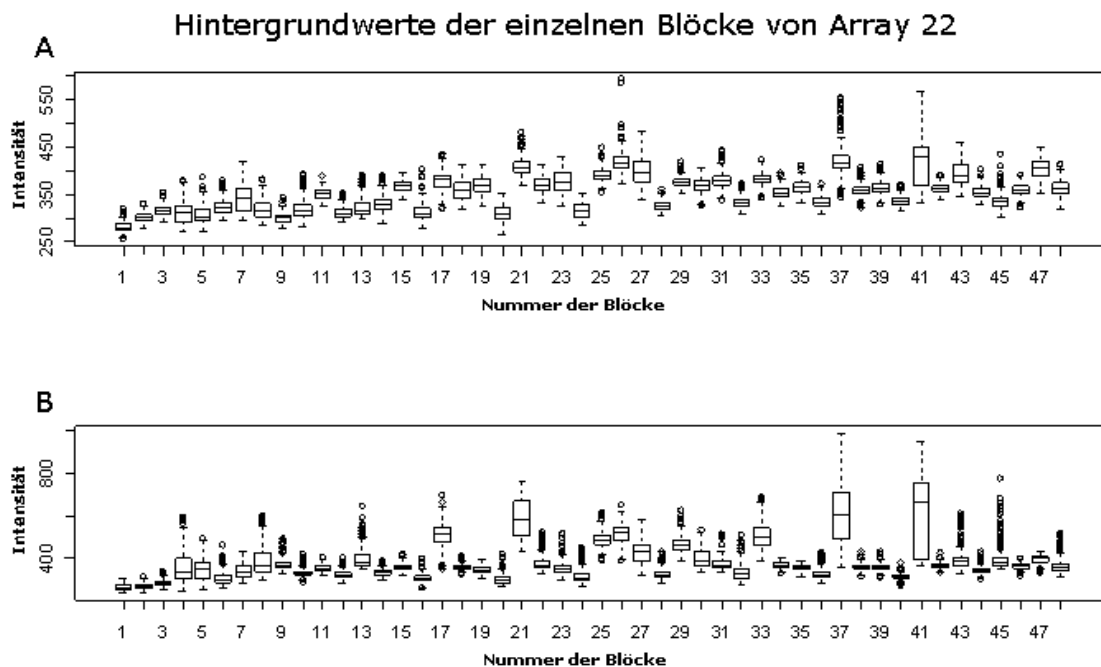


Abb. 31: Box-Grafik der Hintergrundwerte für die Probe/Cy3-Kanal (A) und die Referenz/Cy5-Kanal (B) getrennt nach den Blöcken (je 320 Einzelwerte) aufgetragen. Die horizontale Linie in der Mitte der Boxen zeigt den Median. Die obere Linie zeigt 75%, die untere 25% des Quartils an. Die Linien außerhalb der Boxen zeigen die experimentellen Maximal- und Minimalwerte an. Die Kreise repräsentieren Extremwerte einzelner spots, die außerhalb der Grenzen liegen.

In den gezeigten Box-Diagrammen der Abb. 31, sind die lokalen Hintergrundwerte beider Kanäle gegen die Anzahl der Drucknadeln bzw. Blöcke entsprechend der Anordnung auf dem *H. sal.* R1 Mikroarray dargestellt. Dabei entspricht ein Block einer  $n(16) \times m(20)$  Punkt-Matrix und wird bei der Mikroarray-Herstellung von einer Drucknadel erzeugt, wie in Abschnitt 3.1.2 erwähnt. In den Box-Diagrammen der Referenz und Probe ist deutlich ein räumlicher Effekt zu beobachten, der in beiden Kanälen gleichermaßen auftritt. Dieser räumliche Effekt zeigt sich dadurch, dass man einen leichten Anstieg der Mediane in den Kästchen jedes vierten Blocks sehen kann, was eindeutig mit der Position des Blocks auf dem Chip korreliert. Der gleiche Effekt, zwar weniger ausgeprägt, tritt auf, wenn die Vordergrundwerte für jeden Block des Mikroarray-Chips 22 des zweiten Zeitreihenexperiments in einem Box-Diagramm aufgetragen werden.

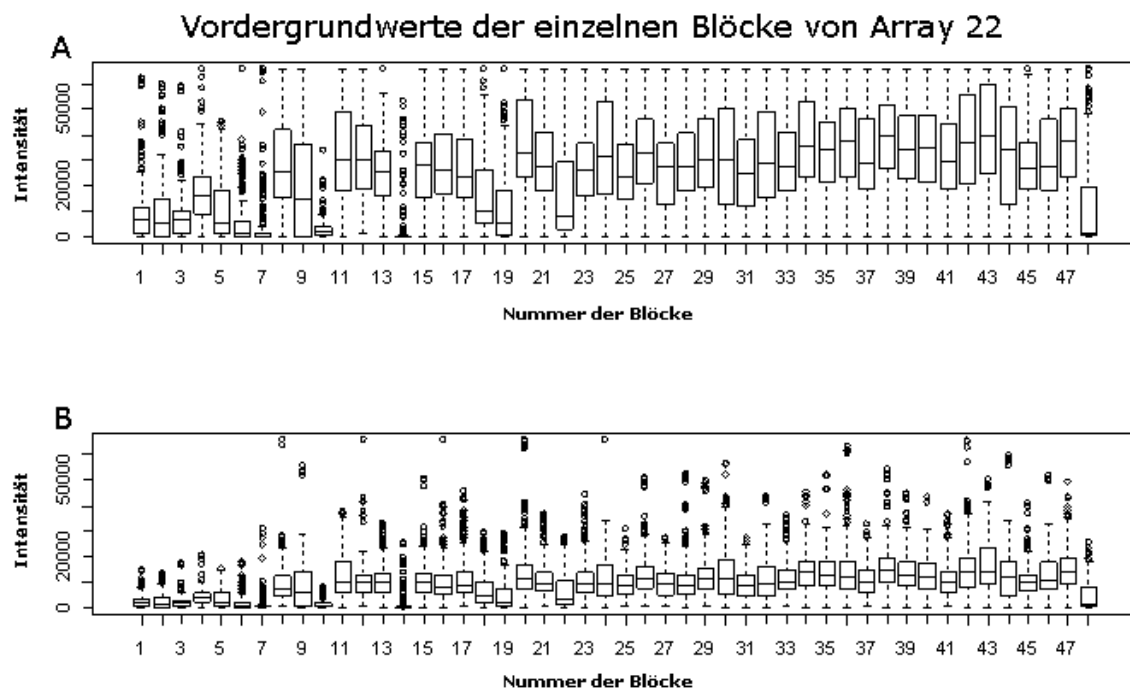


Abb. 32: Box-Grafik der Vordergrundwerte für die Probe/Cy3-Kanal (A) und die Referenz/Cy5-Kanal (B) getrennt nach den Blöcken aufgetragen.

Eine eindeutigere Darstellungsform von räumlichen Effekten sind Höhenliniendiagramme. Als Beispiel ist in Abb. 33 für Mikroarray-Chip 1 des ersten Transkriptom-Experiments ein Höhenliniendiagramm der M-Werte ( $=\log_2 R/G$ ) aufgetragen worden.

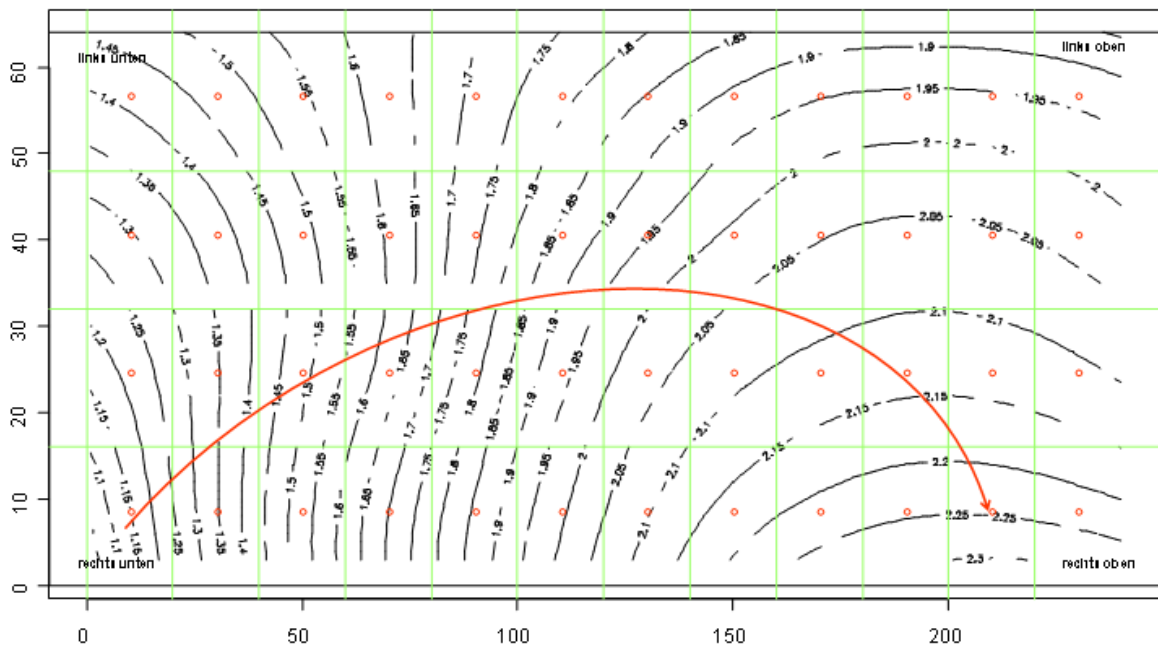
Höhenliniendiagramm von  $M = \log_2 R - \log_2 G$  für Array 1

Abb. 33: Höhenliniendiagramm („contour plot“) der M-Werte für Array 1 nach der räumlichen „block wise“ Glättung mit dem „loess fit“ (span Faktor = 0,75). Die 48 Blöcke des Arrays werden durch die grünen Kästchen mit dem roten Punkt in der Mitte dargestellt. Durch den roten Pfeil wird der steigende Trend der Höhenlinien dargestellt.

Der für den Mikroarray 1, aus dem ersten Experiment, dem Vergleich aerob/phototroph, sichtbare räumliche Effekt, ist für alle Arrays dieses Transkriptom-Experiments, in ähnlicher Weise vorhanden. Die M-Werte steigen von unten zur oberen rechten Ecke des Mikroarrays an.

Zusammenfassend kann man sagen, dass ein deutlicher räumlicher Effekt zu beobachten ist, der mit dem Block und damit mit der Position auf dem Mikroarray zusammenhängt und für den Referenz- und Probenkanal in gleicher Weise zu verzeichnen ist. Somit ist es möglich, nach den oben erwähnten Faustregeln, eine Hintergrundkorrektur mit den Daten durchzuführen.

Zur Festlegung des Hintergrundbereichs, wurde darauf verzichtet, einen Grenzwert zu bestimmen. Stattdessen wurde eine Methode angewandt, die auf der Basis von Höhenlinien einer zweidimensionalen Normalverteilung einen Abstand für jedes Paar von Expressionswerten (Probe und Referenz) berechnet. Dazu wurde eine zweidimensionale Normalverteilung der Hintergrundwerte für jeden einzelnen Block vorausgesetzt. Für jeden Block (i) wurde das arithmetische Mittel und die Standardabweichung für die Hintergrundwerte der Referenz (Kanal 1,  $x_1$ ) und der Probe (Kanal 2,  $x_2$ ) berechnet. Dazu wurde die „Minimum Volumen Ellipsoid“ Methode verwandt, da sie eine einfache Erstellung der elliptischen

Höhenlinien erlaubt und als ein Schätzer anzusehen ist, der durch stark abweichende Werte (Ausreißer) kaum beeinflusst wird. Somit stellt diese Methode ein zuverlässiges Verfahren zur Bestimmung des Mittelwerts und der Standardabweichung dar. Der Mittelpunkt der zu erzeugenden Ellipse wird derart geschätzt, dass bei minimalem Volumen der Ellipse, eine bestimmte Anzahl an Punkten abgedeckt wird. Zur Berechnung der Ellipse wurde folgende Gleichung benutzt:

$$\left(\frac{x-m_{1i}}{sd_{1i}}\right)^2 - 2\rho\left(\frac{x-m_{1i}}{sd_{1i}}\right)\left(\frac{x-m_{2i}}{sd_{2i}}\right) + \left(\frac{x-m_{2i}}{sd_{2i}}\right)^2 = -2*(1-\rho)^2*\log(\alpha) \quad (1)$$

$m_{1i}$ = arithmetisches Mittel des Hintergrundwertes Kanal 1 (Referenz) für Block i

$sd_{1i}$ = Standardabweichung des Hintergrundwertes Kanal 1 (Referenz) für Block i

$m_{2i}$ = arithmetisches Mittel des Hintergrundwertes Kanal 2 (Probe) für Block i

$sd_{2i}$ = Standardabweichung des Hintergrundwertes Kanal 2 (Probe) für Block i

$\rho_1$ = Korrelationskoeffizient ( $x_1, x_2$ ) für Block i

Diagramm der Hintergrundwerte mit Ellipsoid für Array 22 Block 1 / Nadel 1

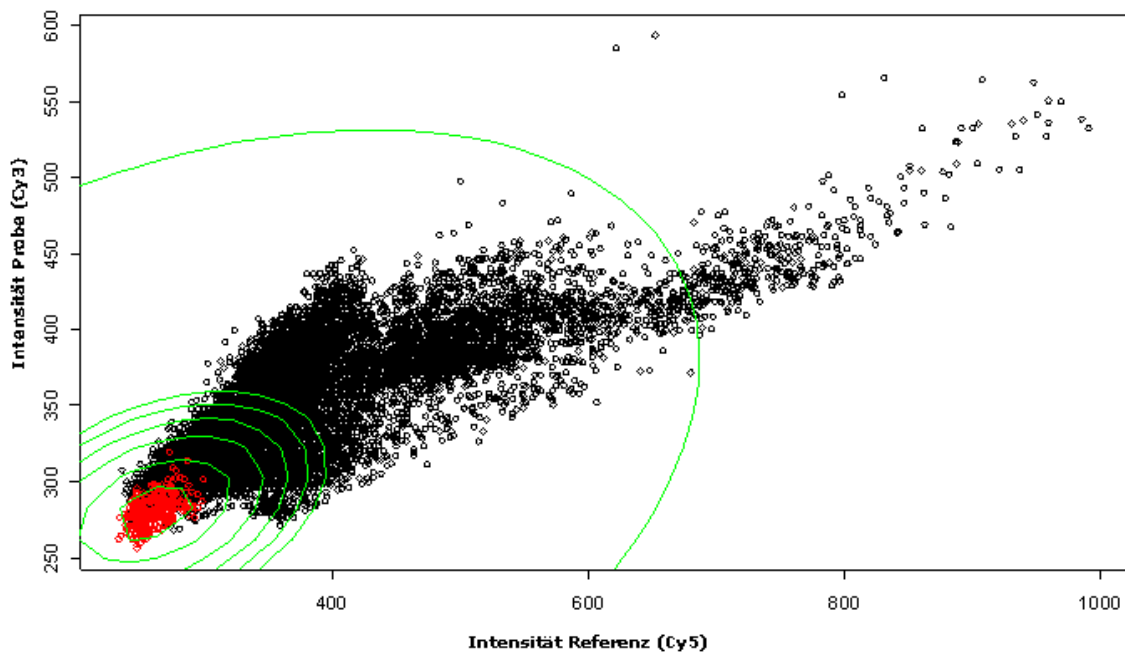


Abb. 34: Diagramm der Hintergrundwerte für die Probe/Cy3-Kanal und die Referenz/Cy5-Kanal des Arrays 22, in dem die Werte des Blocks 1 in rot gezeichnet wurden. In grün sind die Höhenlinien einer zweidimensionalen Normalverteilung eingezeichnet, die berechnet wurden auf der Basis der „minimum volume ellipsoid“ Methode (Gleichung (1)) unter Verwendung verschiedener Werte für  $\alpha$ .

Basierend auf der Gleichung (1) wurden für jedes Paar von Hintergrundwerten (Referenz und Probe (Abb. 34)) und Vordergrundwerten (Referenz und Probe (Abb. 35)) sämtlicher Blöcke aller DNA-Mikroarrays der drei Transkriptom-Experimente der Wert für  $\log \alpha$  berechnet. In den Abbildungen 34 und 35 sind als Beispiel die elliptischen Höhenlinien des Blocks 1 von Mikroarray 22 des zweiten Zeitreihenexperiments für die Hintergrund- und die Vordergrundwerte gezeigt. Zur Veranschaulichung ist in Abb. 35 dargestellt, wie gut die berechnete Ellipse den Bereich der niedrigen Intensitäten abdeckt, der in der Abbildung durch die roten Punkte repräsentiert wird.

In Abb. 35 wird dargestellt, welche von den Vordergrundwerten des Blocks 1 vom Arrays 22 auf Grund des berechneten  $\alpha$ -Wertes in den Hintergrundbereich eingestuft und somit von der weiteren Datenanalyse ausgeschlossen wurden, da sie nicht vom Hintergrundsignal zu unterscheiden waren.

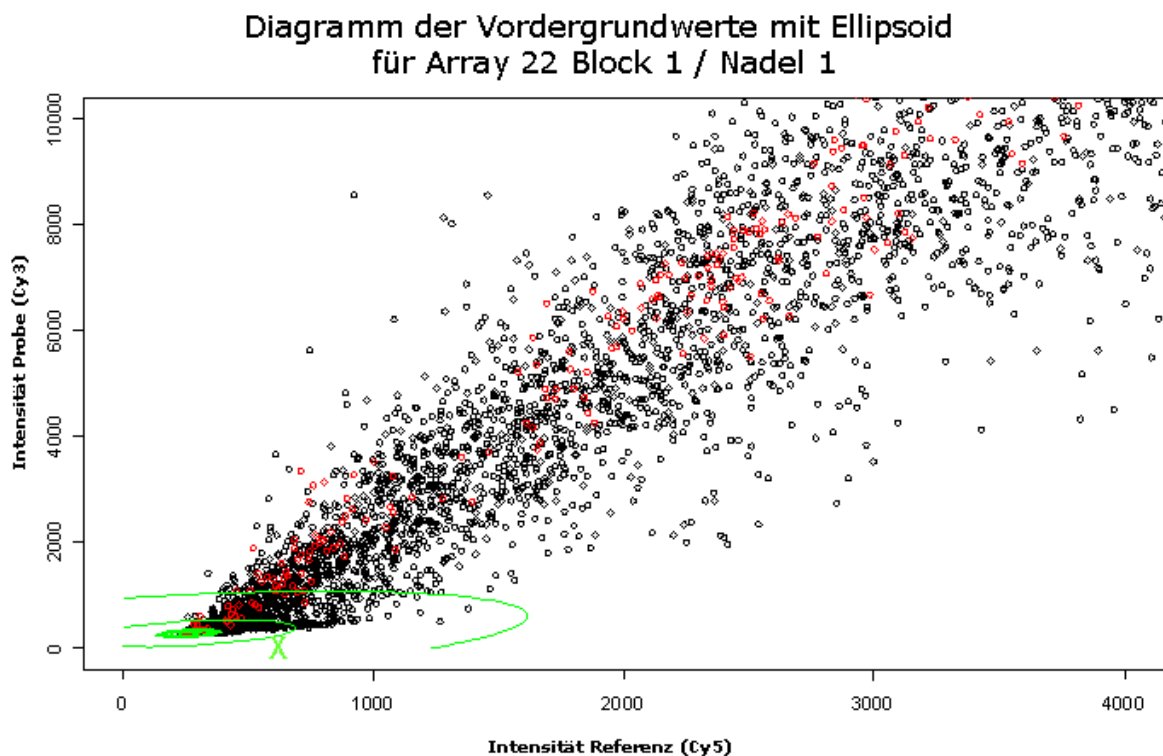


Abb. 35: Diagramm der Vordergrundwerte für die Probe/Cy3-Kanal und die Referenz/Cy5-Kanal des Arrays 22, in dem die Werte des Blocks 1 in rot gezeichnet wurden. In grün sind dieselben Höhenlinien eingezeichnet wie in Abb. 25. Alle Vordergrundwerte, die innerhalb einer ausgewählten Ellipse liegen, sind vom Hintergrundsignal nicht zu unterscheiden und werden von der weiteren Analyse ausgeschlossen. (Die mit einem X gekennzeichnete Höhenlinie ist die größte aus Abb. 34.)

Ein sehr kleiner Wert von  $\log \alpha$  markiert die Vordergrundwerte, die sich vom Hintergrund unterscheiden. Infolgedessen wurden die elliptischen Höhenlinien

eines bestimmten Wertes für  $\log \alpha$  dazu verwendet, die Expressionswerte auszuschließen, deren Wert von  $\log \alpha$  größer als eine bestimmte Konstante ( $c$ ) ist. Den Einfluss des Ausschlusses der Werte mit einem  $\log \alpha > c = -500$  ist in den folgenden Abbildungen zu sehen.

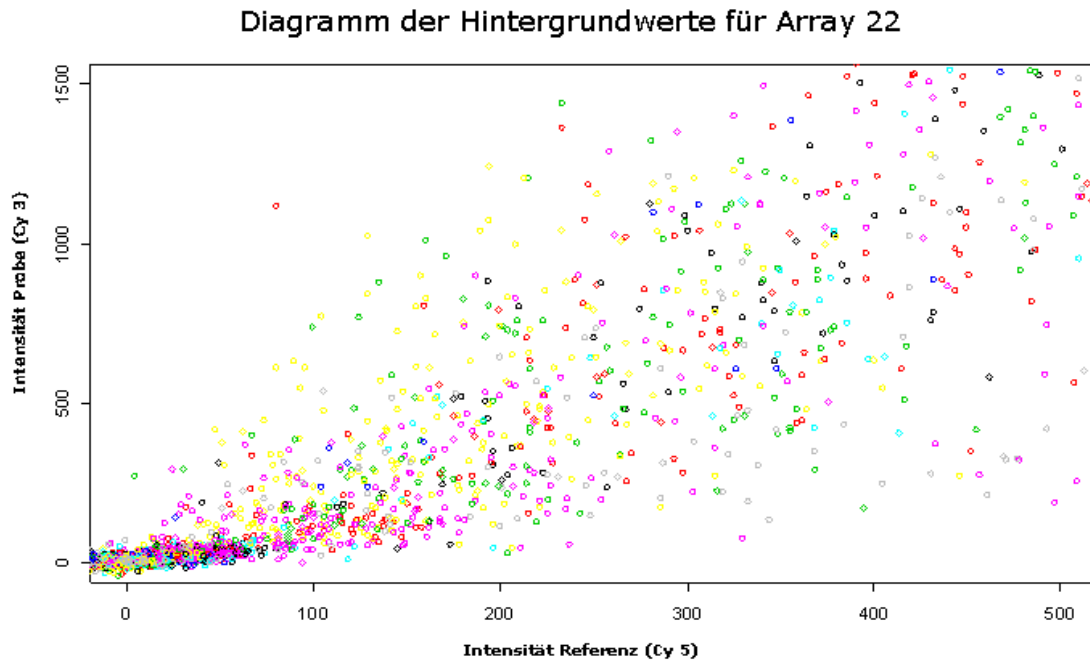


Abb. 36: Diagramm der Hintergrundwerte für die Probe/Cy3-Kanal und die Referenz/Cy5-Kanal des Arrays 22, vor der Berechnung des Hintergrundniveaus. Die Werte im Diagramm sind für jeden Block in einer anderen Farbe dargestellt. Im unteren Intensitätsbereich ist deutlich eine Anhäufung von Werten zu sehen, wodurch sich die „Hockeyschläger-Form“ der Punktwolke ausbildet.

In dem Diagramm der Abb. 36 sind die Vordergrundwerte der Referenz gegen die Probe aufgetragen. Im unteren Intensitätsbereich ist eindeutig eine Neigung bzw. ein Abknicken der Werte zu sehen, wodurch sich die so genannte „hockey stick“ (Hockeyschläger) Form der Punktwolke ausbildet.

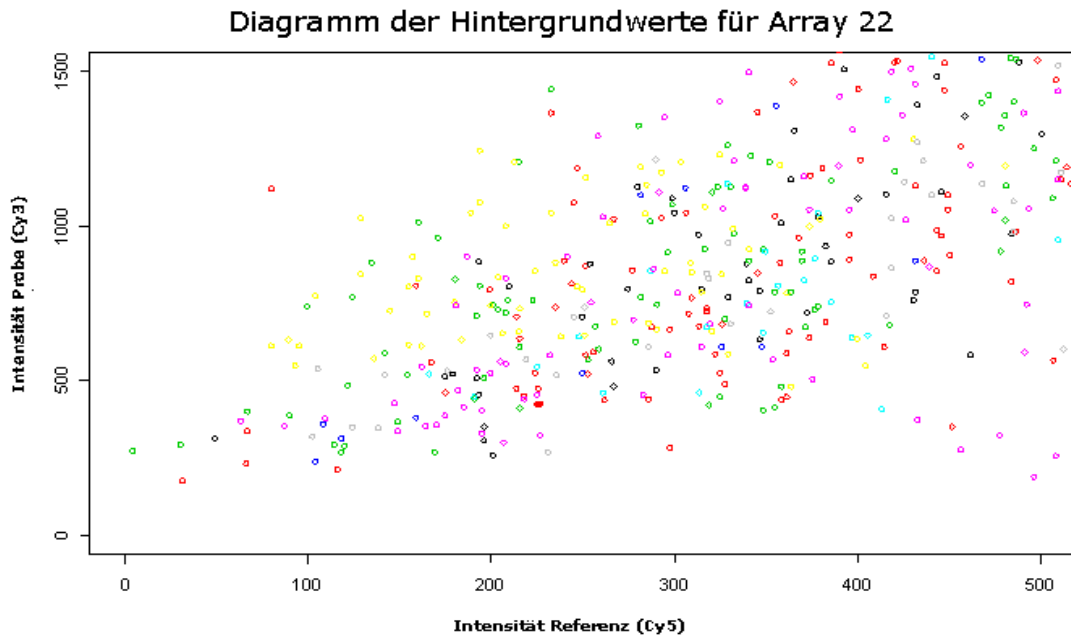


Abb. 37: Diagramm der Hintergrundwerte für die Probe/Cy3-Kanal und die Referenz/Cy5-Kanal des Arrays 22, nach der Berechnung des Hintergrundniveaus. Die Werte im Diagramm sind für jeden Block in einer anderen Farbe dargestellt.

Im Gegensatz zur Abb. 36, ist hier in Abb. 37 die Hockeyschläger-Form nicht mehr zu beobachten, weil durch die Festlegung des Hintergrundniveaus für den unteren Intensitätsbereich unzuverlässige Expressionswerte ausgeschlossen worden sind, wodurch sich für diesen Bereich eine Art Schärfung ergibt.

Diese Vorgehensweise hat den Vorteile, dass die untere Ausschlussgrenze der Expressionswerte nicht durch einen willkürlich festgelegten Schwellenwert bestimmt, sondern individuell durch eine mathematische Methode berechnet wird.

## 4.2 Normalisierung der Daten

Nach der Hintergrundbereinigung müssen andere experimentell begründete systematische Fehler beseitigt werden, die entstanden sind durch die Fluoreszenzfarbstoffe, die Fluoreszenzintensitäten, die Drucknadeln, Markierungsunterschieden oder den Druckvorgang (Yang YH, et al., 2002, Schuchhardt J, et al., 2000). Solche experimentellen Effekte zu identifizieren und zu entfernen ist das Ziel einer Normalisierung. Selbstverständlich ist es notwendig eine Korrektur der Fluoreszenzsignale mittels Normalisierung vor einer weiteren Analyse durchzuführen, um sicherzustellen, dass die Expressionsunterschiede einen biologischen Ursprung haben und nicht durch systematische Effekte hervorgerufen wurden. Eine Normalisierung von



Mikroarraydaten ist generell als ein Zweistufenprozess anzusehen. Auf der ersten Ebene werden die Intensitätssignale der einzelnen DNA-Mikroarrays normalisiert, um auf der nächsten Ebene eine Normalisierung der DNA-Mikroarrays untereinander durchzuführen.

#### 4.2.1 Normalisierung der einzelnen Mikroarrays

Auf dieser Stufe der Normalisierung werden nur die Daten von den einzelnen Mikroarrays bearbeitet. Um systematische Variationen in den Daten sichtbar zu machen eignen sich verschiedene grafische Darstellungen der Daten. So bieten MA-Diagramme, bei denen der Logarithmus des Quotienten von R/G (Referenz/Probe) gegen den Logarithmus des Mittelwerts aufgetragen wird oder eine Höhenliniendarstellung, zwei Möglichkeiten zur Darstellung von systematischen Varianzen.

Werden in MA-Diagrammen die einzelnen Blöcke eines Mikroarrays mit unterschiedlichen Farben dargestellt, wie in der folgenden Abbildung, bekommt man einige Hinweise auf mögliche räumliche Effekte.

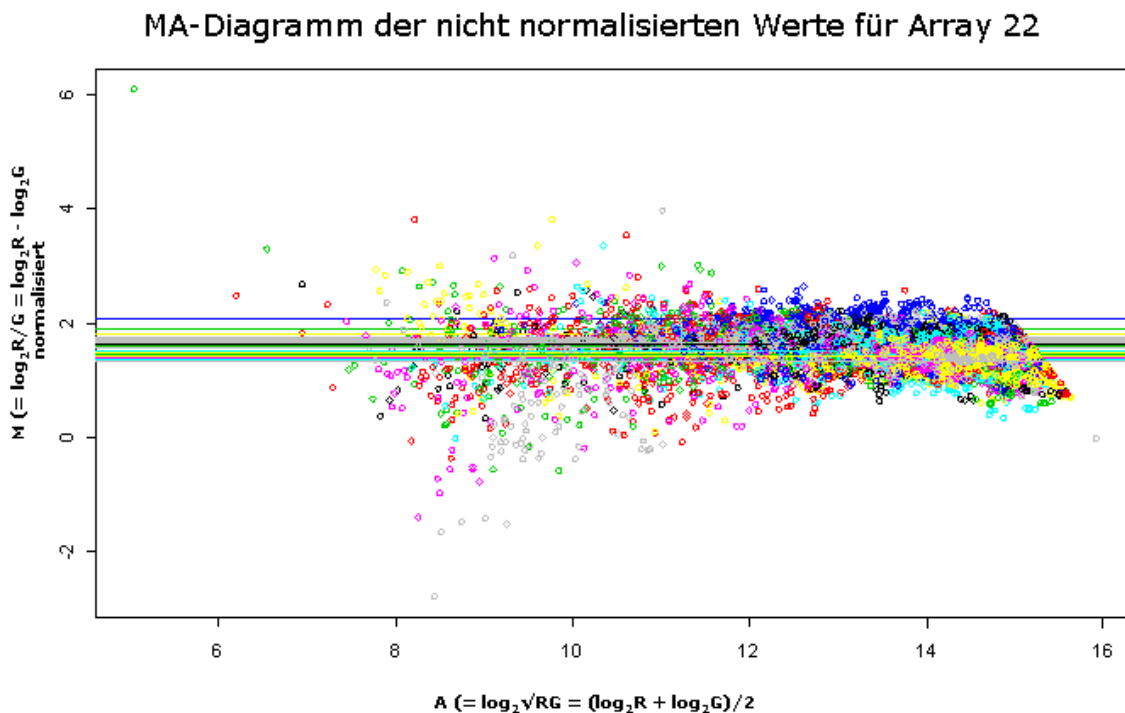


Abb. 38: Ma-Diagramm normalisierter Werte für Array 22. Jede Farbe repräsentiert einen Block. Die Linien entsprechen den Medianen der einzelnen Blöcke.

Weil in den grafischen Darstellungen eindeutig zu sehen ist, dass ein räumlicher Effekt auftritt, ist zur Normalisierung eine Methode zu empfehlen, die eine räumliche Glättung der Werte vornimmt. Einige solcher räumlichen

Glättungsverfahren sind beschrieben und die folgenden sind in Betracht gezogen worden.

- einfache Logarithmustransformation der Intensitätsquotienten
1. räumliche (blockweise) Mediantransformation
  2. räumliche (blockweise) „loess“ Glättung (lokale polynomale Regressionsanpassung)
  3. räumliche Glättung (bezüglich der Koordinaten der Punkte des Arrays)
  4. ANOVA basierende Abstände, mit Block, Reihe und Spalte als Faktoren

Die Methoden 1 bis 3 wurden für die Normalisierung der einzelnen Mikroarrays angewandt, da es alles Methoden sind, die eine lokale Komponente besitzen und mit einem vertretbaren Rechenaufwand durchzuführen sind. Die Ergebnisse dieser Verfahren können mit dem Kommando „normcheck“ in der R Umgebung grafisch dargestellt werden, um ihre Anwendbarkeit zu überprüfen. In der folgenden Abbildung wird, als ein Beispiel, für Block 1 des DNA-Mikroarrays 22, ein MA-Diagramm mit den Ergebnissen der drei verschiedenen Methoden gezeigt.

**MA-Diagramm von Block 1 des Array 22**

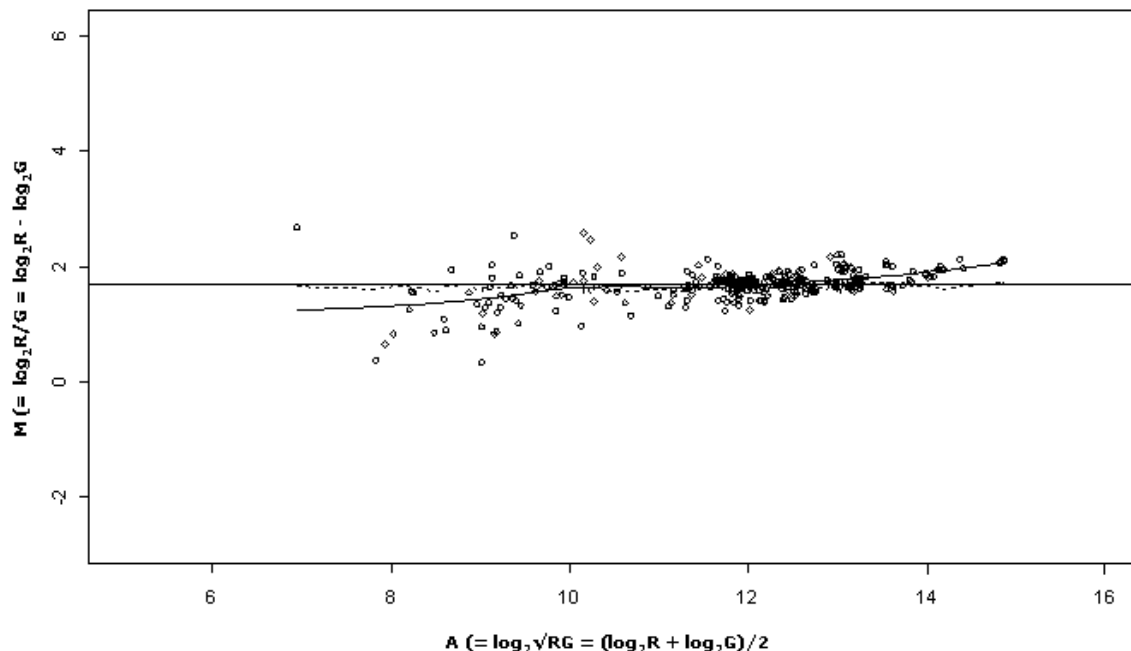


Abb. 39: Ma-Diagramm für Block 1 von Array 22. Es wurden verschiedene Methoden zur Normalisierung der Werte benutzt. Die Methode der räumlichen Mediantransformation wird mit der durchgezogenen Linie dargestellt. Die geschwungene Linie repräsentiert die Methode der blockweisen „loess“ Glättung und die gestrichelte Linie steht für die Methode der räumlichen „loess“ Anpassung („loess“ = lokale polynomale Regressionsanpassung).

Durch die gerade Linie in Abbildung 39 wird das Ergebnis der räumlichen Mediantransformation gezeigt, die geschwungene Linie repräsentiert die blockweise „loess“ Glättung und die gestrichelte Linie gibt das Ergebnis der räumlichen „loess“ Anpassung wieder. Als Ergebnis der Normalisierung erhält man die Abstände bzw. die Differenzen des originalen M-Wertes zu der jeweiligen Kurve, die an Hand der angewandten Methode, durch die Punktwolke gelegt wurden. Die Methode der räumlichen Mediantransformation und die blockweise „loess“ Glättung, liefern annähernd identische Werte, was eine Mediantransformation als ausreichend erscheinen lässt. Des Weiteren scheinen keine blockspezifischen Intensitätsabhängigkeiten der M-Werte vorhanden zu sein und auch keine Unregelmäßigkeiten im unteren Intensitätsbereich. Allerdings muss bei der Verwendung der räumlichen „loess“ Glättung die Gefahr des so genannten „overfittings“ berücksichtigt werden, bei dem die Daten zu sehr normiert werden. D.h. Es besteht durch das overfitting die Gefahr, dass biologisch begründete Expressionsunterschiede eines Gens durch das Normalisierungsverfahren beseitigt werden und somit dieses Gen nicht als ein differentiell exprimiertes Gen detektiert wird.

Das Phänomen des „overfittings“ ist in der Methode des „loess“ selbst begründet, denn bei der „loess“ Glättung, werden Polynome iterierend mittels der Methode lokal gewichteter kleinster Quadrate angepasst. Dabei werden zur Anpassung des Punktes  $x$ , aus seiner Nachbarschaft Punkte herangezogen, die aufgrund der Entfernung zum Punkt  $x$  gewichtet werden. Wie viele Punkte zu dieser Nachbarschaft gehören, wird durch einen Abstandsfaktor („span factor“) bestimmt. Hier wurden die Berechnungen mit dem Standardwert für den Glättungsfaktor von 0,75 ausgeführt.

Tritt der Fall ein, dass in der Umgebung des anzupassenden Gens nur wenige weitere spots vorhanden sind, wie in der Abbildung 26 gezeigt, ist die Gefahr des „overfitting“ (Überinterpretation) offenkundig. Denn in der Nachbarschaft dieses Gens sind nur sehr wenige Punkte vorhanden, die zur Normalisierung benutzt werden können, was dazu führt, dass dieses Gen als Ausreißer eingestuft und damit nach der räumlichen Glättung nicht mehr als mögliches differentiell exprimiertes Gen angesehen wird.

## MA-Diagramm von Block 12 des Array 30

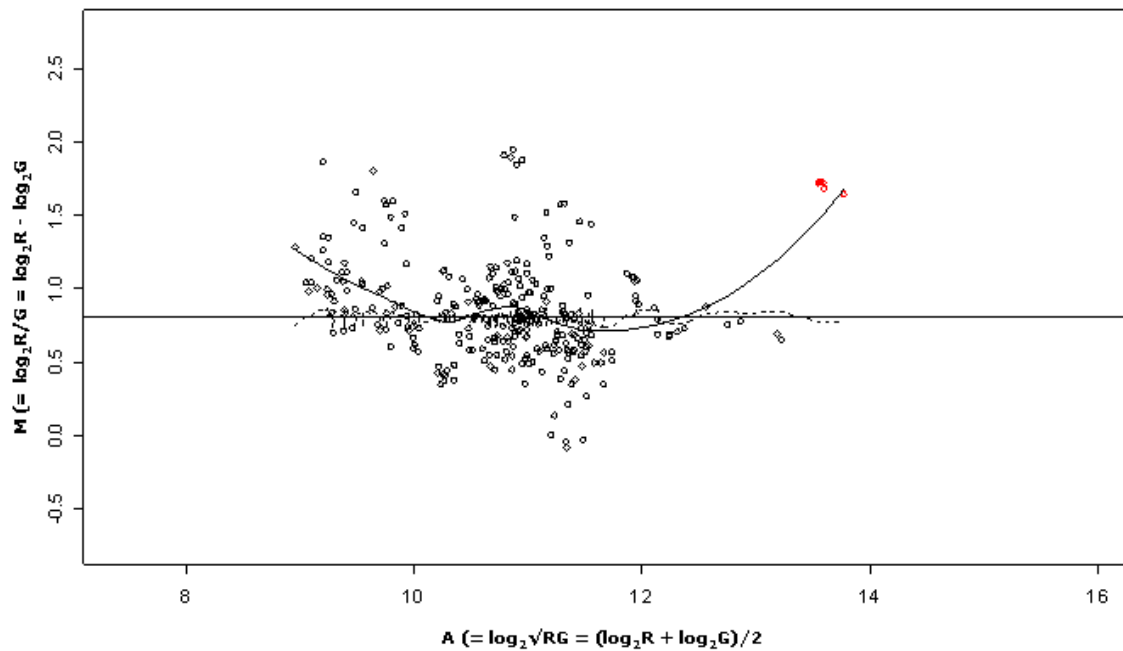


Abb. 40: Ma-Diagramm für Block 12 von Array 30. Es wurden die gleichen Methoden zur Normalisierung angewandt, wie in Abb. 39. Die Werte für das Gen OE3106F (*bop*) sind in rot hervorgehoben. Man kann deutlich sehen, dass bei der Normalisierung die Gefahr eines „overfittings“ der Daten gegeben ist.

Alle grafischen Darstellungen zeigen, dass generell keine Varianzstabilisierung notwendig ist und somit, wie erwähnt, eine räumliche Mediantransformation als ausreichend erachtet wird.

### 3.2.2 Normalisierung der Mikroarrays untereinander

Aufgrund der zufälligen Anordnung von den rund 2700 Genen von *H. sal.* R1 auf dem Mikroarray und der Annahme, dass der größte Teil der Gene nicht differentiell exprimiert wird, ist es möglich eine Normalisierung über alle Gene anzuwenden. So sind z.B. beim ersten Transkriptom-Experiment, dem Vergleich aerob/phototroph, am Ende weniger als 10% der Gene als differentiell exprimiert anzusehen.

Im Verlauf einer Normalisierung der einzelnen Mikroarrays untereinander, findet häufig eine Zentrierung der M-Werte statt. Dies ist der erste Schritt in der Berechnung des Algorithmus der ANOVA Schätzwerte (Methode 4 der oben erwähnten Transformationsmethoden). So repräsentieren die ersten vier Terme des MAANOVA Modells ein Normalisierungsmodell und können wie folgt formuliert werden:

$$Y_{ijkgr} = \mu + \text{Array}_i + \text{Farbstoff}_j + \text{Array} * \text{Farbstoff}_{ij} \quad (2)$$

$Y_{ijkgr}$  steht für die normalisierten Intensitätswerte,  $\mu$  für die durchschnittliche Intensität und die anderen Terme wie  $\text{Array}_i$ ,  $\text{Farbstoff}_j$ , und  $\text{Array} * \text{Farbstoff}_{ij}$  geben die Varianz bezüglich des Arrays, des Farbstoffes und der Wechselwirkung zwischen Array und Farbstoff an. Der Index (i) steht für die Mikroarrays, (j) für den Farbstoff, (g) für die Gene und (r) ist kombiniert mit (i) und (j), um weiterhin bei Replikaten eines Genes, einen speziellen Datenpunkt zu definieren. Die Wiederholungen der Mikroarrays sind durch den Index (k) von einander zu unterscheiden, wobei die Angabe der drei Indices (i,j,k) redundant ist, da bei einem gegebenen Mikroarray (i) und einem Farbstoff (j) nur ein (k) möglich ist. Bei dieser Auswertung wurde den Replikaten der Mikroarrays erstens dadurch Rechnung getragen, dass die Daten in der Ausgangsmatrix in einer bestimmten, den Wiederholungen entsprechenden, Form angeordnet wurden und zweitens durch einen speziellen Vektor, der die Position eines Genes in der Ausgangsdatenmatrix definiert.

Wie oben erwähnt ist die Normalisierung der erste Schritt im Verlauf der Lösung des MAANOVA Modells. Der Rest des Modells und der zweite Schritt der Anpassung an das Modell wird in dem folgenden Abschnitt 4.3 beschrieben.

### **4.3 Statistische Datenanalyse / Identifizierung differentiell exprimierter Gene**

In dieser Phase der statistischen Analyse ist es das Ziel, die differentiell exprimierten Gene zu identifizieren. Das R/MAANOVA Packet bietet einen kompletten Arbeitsablauf zur Mikroarraydatenanalyse. Unter anderem ist ein „mixed model ANOVA“ implementiert, welches die Bestimmung von Varianzen erlaubt und F- bzw. t-Tests zur Identifizierung differentiell exprimierter Gene ausführt. In einem „mixed“ ANOVA Modell wird für einige Terme angenommen, dass das Ergebnis eine Realisierung eines Zufallsverfahrens ist, während hingegen bei einem „fixed“ ANOVA Modell alle Terme als fest angesehen werden.

#### **4.3.1 Vervollständigung des Datensatzes: „missing value imputation“**

Die Algorithmen in MAANOVA dulden keine fehlenden oder negativen Werte und Nullen. Fehlende Werte entstehen bei einem Mikroarray-Experiment z.B. durch technische Probleme beim Drucken. Fehlen Daten für ein Gen, so müssten alle

dazugehörigen vorhandenen Daten von Hand entfernt werden, wodurch eine starke Reduktion der Ausgangsmatrix eintreten würde. Deshalb ist es notwendig, Methoden anzuwenden, mit deren Hilfe fehlende Daten in passender Weise ersetzt werden können.

Folgende drei Methoden zur Vervollständigung von Datensätzen sind möglich:

1. Reihenmittelwert, d.h. der fehlende Wert wird durch den Mittelwert der verbleibenden Expressionsdaten eines Gens ersetzt.
2. k nächste Nachbarn („k nearest neighbor“), d.h. es werden k Gene mit einem ähnlichen Expressionsverhalten genommen, um aus diesen einen Mittelwert zu berechnen, der den fehlenden Wert ersetzt.
3. Hauptkomponentenanalyse („single value decomposition“ - SVD) um den fehlenden Wert an die vorhandenen anzugleichen.

Die Anwendbarkeit der verschiedenen Methoden hängt stark von der Struktur der unvollständigen Daten ab. Deshalb wurde basierend auf der Datenmatrix der M-Werte mit den technischen Replikaten, die fehlenden Werte nach einem mehrstufigen Verfahren ersetzt.

1. Ausschluss der Gene mit einem Prozentsatz an fehlenden Werten, der höher ist als eine gegebene Konstante, die in Bezug zu allen Wiederholungen und Mikroarrays steht.

⇒ Berechnung der Mediane aus allen Wiederholungen auf den verschiedenen Mikroarrays nur für die Gene, die nicht mehr als 33% fehlende Werte haben. Für das erste Experiment, in dem die Wachstumsbedingungen aerob und phototroph miteinander verglichen wurden, sind fünf Wiederholungen für jede Bedingung vorhanden und auf den Mikroarray-Chips fünf technische Replikate für jedes Gen. Das bedeutet, dass Gene mit mehr als 17 fehlenden Werten von den 50 möglichen (33%), von der Analyse ausgeschlossen werden. Für das erste Experiment sind dies nur 72 Gene, was 2,7% des gesamten Genbestandes entspricht.

2. Vervollständigung der fehlenden Daten in der Medianmatrix durch einen Algorithmus basierend auf einer Hauptkomponentenanalyse.

⇒ Jeder fehlende Wert in der Originaldatei wurde durch eine Zufallszahl ersetzt, die auf einer Normalverteilung der Mittelwerte beruht. Die Matrix

wurde berechnet aus den vorhergehenden Schritten und den gemittelten Standardabweichungen des entsprechenden Gens über alle Mikroarrays.

#### **4.3.2 Berechnung des MAANOVA Modells**

Eine Varianzanalyse ist eine Technik zur Analyse von Experimenten, bei denen kontinuierliche Zielgrößen unter verschiedenen experimentellen Bedingungen gemessen werden. Die Anwendung von Varianzanalysen auf Mikroarraydaten ist in der Literatur an verschiedenen Stellen vorgeschlagen, beschrieben und diskutiert worden (Kerr MK, et al., 2000, 2002). Ein ANOVA Modell wird auf transformierte, also vorverarbeitete, Intensitätsdaten angewandt und ermöglicht es, die Quelle von Abweichungen zu bestimmen, die nicht ihren Ursprung in der differentiellen Expression haben. Diese Faktoren müssen z.B. durch eine Normalisierung herausgefiltert werden. Die Normalisierung ist der erste Schritt des Algorithmus der Anpassung an das ANOVA Modell.

Das vollständige Varianzanalysemodell für Mikroarrays wird in der folgenden Gleichung beschrieben:

$$Y_{ijkgr} = \mu + \text{Array}_i + \text{Farbstoff}_j + \text{Array} * \text{Farbstoff}_{ij} + G_g + Vg_{kg} + DG_{jg} + AG_{ig} + S_{r(ig)} + \varepsilon_{ijkgr} \quad (3)$$

Die ersten vier Terme der Gleichung sind aus dem Normalisierungsmodell Gleichung (2) bekannt. Der zusätzliche Gen-Term  $G_g$  bezieht sich auf die durchschnittliche Intensität eines einzelnen Gens. Der (Unterschied-bei-Gen) Term  $VG_{kg}$  des relativen Expressionswertes, gibt das Expressionsniveau eines Gens ( $g$ ) einer Wiederholung ( $k$ ) relativ zur gewichteten durchschnittlichen Expression des Gens über alle Wiederholungen des Transkriptionsversuches an. Der (Farbstoff-bei-Gen) Term  $DG_{jg}$  steht für die verschiedenen Gen abhängigen Einbaueffizienzen der Fluoreszenzfarbstoffe. Die Variation (Verschiedenheit) eines „spots“ auf unterschiedlichen Mikroarrays wird durch den (Array-bei-Gen) Term  $AG_{ig}$  repräsentiert und die Heterogenität zwischen den Replikaten der Gensonden, wird in dem Term  $S_{r(ig)}$  zusammengefasst. Bei der Berechnung der individuellen Terme des ANOVA Modells geht man vom Prinzip der kleinsten Fehlerquadrate aus, unter Annahme der Normalverteilung des Fehlers. Dieser Fehler ist in Gleichung (3) durch  $\varepsilon_{ijkgr}$  berücksichtigt. Die Nomenklatur der Indices ist die selbe, wie für Gleichung (2).

Der Vorteil eines ANOVA Modells liegt darin begründet, dass es erlaubt, die Informationen von mehreren Mikroarray-Hybridisierungen in allen erdenklichen Kombinationen in einem Evaluierungsprozess zusammenzufassen. Die

herkömmliche Vorgehensweise zur Anpassung der Daten an ein ANOVA Modell, mittels der kleinsten Fehlerquadrate, beinhaltet die Berechnung des Inversen der Designmatrix. Da diese Designmatrix für Mikroarrayexperimente sehr groß sein kann, findet die Anpassung unter MAANOVA in zwei Schritten statt. Das Normalisierungsmodell wird zuerst angepasst, gefolgt von den weiteren Termen, die in einem iterativen Gen-für-Gen Verfahren angepasst werden.

Die Parameter, die man im Anschluss an eine Anpassung erhält, sind Schätzwerte und unterliegen somit einem gewissen Fehler. Deshalb müssen zur statistischen Schlussfolgerung zwei Modelle aufgestellt werden, die durch Hypothesentests beurteilt werden.

Im MAANOVA Paket sind mehrere Hypothesentests (genannt F1, F2, F3, FS) enthalten:

F1 ist eine normale F-Statistik, die dann berechnet würde, wenn man nur Daten für ein Gen zur Verfügung hätte.

F3 ist eine Methode, die für alle Gene eine gleiche Fehlervariation annimmt.

F2 stellt ein Hybrid aus F1 und F3 dar und benutzt eine gewichtete Kombination einer globalen und genspezifischen Varianz.

FS benutzt einen Shrinkage-Schätzer für die genspezifische Varianz basierend auf einem James-Stein-Schätzer (Cui XQ und Churchill GA, 2003; Cui XQ, et al., 2003)

Routinemäßig werden bei der Analyse alle drei Methoden angewandt, weil sie jeweils unterschiedliche Aspekte der Daten berücksichtigen. Alle Tests basieren auf dem genspezifischen Rest der Summe der Abweichungsquadrate und dem Rest der Freiheitsgrade.

Die Anwendung der oben erwähnten Teststatistiken liefert eine statistische Signifikanz, einen so genannten p-Wert. Der p-Wert vom Test auf eine differentielle Expression, fasst die statistische Signifikanz des Tests zusammen, der auf der Variation in der Genexpression und seiner Fehlerschwankung basiert. Technisch gesprochen ist der p-Wert ein abnehmender Index der Zuverlässigkeit eines Ergebnisses. Je größer der p-Wert ist, desto weniger ist dem beobachteten Zusammenhang zu trauen. Dabei ist jedoch zu bedenken, dass eine große statistische Signifikanz nicht unmittelbar auch eine große biologische Relevanz



beinhaltet. Außerdem muss ein Signifikanzniveau festgelegt werden, um ein Ergebnis als relevant zu bestimmen. Ein solches Signifikanzniveau gibt die Genauigkeit des Größenunterschiedes der relativen Expression an.

Eine häufig verwandte Methode zur Kontrolle des Fehlers beim multiplen Testen ist die „family-wise error rate“. Die „family-wise error rate“ (FWER) ist die Wahrscheinlichkeit wenigstens einen Fehler 1. Art zu begehen. Das bedeutet die Ausgangshypothese  $H_0$  wird abgelehnt, obwohl sie richtig ist. Die Methode der FWER ist sehr stringent, was dazu führt, dass nur sehr wenige Kandidaten ausgewählt werden und deshalb für die Anwendung auf Mikroarrays nicht als optimal anzusehen ist. Denn das Ziel einer Transkriptom-Analyse ist es, eine Liste von differentiell exprimierten Genen aufzustellen, in der wenige falsch positive Gene zu tolerieren sind. Man akzeptiert also eher Gene, die fälschlicherweise als differentiell exprimiert angesehen wurden, als Gene von vornherein nicht zu finden, die differentiell exprimiert werden.

Eine Alternative zur FWER stellt die Methode der „false discovery rate“ (FDR) da (referenz). Die FDR ist definiert als der erwartete Anteil von Fehlern 1. Art der abgelehnten Null-Hypothesen.

### **4.4 Visualisierung der Daten**

Die Kommandozeilensoftware R besitzt den großen Vorteil der vielseitigen Einsatz- und Anwendbarkeit. Gleichzeitig allerdings beinhaltet diese universelle Verwendbarkeit eine relativ große Komplexität, die einer einfachen Handhabbarkeit entgegensteht. Zur Beseitigung dieses Nachteils von R, wurden im Rahmen dieser Transkriptom-Analyse zur Visualisierung der Daten Kommandofolgen implementiert, die es ermöglichen, die zur Interpretation während des Analyseprozesses erforderlichen Abbildungen zu generieren. Die im Abschnitt 4 zur Verdeutlichung der statistischen Datenanalyse eingesetzten Abbildungen, wurden mit diesen implementierten Funktionen generiert.

Im Rahmen einer exploratorischen Datenanalyse, bei der es das Ziel ist, durch eine weiterführende statistische Datenanalyse, die in den Daten enthaltenen Muster und Auffälligkeiten zu identifizieren, findet ebenfalls eine Visualisierung dieser Analyseergebnisse statt. Eine Möglichkeit einer exploratorischen Datenanalyse ist die Anwendung einer Hauptkomponentenanalyse (Principal Component Analysis; PCA), die auch mit den *H. sal.* R1 Transkriptom-Daten durchgeführt wurde. Da die Durchführung einer PCA einen ausgewerteten Datensatz voraussetzt und in diesem Kapitel der generelle Weg der statistischen

Auswertung beschrieben wurde, werden die Ergebnisse der Hauptkomponentenanalyse unter Kapitel 5.3.1 dargestellt.

### **4.5 Diskussion**

In dem zweiten großen Teil der Ergebnisse wurde die Datenanalyse der Transkriptomexperimente beschreiben. Bei der Wahl der Datenanalyse wurde auf ein sehr flexibles und leistungsstarkes System gesetzt. All diese Voraussetzungen erfüllt eine Auswertung der Expressionsdaten mit dem statistischen Auswertungssystem R (siehe 3.4.5).

Um die große Flexibilität und Leistungsstärke der statistischen Auswertung mit R in vollem Umfang nutzen zu können, wurde auf die Erfahrung und Expertise von Dr. G. Welzl (GSF-Forschungszentrum für Umwelt und Gesundheit, GmbH) und Dr. J. Landgrebe (Universität Göttingen) gesetzt (Landgrebe J, et al., 2002).

Eine nähere Ausführung zu R befindet sich in Kapitel 11.6.1.

Die Wahl des Auswertungssystems ist ein nicht unerheblicher Punkt, denn bei einem sorgfältig geplanten Transkriptomexperiment fällt eine sehr große Datenmenge an.

Betrachtet man das erste Transkriptomexperiment, den Vergleich der aeroben und phototrophen Wachstumsbedingung von *H. sal.* R1, bei dem die Hybridisierungen je Bedingung fünfmal wiederholt wurden, so sind hier für 153.600 Datenpunkte, der 10 Hybridisierungen, die statistischen Kenngrößen der Vorder- und Hintergrundwerte erhoben worden. Für das zweite Zeitreihenexperiment steigt diese Komplexität des Datensatzes noch weiter an, denn für diesen Versuch wurde die statistische Auswertung für eine Gesamtzahl von 384.000 Datenpunkten berechnet.

Aus diesen Gründen weist das R/MAANOVA Packet in der interaktiven Umgebung von R alle Voraussetzungen für eine statistische Datenanalyse von Mikroarrayexperimenten auf. Doch nicht nur die Fähigkeit der Handhabung solch großer Datenmengen, auch die Möglichkeit der Nachvollziehbarkeit und Dokumentation jeden Schritts der Auswertung machen dieses System zu einem hervorragendem Hilfsmittel der Transkriptomdatenauswertung.

Eine Beschreibung der einzelnen Schritte erfolgte bereits bei der Vorstellung der Ergebnisse, deshalb dient dieses Kapitel dazu die Komplexität und Verschachtelung der Auswertung näher zu beschreiben.

Wie unter 4.0 beschrieben und in Abb.29 dargestellt, basiert die Auswertung auf verschiedenen Arbeitsschritten. Zur Erinnerung erfolgt eine Zusammenfassung der Arbeitsschritte (siehe Tab.4):

Analyseschritt	Beschreibung
Vorverarbeitung der Daten	<ul style="list-style-type: none"> <li>• Ausschluss von "spots" mit schlechter Qualität</li> <li>• Hintergrundkorrektur der Daten</li> <li>• Berechnung der Ausschlussgrenze für "spots" im unteren Intensitätsbereich (diese sind vom Hintergrundsignal nicht zu unterscheiden)</li> <li>• Logarithmustransformation der Intensitätswerte</li> </ul>
Normalisierung der Daten	<ul style="list-style-type: none"> <li>• Anwendung geeigneter grafischer Darstellungen der Daten zur Datenvalidierung</li> <li>• Anwendung räumlicher "loess" Glättung (=lokale polynomale Regressionsanpassung)</li> </ul>
Statistische Datenanalyse und Visualisierung	<ul style="list-style-type: none"> <li>• Vervollständigung der Datensätze "missing value imputation"</li> <li>• Berechnung des MAANOVA-Modells mit Hilfe verschiedenen Hypothesentests = Berechnung der signifikanten statistischen Rangfolge der Gene</li> <li>• Bestimmung eines Ausschlusswertes zur Ablehnung der Nullhypothese</li> <li>• Anwendung einer Hauptkomponentenanalyse (PCA) zur grafischen Veranschaulichung der signifikanten Unterscheide in des Datensatzes.</li> </ul>

Tab. 4: Zusammenfassung der Arbeitsschritte der statistischen Datenanalyse

Beim Punkt der Datenvorverarbeitung ist hervorzuheben, dass die Festlegung des oberen Grenzbereichs der Intensitäten durch den Schwellenwert der Sättigung des Fluoreszenzsignals bei 65000 ( $2^{16}$ ) Einheiten gegeben ist und nicht verändert werden kann. Die Bestimmung des unteren Intensitätsbereich der Fluoreszenzsignale erfolgte jedoch nicht durch eine willkürliche Festlegung eines Schwellenwertes. Denn die Grenze der unteren detektierbaren Signale wurde durch die so genannte "minimum volumn ellipsoid" Methode berechnet.

Diese Methode kalkuliert auf der Basis der Mittelwerte und Standardabweichungen der Vorder- und Hintergrundwerte jedes Blocks auf dem Mikroarray-Chip eine Ellipse. Dabei wird die Ellipse so erstellt, dass in deren Mittelpunkt diejenigen Punkte liegen, deren Vordergrundwerte nicht vom Hintergrundsignal zu unterscheiden sind und für eine weitere Auswertung nicht benötigt werden (siehe Abb.35). Der Vorteil einer solchen dynamischen Festlegung des unteren Bereichs liegt darin, dass es keine starre scharfe Grenze gibt wie im oberen Bereich. Dadurch wird die Gefahr erheblich herabgesetzt signifikante Signale im niedrigen Intensitätsbereich von der Datenanalyse von vornherein ausgeschlossen zu haben. Dies ist im Hinblick auf den kleineren dynamischen Bereich von Mikroarrays ein besonders kritischer Punkt. So ist in der Literatur beschrieben, dass bei Mikroarrayexperimenten die Induktionsraten der Gene um den Faktor zwei bis zehn niedriger liegen können, als im Vergleich zu RT-PCR Experimenten (Conway T und Schoolnik GK, 2003). Die Raten dieser Methoden liegen bei 10-1000, während man für Mikroarrayexperimente häufig eine Induktion von <10 findet.

Da auch in allen Transkriptomexperimenten dieses Projekts die Induktion unter dem Faktor 10 liegen, stellt die erwähnte Methode der "minimum volumn ellipsoide" eine hervorragende Möglichkeit dar, um im niedrigen Intensitätsbereich möglichst keine signifikanten Signale zu verlieren.

Durch die Festlegung der unteren Grenze des Hintergrundniveaus der niedrigen Signale erfolgte eine Schärfung in diesem Intensitätsbereichs, so dass die vorher auftretende Hockeyschläger-Form in der Auftragung der Signale nicht mehr zu beobachten ist. Was eine Voraussetzung für die weiteren Korrekturmethode war (siehe 3.1.1)

Nach der Hintergrundkorrektur wurden durch die Normalisierung alle weiteren experimentell bedingten Fehler beseitigt. Der Ablauf der Normalisierung wurde in zwei Teile gegliedert, wobei im ersten Schritt die einzelnen Mikroarray-Chips normalisiert wurden, um dann im zweiten Abschnitt die Mikroarray-Chips untereinander zu normalisieren.

Bei der Normalisierung der einzelnen Mikroarray-Chips ist bei der Wahl der Methode die Gefahr des so genannten "overfittings" gegeben. Hierbei würden durch die angewandte Normalisierungsmethode die Expressionsunterschiede eines Gens nicht als biologisch begründet betrachtet und nivelliert werden (siehe

Abb.41). Durch diese Überinterpretation würden auf diese Weise wichtige Ergebnisse nicht erkannt.

Der Abgleich der Mikroarrays untereinander wurde durch den zweiten Normalisierungsabschnitt erreicht. Als notwendige Voraussetzung zur Durchführung der Normalisierung wurde davon ausgegangen, dass die Mehrzahl der Gene nicht differentiell exprimiert wird.

Das aufgestellte Normalisierungsmodell entspricht den ersten vier Termen des MAANOVA-Modells und beinhaltet, die Varianzen bezüglich der Arrays, der Farbstoffe und die Wechselwirkungen zwischen Array und Farbstoff, sowie die durchschnittliche Intensität (siehe Gleichung 2 in 4.2.2).

Somit stellt die Normalisierung der Daten auch gleichzeitig den Beginn der statistischen Auswertung der Datenmatrix da.

Bevor mit der Berechnung des MAANOVA-Modells begonnen wird, muss ein weiterer Punkt der Datenauswertung durchgeführt werden. Dieser Punkt ist die Vervollständigung der Datenmatrix und heißt im englischen „missing value imputation“. Die große Datenmatrix der vorverarbeiteten Expressionsdaten enthält üblicherweise fehlende Werte. Diese Lücken in den Datensätzen müssen für die nachfolgende Berechnung des MAANOVA-Modells geschlossen werden, da für die Anwendung der notwendigen Algorithmen ein vollständiger Datensatz eine wesentliche Voraussetzung ist. In der Literatur findet man einige Beispiele und Vorschläge für Methoden der „missing value imputation“ (Troyanskaya O, et al., 2001; Donders AR, et al., 2006). Wichtig ist bei diesem Prozess, dass fehlende Werte in einer moderaten Art ersetzt werden. Aus diesem Grund wurden für diese Arbeit der unter (4.3.1) beschriebene Ablauf aufgestellt und eingehalten. Eine wesentliche Vorgabe war dabei, dass die Gene von der Auswertung ausgeschlossen wurden, die mehr als 33% fehlende Werte besaßen. Die Vervollständigung erfolgte mit Hilfe eines Algorithmus basierend auf einer Hauptkomponentenanalyse (Liu L, et al., 2003).

Da dieser Punkt der statistischen Auswertung einen gewissen Einfluss auf die Ergebnisse der Datenanalyse ausüben kann, zeigt dieser Aspekt der Auswertung bei Statistikern ein großes Interesse und unterliegt damit weiterhin intensiven Weiterentwicklungen (Scheel I, et al., 2005; Donders AR, et al., 2006; Johansson P und Hakkinen J, 2006).

Nachdem der Datensatz vorverarbeitet und vervollständigt wurde, konnte das MAANOVA-Modell berechnet werden. Die Gleichung (3) in Kapitel 4.3.2) zeigt das gesamte Varianzanalysenmodell zur Mikroarrayauswertung. Die Berechnung des Modells erfolgte in zwei Schritten. Als erstes wurde das Normalisierungsmodell angepasst und im zweiten Schritt in einem iterativen Gen-für-Gen Verfahren die weiteren Terme der Gleichung berechnet.

Die resultierenden Parameter der Anpassung an das Modell sind Schätzwerte und als solche unterliegen sie einem gewissen Fehler. Zur Beurteilung dieser Parameter sind verschiedene Hypothesentests im MAANOVA-Modell enthalten. Diese Teststatistiken liefern so genannte p-Werte, die die Signifikanz eines Ergebnisses widerspiegeln. Der p-Wert ist damit ein Instrument die Zuverlässigkeit eines statistischen Ergebnisses zu interpretieren. Auf der Basis der so erhaltenen p-Werte wurde beginnend mit dem kleinsten Wert eine Rangliste von Genen erstellt, die eine signifikant differentielle Expression aufweisen.

Bei der Interpretation dieser Ranglisten ist zu bedenken, dass eine statistische Signifikanz nicht unmittelbar auch eine biologische Relevanz beinhaltet. Deshalb ist es wichtig ein Signifikanzniveau festzulegen, um die Genauigkeit des Ergebnisses abschätzen zu können. Bezogen auf diese Arbeit bedeutet ein solches Signifikanzniveau, wie verlässlich bzw. wahrscheinlich der detektierte Expressionsunterschied ist.

Durch eine Kontrolle des Fehlers, der beim multiplen Testen während der statistischen Analyse auftritt, kann das Signifikanzniveau einer Auswertung festgelegt werden. Wobei die Kontrolle ihrerseits stringent oder weniger stringent erfolgen kann. Dies bedeutet, dass wenn man eine stringente Methode anwendet, wie z.B. die „family-wise error rate“ (FWER), nur eine kleine Anzahl von differentiell exprimierten Genen ausgewählt werden. Bei der FWER wird die Wahrscheinlichkeit wenigstens einen Fehler 1. Art begangen zu haben bestimmt. Also die Ausgangshypothese  $H_0$  abzulehnen, obwohl sie richtig ist. Man hätte zwar eine Liste mit sehr signifikanten Kandidatengen, die allerdings auf Grund ihrer stringenten Erstellung für eine biologische Interpretation zu wenig umfangreich ist.

Bei einem Ansatz mit geringerer Stringenz, könnte man als Ergebnis der Auswertung eine Kandidatengenliste akzeptieren, bei der ein gewisser Anteil der Gene falsch positiv ist. Es werden demnach eher Kandidatengene toleriert, die fälschlicherweise als differentiell exprimiert angesehen werden, als von

vornherein differentiell exprimierte Gene auszuschließen. Als Beispiel einer weniger stringenten Festlegung des Signifikanzniveaus, kann die Methode der „false discovery rate“ (FDR) betrachtet werden, bei der der Anteil der fälscherweise ausgewählten Gene kontrolliert wird. Dies bedeutet, bei der FDR definiert man den Anteil der erwarteten Fehler der 1. Art. Dieser Anteil kann von 5-20% variieren. Mit anderen Worten, es werden 5 bis 20 falsch positive Gene von einer 100 Gene umfassenden Liste akzeptiert. Die Wahl des FDR-Wertes ist eine individuelle Entscheidung des Experimentators, da kein einheitliches Verfahren zur Festlegung der FDR-Rate existiert.

Auch bei der Auswahl der verschiedenen Methoden des Hypothesentests, mit deren Hilfe die Kandidatengenenlisten erstellt wurden, gibt es keine eindeutig festlegbaren Kriterien. Das MAANOVA-Paket in R umfasste zum Zeitpunkt der Erstellung dieses Manuskripts die Hypothesentests F1, F2, F3 und FS (siehe 4.3.2). Routinemäßig wurden die Hypothesentests F2, F3 und FS für alle Experimente berechnet. Die daraus abgeleiteten Listen der p-Werte, dienen zur Festlegung der Kandidatengene, die als differentiell exprimiert gelten. Das Signifikanzniveau dieser Listen wurde dabei auf der Basis der FDR bestimmt. Welche Kriterien zur Festlegung der Listen der differentiell exprimierten Gene gewählt wurden, ist der Diskussion der folgenden Kapitel zu entnehmen. Denn je nach Ansatz der Interpretation der Expressionsdaten, mussten unterschiedliche Aspekte berücksichtigt werden, die Einfluss auf die jeweilige Auswahl besaßen.

## **5. Differentiell exprimierte Gene**

Nach der Untersuchung der Datenqualität und Transformation der Daten, sind die intensitätsabhängigen und räumlichen Schwankungen entfernt worden, um mit der Anpassung an das MAANOVA-Modell die signifikant differentiell exprimierten Gene identifizieren zu können. Im vorhergehenden Kapitel wurden diese Schritte der statistischen Analysestrategie der Transkriptom-Daten näher beschrieben.

Im Verlauf dieser Arbeit wurden zur Genexpressionsanalyse von *H. sal.* R1 unterschiedliche Experimente durchgeführt. Zur Veranschaulichung der verschiedenen Transkriptom-Experimente, sind diese in dem nachfolgenden Flussdiagramm graphisch zusammengefasst worden.

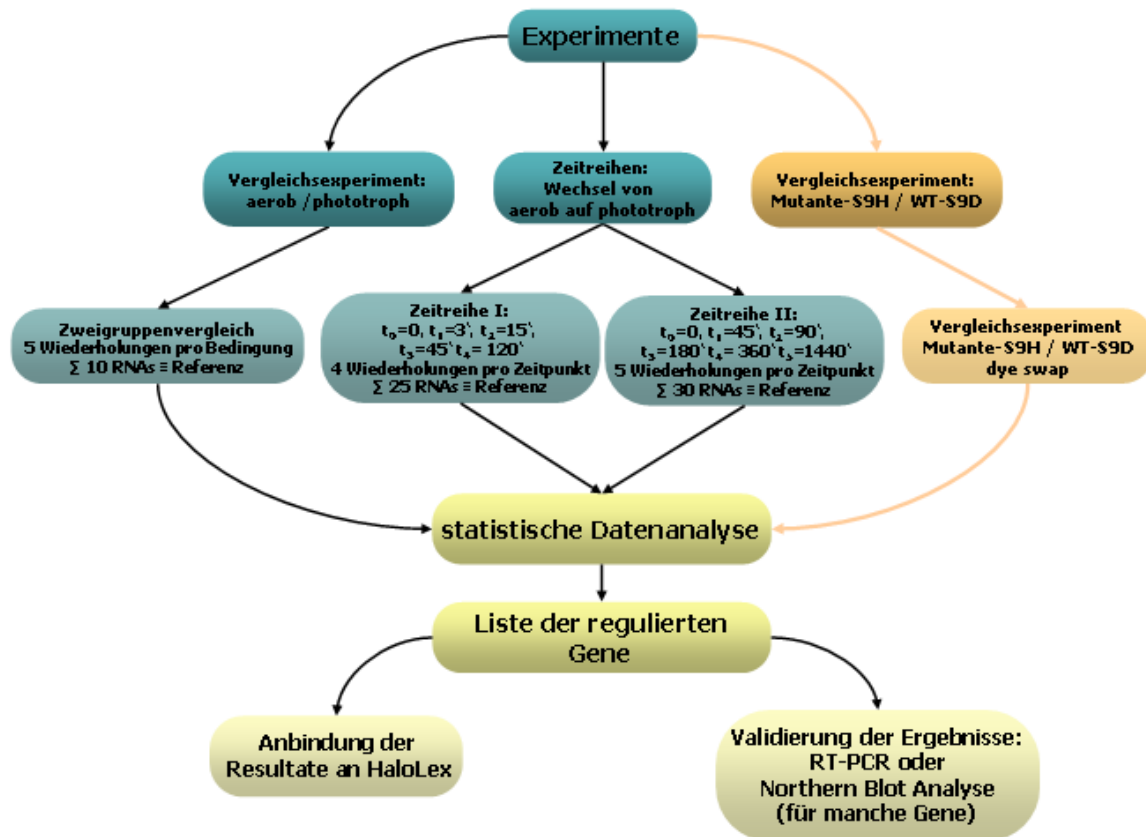


Abb. 42: Flussdiagramm der durchgeführten Transkriptom-Experimente mit den *H. sal.* R1 DNA-Mikroarrays.

In dem Diagramm sind die einzelnen Experimente und weiteren Ebenen in verschiedenen Farben dargestellt worden. Auf der Ebene der Experimente gab es einen Hauptunterschied. Die blau gezeigten Versuche wurden nach dem common reference Experimentdesign mit Zellen von *H. sal.* R1 durchgeführt, während die in orange gehalten Versuche zur Expressionanalyse der S9H-Mutante auf dem dye swap (Farbwechsel) Experimentdesign basieren. Desweiteren beziehen sich die zwei Farben nicht nur auf die unterschiedlichen Experimentdesigns und die verwendeten Zellen, sondern auch auf die verschiedenen Datenanalysestrategien. So wurden die in blau dargestellten Transkriptom-Experimente von *H. sal.* R1 mit Hilfe von R und dem MAANOVA-Paket analysiert, während die Expressionsanalyse der dye swap Experimente unter Anwendung von Excel durchgeführt wurden. Da zu dieser Analyse eine andere Methode und andere Stämme verwandt wurden, wie für den Hauptteil der Transkriptom-Experimente, werden die Ergebnisse dieser Versuche und Details zur Entstehung der Mutante S9H in Kapitel 5.4.3 gezeigt. Die dye swap Experimente wurden zur



vollständigen Aufzählung der durchgeführten Hybridisierungen aber in die Abb. 42 mit aufgenommen.

Die in gelb dargestellte Ebene der statistischen Auswertung und Generierung der Liste mit den differentiell exprimierten Genen, der Anbindung dieser Resultate an die HaloLex-Datenbank und die Validierung der Mikroarray-Ergebnisse mittels RT-PCR, sind Bestandteil dieses Kapitels fünf. In den nächsten Abschnitten sollen die Ergebnisse dieser Datenauswertung dargestellt werden, d.h. es soll die Frage nach den signifikantesten Unterschieden in den durchgeführten Experimenten beantwortet werden. Die Festlegung der Anzahl der differentiell exprimierten Gene ist für diese Arbeit so erfolgt, dass für alle im Folgenden beschriebenen Transkriptom-Experimente eine Anzahl von 242 Genen als differentiell exprimiert angesehen wurde. Für die einzelnen Transkriptom-Experimente liefert diese Auswahl einen jeweils anderen FDR-Wert (false discovery rate), aber da die Auswahl aufgrund eines bestimmten FDR-Wertes ebenfalls eine individuelle Entscheidung darstellt, ist für dieses Projekt entschieden worden, für alle Experimente die gleiche Anzahl von Genen auszuwählen. Dazu muss erwähnt werden, dass bei dieser Auswahl für jedes Experiment die Rate der falsch positiven Gene im statistisch abgesicherten Bereich liegt.

Vorher muss noch angemerkt werden, dass die Resultate von Mikroarray-Experimenten hier als qualitative Aussage einer Expressionsanalyse angesehen werden. Die Aufgabe der Expressionsanalyse mittels DNA-Mikroarrays ist die Erstellung einer Liste mit differentiell exprimierten Genen, in der möglichst wenig falsch positive Kandidatengene enthalten sind. Möglichst wenig falsch positive Kandidaten ist eine Abwägung von statistischer Signifikanz und biologischer Relevanz, die für jedes Projekt neu festgelegt werden muss.

### **5.1 Anbindung an die HaloLex-Datenbank**

Die gesamten Ergebnisse der statistischen Berechnung unter R, sind an die HaloLex-Datenbank angebunden ([http://:www.halolex.de](http://www.halolex.de)) worden.

Dort findet man unter dem Link (Verknüpfung) „Transcriptomic tools“ eine kurze englische Einleitung über die Versuchsdurchführung, sowie Angaben und Informationen welche Transkriptom-Experimente durchgeführt worden sind.

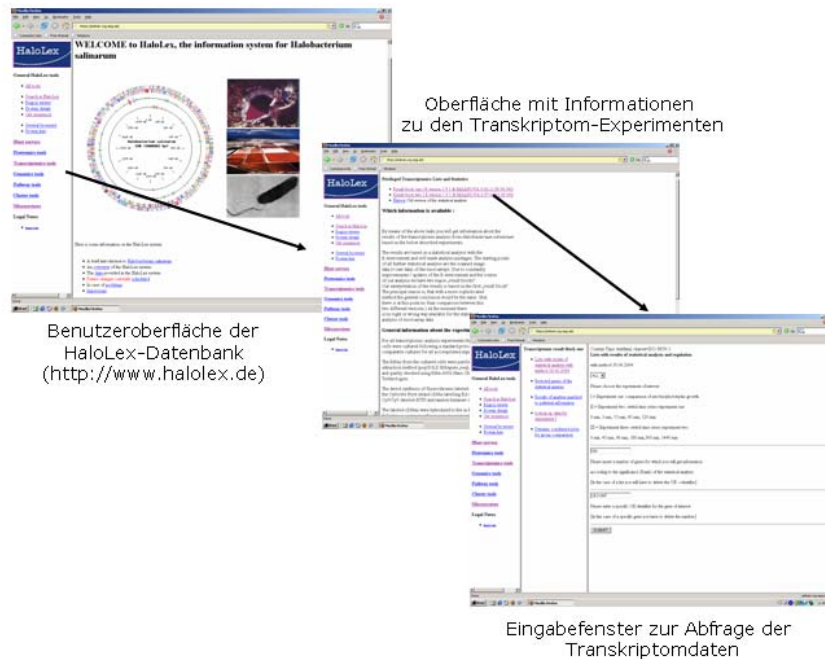


Abb. 43: Screenshots von der Benutzeroberfläche der HaloLex-Datenbank. Gezeigt wird unter welcher Verknüpfung die Transkriptom-Daten dieses Projektes hinterlegt sind.

Nach dieser Einleitung kommt man zu einem Eingabefenster, in welchem man entscheiden kann nach welcher Möglichkeit man die Daten abrufen möchte.

Bei der ersten Abfrage ist es möglich sich eine bestimmte Anzahl von regulierten Genen anzeigen zu lassen. Dabei erfolgt die Ausgabe der Gene analog der statistischen Signifikanz der angewandten Teststatistik zur Lösung des MAANOVA-Modells, wobei zu beachten ist, dass je größer der p-Wert ist, desto geringer ist die Signifikanz bzw. Zuverlässigkeit des Ergebnisses (siehe 4.3.2).

Die zweite Möglichkeit ist, die direkte Abfrage eines Gens, die aus Gründen der Eindeutigkeit der Einträge, mit der exakten Angabe des OE-Identifizierungsmerkmals erfolgen muss. Des weiteren sind die Transkriptom-Daten des Zweigruppenvergleichs aerob/phototroph kombiniert worden mit den Daten für die Stoffwechselwege von *H. sal.* R1. Da diese Art der Abfrage der Transkriptom-Daten zu dem Abschnitt der Interpretation gehört, werden unter 5.4.2 Beispiele dieser Abbildungen der Transkriptom-Daten auf die Stoffwechselwege gezeigt.

Die Listen mit den differentiell exprimierten Genen, die in den nachfolgenden Abschnitten näher beschrieben werden, basieren auf dem Hypothesentest der FS-Statistik. Die Entscheidung, die differentiell exprimierten Gene nach diesem Hypothesentest auszuwählen beruht darauf, dass diese Methode einen Shrinkage-Schätzer zur Bestimmung der genspezifischen Varianz benutzt, der

sich wiederum auf einen James-Stein-Schätzer bezieht. Dieses Verfahren liefert sehr zuverlässige Schätzwerte und wird deshalb z.B. auch in der Wirtschaft zur Renditenschätzung und Portfolio Optimierung von Aktienfonds eingesetzt.

Zum Zeitpunkt der Beendigung dieses Manuskripts, sind diese Abfragemöglichkeiten noch nicht mit der öffentlich zugänglichen Version der HaloLex-Datenbank möglich.

### **5.2      Differentiell      exprimierte      Gene      des Zweiggruppenvergleichs**

Wie unter 5.1 erwähnt, sind die Ergebnisse des Zweigruppenvergleichs (aerobes / phototrophes Wachstum) in der HaloLex-Datenbank unter der Verknüpfung „transcriptomic tools“ zu finden. Die vollständige Liste ist dort hinterlegt, deshalb enthält die folgende Tabelle nur die ersten 50 signifikantesten differentiell exprimierten Gene. Als signifikant differentiell exprimierte Gene des Zweigruppenvergleichs werden 242 Gene angesehen.

Die Anzahl von 242 Genen entspricht ca. 9% des Genoms, wenn man eine Gesamtanzahl von 2784 Proteinen für *H. sal.* R1 festlegt. Somit ist die zur durchgeführten globalen Normalisierung der Transkriptomdaten notwendige Voraussetzung erfüllt, dass die Mehrzahl der Gene nicht reguliert werden und die Durchführung der Normalisierungsmethode damit legitim ist.

## Ergebnisse und Diskussion

OE-Nummer	Stat. Rang	aerob/ phototrop	Funktions- klasse	Genname	Protein- Identifizierung	Proteinname
OE3822R	1	2.37	CHY	-	-	conserved hypothetical protein
OE3106F	2	2.09	EM	bop	Trusted	bacteriorhodopsin precursor
OE5134F	3	1.56	NOF	-	Trusted	conserved protein
OE3093R	4	1.55	LIP	crtB1	-	geranylgeranyl-diphosphate geranylgeranyltransferase (EC 2.5.1.32) (phytoene synthase)
OE6063F	5	1.27	TP	lctP	-	probable L-lactate permease
OE4132R	6	1.25	TL	trpS2	Trusted	tryptophan--tRNA ligase (EC 6.1.1.2)
OE3116F	7	1.31	NOF	-	Trusted	conserved protein
OE5276F	8	0.74	NOF	-	Trusted	conserved protein
OE2780F	9	0.29	HY	-	Questionable	hypothetical protein
OE3107F	10	1.6	HY	-	-	hypothetical protein
OE3101R	11	1.39	MIS	bat	Questionable	bacterioopsin activator
OE3556R	12	1.55	AA	carA	Insecure	carbamoyl-phosphate synthase (EC 6.3.-.-) small chain ((ammonia) (EC 6.3.4.16)
OE5102R	13	1.35	ISH	-	-	probable IS1341-type transposase (ISH12)
OE3554F	14	1.25	AA	carB	Trusted	carbamoyl-phosphate synthase (EC 6.3.-.-) large chain ((ammonia) (EC6.3.4.16 )
OE1019R	15	1.38	ISH	-	-	probable IS1341-type transposase (ISHnew88)
OE4121R	16	1.27	AA	ocd1	Believed	probable ornithine cyclodeaminase (EC 4.3.1.12)
OE4500R	17	0.76	LIP	acd6	Trusted	probable acylbutyryl-CoA dehydrogenase (EC 1.3.99.-)
OE2778R	18	0.58	HY	-	Questionable	hypothetical protein
OE1202F	19	0.76	TP	trp1	Trusted	probable ABC-type transport system ATP-binding/permease protein
OE2779F	20	0.48	TP	pheP	Trusted	amino acid transport protein (probable phenylalanine transport protein)
OE3324R	21	1.18	CHY	-	-	conserved hypothetical protein
OE3842R	22	1.11	HY	-	-	hypothetical protein
OE3930R	23	1.19	CHY	-	Questionable	protein synthesis inhibitor homolog
OE4668R	24	0.74	NOF	-	Insecure	conserved protein
OE3314R	25	1.45	HY	-	-	hypothetical protein
OE4651F	26	1.45	COM	thi1	Trusted	thiamine biosynthesis protein (thiazole biosynthetic enzyme)
OE1440F	27	1.22	ISH	-	-	probable IS1341-type transposase (ISHnew81)
OE1455R	28	0.62	NOF	-	Trusted	conserved protein
OE5248F	29	0.79	NOF	-	Trusted	conserved protein
OE1119F	30	1.23	CHY	-	Questionable	dTDPglucose 4,6-dehydratase C-terminal region homolog
OE5206R	31	0.81	AA	arcC	Trusted	carbamate kinase (EC 2.7.2.2)
OE5417R	32	1.19	CHY	-	-	conserved hypothetical protein
OE3270R	33	0.75	NOF	-	Trusted	conserved protein
OE5062R	34	1.22	ISH	-	-	probable IS1341-type transposase (ISHnew82)
OE1405R	35	1.37	NOF	-	Trusted	conserved protein
OE3959R	36	1.23	AA	ilvE	Trusted	branched-chain-amino-acid transaminase (EC 2.6.1.42)
OE4748F	37	0.7	NOF	-	Trusted	conserved protein
OE2186R	38	1.28	NOF	-	Trusted	conserved protein
OE7023R	39	1.39	MIS	gvpL1	Questionable	gas-vesicle operon protein gvpL1
OE1420F	40	0.89	TP	-	-	probable ABC-type transport system permease protein
OE3030R	41	0.84	NOF	-	Insecure	conserved protein
OE3057F	42	1.18	CHY	-	-	conserved hypothetical protein
OE4382R	43	1.42	HY	-	Insecure	hypothetical protein
OE2057F	44	1.63	COM	thiC	Trusted	thiamin biosynthesis protein thiC
OE5110F	45	1.19	CHY	-	-	conserved hypothetical protein
OE4336R	46	1.32	TP	nosY	Trusted	probable ABC-type transport system permease protein
OE3186F	47	0.9	COM	birA1	Insecure	biotin--[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) 1
OE5039R	48	1.14	CHY	-	-	conserved hypothetical protein
OE2170R	49	0.81	MIS	tmpC	Trusted	probable periplasmic substrate-binding protein (membrane lipoprotein tmpC homolog)
OE5158F	50	1.07	HY	-	-	hypothetical protein

Tab. 7: Liste mit den ersten 50 signifikantesten Genen des Zweigruppenvergleichs (aerob/phototrop).orangene Färbung: Induktion unter phototrophen Wachstumsbedingungen; hellgrüne Färbung: Repression unter phototrophen Wachstumsbedingungen.

Die Auswertung der DNA-Mikroarraydaten für den Zweigruppenvergleich von aerobem und phototrophen Wachstum ergab, dass nach der nicht überwachten statistischen Datenanalyse die Ergebnisse mit vorher durchgeführten Experimenten im Einklang stehen. So sind z.B. unter den ersten 11 Genen das *bop*, *crtB1* und das *bat* Gen zu finden, für die in früheren Experimenten eine entscheidende Beteiligung an der Regulation des phototrophen Wachstums von *H. sal.* R1 gezeigt wurde (Shand R und Betlach M, 1994; Baliga NS, et al., 2000). Dies ist eine erste Bestätigung, dass die für diese Arbeit konstruierten Gesamtgenom-DNA-Mikroarrays und die speziell entwickelte statistische Datenanalyse verlässliche Resultate liefert.

Im folgenden Diagramm wird die Verteilung dieser 242 differentiell exprimierten Gene auf die Funktionsklassen von *H. sal.* R1 gezeigt.

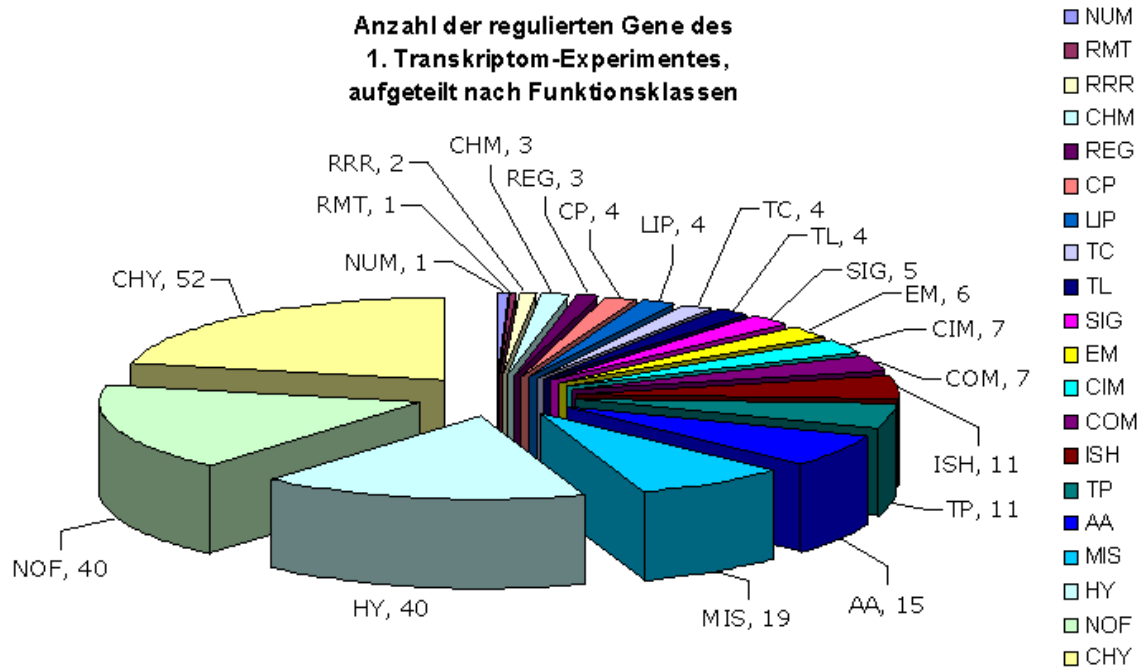


Abb. 44: Tortendiagramm der regulierten Gene des Transkriptom-Experimentes vom Vergleich aerober und phototropher Zellen. Dargestellt sind die 242 signifikantesten Gene berechnet auf der Basis der FS-Statistik (siehe Seite 91). Die Anzahl der Gene in den jeweiligen Funktionsklassen ist neben den Abkürzungen angegeben. Die Abkürzungen bedeuten: AA = Aminosäurestoffwechsel, CHM = Kohlenhydratstoffwechsel, CHY = konserviert hypothetisches Protein, CIM = zentraler Intermediatstoffwechsel, COM = Koenzym Stoffwechsel, CP = zelluläre Prozesse, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, ISH = ISH-kodierte Transposasen, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes, NOF = noch keine Funktion, NUM = Nukleotid-Stoffwechsel, REG = Genregulation, RRR = Reproduktion Reparatur Rekombination, SIG = Signalübertragung, TC = Transkription, TL = Translation, TP = Transport.

Wie dem Tortendiagramm aus Abb. 44 deutlich zu entnehmen ist, gehören den drei Funktionsklassen CHY, HY und NOF 132 Gene an. Dies entspricht einem großen Anteil von 54,5% der regulierten Gene. Bis jetzt können den Mitgliedern dieser Funktionsklassen keine homologen Proteine oder Funktionen anderer Organismen zugewiesen werden. Das bedeutet, dass für mehr als die Hälfte der regulierten Gene des Zweigruppenvergleichs keine eindeutige Aussage über ihre Zugehörigkeit bezüglich ihrer Rolle bei der Adaptation an das phototrophe Wachstum getroffen werden kann. Im Umkehrschluss bieten diese Proteine die Möglichkeit, in weiteren Versuchen, neue Wege bzw. Netzwerke, beim Prozess der Adaptation an das phototrophe Wachstum in *H. sal.* R1, zu entdecken und aufzuklären.

Schließt man die Funktionsklassen mit mehr als zehn und weniger als zwanzig differentiell exprimierten Genen zur nächst größeren Gruppe zusammen, bilden die Funktionsklassen MIS, AA, TP und ISH eine Gruppe mit 56 differentiell

exprimierten Genen, was einem Anteil von 23% entspricht. Diese Funktionsklassen sind jedoch sehr unterschiedlich, so dass eine getrennte Betrachtung erfolgt.

Die Funktionsklasse MIS weist mit 19 differentiell exprimierten Genen einen Anteil von 6,6% der differentiellen Gene auf. Die Mitglieder dieser Funktionsklasse besitzen ein breit verteiltes Spektrum von Aufgaben, so sind hier z.B. Enzyme enthalten, denen eine generelle Funktion zugeordnet werden kann, allerdings über die spezifischen Substrate keine Informationen bis zum jetzigen Zeitpunkt vorliegen. Ebenfalls zur MIS-Funktionsklasse gehören Gene, die an der Produktion von Gasvesikel beteiligt sind, von denen hier vier Gene unter den 19 Genen zu finden sind. Allerdings muß erwähnt werden, dass der *H. sal.* R1 Stamm keine Gasvesikel aufweist, die es den Zellen ermöglichen würden an die Oberfläche des Habitats zu gelangen, um eine bessere Versorgung z.B. mit Licht zu gewährleisten (HaloLex-Datenbank; Dassarma S, et al., 2001).

Von der Funktionsklasse des Aminosäurenstoffwechsels (AA) sind 15 Gene in der Liste enthalten, entsprechend eines Anteils von 6,2%. In dieser Funktionsklasse werden Enzyme zusammengefasst, die an der Biosynthese und dem Katabolismus von Aminosäuren beteiligt sind, mit der Ausnahme des Arginin-Deiminase-Stoffwechselweges und des Transports der Aminosäuren. Unter den 15 Kandidatengenen dieser Gruppe sind die zwei Gene *carA* und *carB* hervorzuheben, die für die beiden Untereinheiten der Carbamoyl-Phosphat-Synthase kodieren.

Die Funktionsklasse des Transports (TP) ist mit 11 Kandidatengenen und damit einem Anteil von 4,5% die nächst kleinere Gruppe von differentiell exprimierten Genen. In dieser Gruppe enthalten sind drei Gene die unter phototrophen Wachstumsbedingungen eine Induktion ihres Transkripts aufweisen, während die übrigen eine Repression zeigen. Ein weiteres auffälliges Merkmal dieser Gruppe ist, dass es sich hauptsächlich um ABC-Transportproteine handelt, deren zu transportierendes Substrat noch nicht experimentell bestimmt wurde. Die Gruppe von 11 Genen umfasst Permeasen, Substrat- und ATP- / Phosphat-bindende Proteine.

Den gleichen Anteil von 11% der regulierten Gene zeigt die Funktionsklasse ISH, in der alle Insertionssequenzen und andere mobile Elemente zusammengefasst sind. Hierzu kann man sagen, dass bis auf ein Gen alle anderen Gene eine Induktion der Transkriptmenge unter phototrophen Bedingungen zeigen. In der Induktion dieser mobilen Genabschnitte könnte ein möglicher

Adaptionsmechanismus liegen, der eine bessere Anpassung an die geänderte Energiequelle des halophilen Archaeons ermöglicht, denn durch Insertion von diesen mobilen Elementen wäre z.B. eine Unterbrechung von Transkriptionseinheiten denkbar, die für die Phototrophie nicht relevant sind.

Desweiteren ist aus dem Tortendiagramm (Abb. 46) zu entnehmen, dass es 13 Funktionsklassen gibt, die weniger als zehn Gene enthalten.

Mit je sieben Genen bilden die Funktionsklassen zentraler Intermediatstoffwechsel (CIM) und der Coenzymmetabolismus (COM) die beiden größten Gruppen, gefolgt vom Energiemetabolismus (EM) mit 6 und der Signaltransduktion (SIG) mit 5 Genen.

Bei den Genen der Funktionsklasse CIM sind *korA* und *korB*, dies sind die Untereinheiten der Oxoglutarat-Ferredoxin Oxidoreduktase, leicht induziert unter phototrophen Bedingungen. In der Gruppe des Coenzymmetabolismus sind die Gene *thiC* und *thi1*, die beide an der Thiaminbiosynthese beteiligt sind, induziert im Vergleich zum aeroben Wachstum. Ebenfalls eine Induktion unter phototrophen Bedingungen zeigen die Gene *nuoCD* und *nuoK* (Untereinheiten der NADH-Dehydrogenase) sowie *dmsB* (Kette B der Dimethylsulfoxidreduktase) des Energiemetabolismus. Die stärkste Induktion in dieser Gruppe weist das *bop* Gen auf, das das Apoprotein Bakteriorhodopsin kodiert. Die Gene der Signaltransduktion (SIG) zeigen eine Repression unter phototrophen Bedingungen und sind alle Transducerproteine wie *htr1*, *htr3(basT)* und *htr8*.

Als nächste Gruppe werden die Funktionsklassen zusammengefasst die vier Gene enthalten.

Darunter sind z.B. die Funktionsklassen der Translations- (TL) und Transkriptionsgene (TC) zu finden. Dies ist besonders zu beachten, da die Mitglieder dieser Funktionsklassen für die Regulation der Proteinbiosynthese verantwortlich sind, die sowohl auf der Ebene der Transkription, als auch auf der Ebene der Translation stattfinden kann. So sind unter diesen acht Genen Transkriptionsinitiationsfaktoren, eine Untereinheit der DNA-abhängigen RNA-Polymerase, zwei t-RNA-Ligasen (Leucin, Tryptophan) und ribosomale Proteine. All dies zeigt deutlich, dass eine Neuorganisation des Proteininventars der halobakteriellen Zellen beim Wechsel zwischen den beiden Wachstumsbedingungen stattgefunden hat.

Auch die Funktionsklasse des Lipidstoffwechsels (LIP) weist vier Gene auf, unter anderem die beiden Gene *crtB1* und *crtB2*, die als Phytoensynthasen an der

Carotenoidbiosynthese beteiligt. Da sie die Reaktion zur Bildung einer wichtigen Vorstufe des Retinals katalysieren, welches als Cofaktor über eine Schiff'sche Base in das Apoprotein Bakteriorhodopsin eingebaut wird, sind beide Gene beim Vorgang der Adaptation an die phototrophen Bedingungen entscheidend beteiligt.

Die 242 differentiell exprimierten Gene des Zweigruppenvergleichs verteilen sich auf 20 Funktionsklassen. In der HaloLex-Datenbank sind für *H. sal.* R1 24 Funktionsklassen festgelegt worden. Die vier Funktionsklassen zu denen keines der differentiellen Gene zugeordnet werden konnte sind die Proteinsekretion (SEC), die Motilität (MOT) und Gene die für Chaperone (CHP) und Zellhüllenproteine (CE) kodieren.

### 5.2.1 Diskussion

Es ist seit mehr als drei Jahrzehnten bekannt, dass Archaea eine eigene monophyletische Gruppe bilden, wobei sie im Gegensatz zu ihrer prokaryontischen Morphologie näher mit den Eukaryonten als mit Bakterien verwandt sind. Diese Tatsache konnte durch die Isolierung neuer Arten, der Untersuchung unkultivierter Archaea und durch mehrere Genomsequenzierungen des Öfteren unterstrichen werden.

Dabei ist anzumerken, dass sich Haloarchaea hervorragend eignen die archaeale Biologie zu studieren. Denn es existieren partielle oder komplette Genomsequenzen von mehreren Organismen, die es erlauben *in silico* eine Charakterisierung von haloarchaealen Eigenschaften durchzuführen. Außerdem existieren Methoden chromosomale Gene in-frame zu verändern oder zu deletieren. Es konnten auch einige biochemische, zellbiologische und genetische Arbeitsmethoden an die hohe Salzkonzentration im Zytoplasma von Haloarchaea angepasst werden.

Durch diese Entwicklungen wurde das Verständnis in einigen Bereichen der haloarchaealen Biologie erheblich erweitert. In dieser Tradition ist auch die Etablierung der Transkriptomanalyse für *H. sal.* R1 zu sehen.

Im Rahmen der Etablierung der Transkriptomanalyse, soll das Projekt dazu dienen durch diese Gesamt-Genom Methode das Wissen über den Prozess der Anpassung an das phototrophe Wachstums zu erweitern.

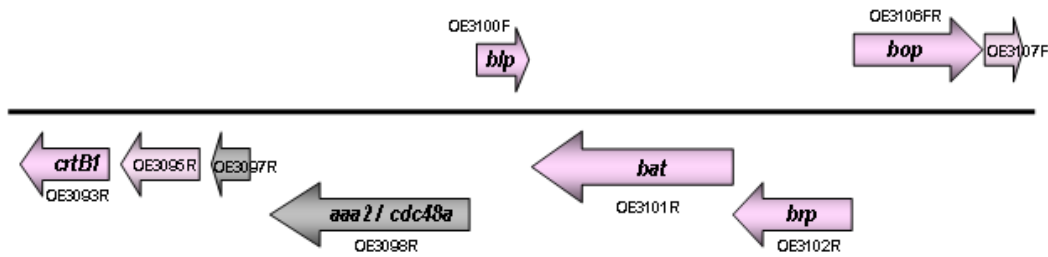
In früheren Untersuchungen konnte gezeigt werden, dass zur Biosynthese der Pupurmembran neben dem *bop* Gene, welches das Bakterioopsin Gen kodiert,



einige weitere benachbarte chromosomale Gene wie das *brp*, *bat* und *crtB1* Gen notwendig sind. Diese vier Gene bilden einen vernetzten Cluster und zeigen eine Koregulation ihrer Expression. Sie sind beteiligt an der Synthese des Chromophors, der genetischen Regulation und oder der Biogenese der Membran.

Die Organisation dieses so genannten *bop* Clusters ist in folgender Abbildung dargestellt.

### Bakterio-Op $\text{sin}$ Gen Cluster



<i>bat</i>	<i>bop</i> „activator gene“	Sauerstoffsensor (?)
<i>blp</i>	<i>bop</i> „linked product gene“	?
<i>bop</i>	„bacterio-opsin gene“	Licht getriebene Protonenpumpe
<i>brp</i>	<i>bop</i> „related protein gene“	Retinal Biosythese (?)
<i>crtB1</i>	„phytoene synthase“	Retinal + Carotinoid Biosythese

Abb. 45: Darstellung der chromosomalen Anordnung der Gene des so genannten *bop* Gen Clusters. Für in grau dargestellte Gene wurde bisher keine Beteiligung an der Biosynthese der Purpirmembran nachgewiesen.

Es konnte nachgewiesen werden, dass das *crtB1* Gen einen Schlüsselschritt in der Carotenoid-Biosynthese katalysiert. Ebenso konnten frühere Arbeiten zeigen, dass eine Insertion in das *bat* (Bakterioopsinaktivator Gen) oder in das *brp* Gen (Bakterioopsin verwandtes Gen) die Transkription des *bop* Gens blockiert. Für das *bat* Gen wird eine regulatorische Funktion vorhergesagt. Diese Vermutung basiert auf Ergebnissen von Versuchen und einer Analyse der Domänenstruktur des Proteins. Nach der Domänenanalyse zeigte sich, dass das 74 kDa große BAT Protein aus einer PAS/PAC Domäne, einer GAF-Domäne und einem DNA-bindendem Helix-Turn-Helix (HTH) Motiv aufgebaut ist.

Es ist bekannt, dass bei *H. sal.* R1 die Transkription des *bop* Gens ab der mittleren bis späten exponentiellen Wachstumsphase induziert wird. Jedoch ist der Organismus auch fähig auf äußere Signale zu reagieren. So tritt bedingt durch eine geringe Sauerstoffkonzentration und oder eine hohe Lichtintensitäten eine Induktion des *bop* Gens ein. Damit ist das halophile Archaeon in der Lage

durch die gesteigerte Biosynthese der lichtgetriebenen Protonenpumpe phototroph zu wachsen.

Die bisher bekannten regulatorischen Zusammenhänge der Bakteriorhodopsin Biosynthese werden in den folgenden Abbildungen dargestellt.

### Regulation der BR Expression durch zwei Signale:

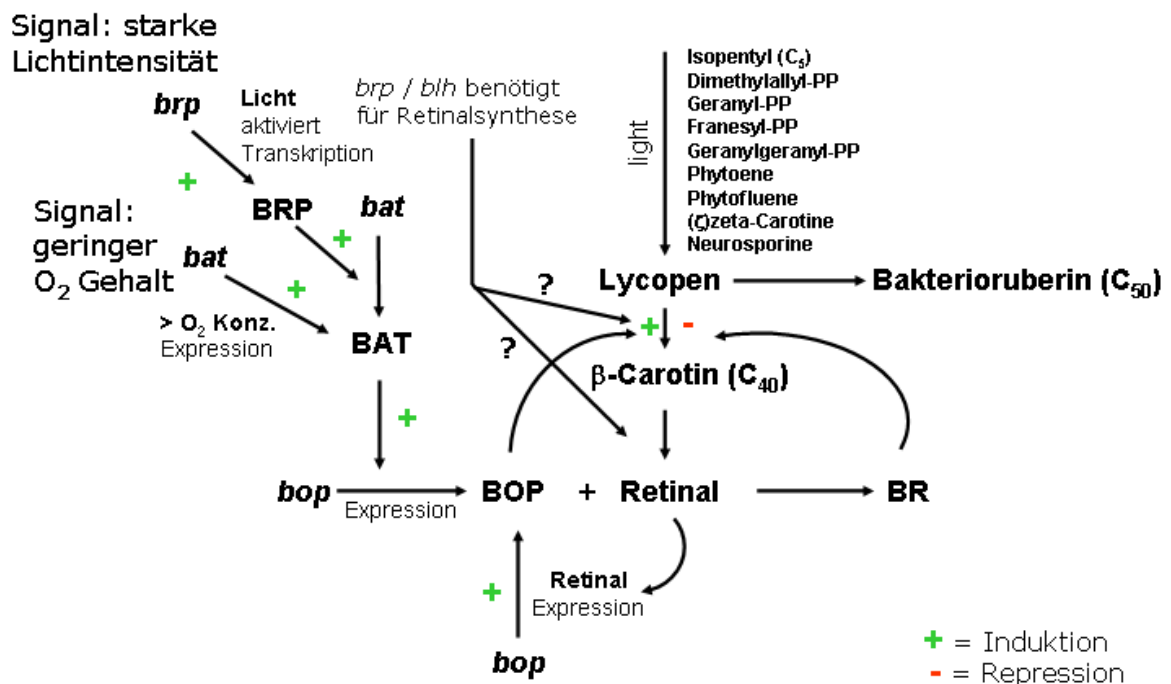


Abb. 46: Grafische Darstellung des bisher bekannten Regulationsnetzwerkes der BR Biosynthese durch die Signale geringe Sauerstoffkonzentration und starke Lichtintensität.

Eine geringe Sauerstoffkonzentration führt zu einer Induktion des *bat* Gens. Das BAT Protein wiederum bewirkt eine Induktion des *bop* Gens, was zu einer Biosynthese des Bakterioopsins (BOP) führt, welches noch kein Retinal als Chromophor enthält. Die Biosynthese des Chromophors Retinal ist Teil der Isoprenoidbiosynthese und setzt auf der Stufe des Lycopens ein. Denn das Lycopene ist sowohl die Vorstufe der Bakterioruberine (C<sub>50</sub> Körper), als auch des  $\beta$ -Carotins (C<sub>40</sub> Körper), was durch Spaltung zum Retinal wird.

Eine gesteigerte Expression des BOP-Proteins, bewirkt eine erhöhte Produktion von  $\beta$ -Carotin und damit Retinal. Dieses wird in das Bakterioopsin als Chromophor eingebaut, wodurch die funktionsfähige Protonenpumpe Bakteriorhodopsin (BR) entsteht. Ein Überschuss an Retinal verursacht eine Induktion des *bop* Gens. In einer Art Feedback Mechanismus reprimiert das

Bakteriorhodopsin wiederum die Bildung von Retinal, denn eine zu hohe freie Konzentration an Retinal würde die Zelle schädigen.

Es wird vermutet, dass die Gene *brp* und *blh* an der Biosynthese des Retinals beteiligt sind, jedoch gibt es dafür bisher noch keine eindeutigen Beweise (Peck RF, et al., 2001).

Bedingt durch eine hohe Lichtintensität wird die Transkription des *brp* Gens aktiviert. Die vergrößerte Rate an BRP Protein seinerseits induziert die Expression des *bat* Gens, wodurch die Kaskade der BR Biosynthese ausgelöst wird (Shand RF, et al., 1991).

Wie unter (5.2) erwähnt, wurden auf der Basis der Datenanalyse für den Transkriptionsversuch des Zweigruppenvergleichs 242 Gene als differentiell exprimiert ausgewählt.

Die Basis der Liste stellt das Ergebnis der FS-Statistik da (siehe Seite 91). Zur Fehlerabschätzung wurde die FDR Methode benutzt (siehe Seite 98), die einen gewissen Anteil an falsch positiven Kandidatengen erlaubt. Wie hoch der Anteil der akzeptierten falsch positiven Gene ist, liegt im Ermessen des jeweiligen Experimentators und umfasst einen Bereich von 5-20%.

Die erhaltene Liste von signifikant differentiell exprimierten Genen, ist so aufgebaut, dass an den ersten Positionen die am höchsten regulierten Gene stehen.

Unter den ersten 11 Genen dieser Liste befinden sich z.B. die Gene *bop*, *bat* und *crtB1*. Dadurch kann eindeutig bewiesen werden, dass die Resultate des ersten großen Transkriptomexperiments mit Ergebnissen früherer Arbeiten übereinstimmen. Dies ist ein wichtiger Punkt, denn die Datenanalyse wurde in einem „unsupervised“ Verfahren ausgeführt. Was bedeutet, dass keinerlei biologisches Vorwissen in die Datenanalyse eingeflossen ist. Somit ist das Vorhandensein der Gene *bop*, *bat* und *crtB1*, als eine indirekte Bestätigung für die Richtigkeit des Transkriptionsexperiments anzusehen.

Ein ebenso offensichtliches Ergebnis dieses Experiments zeigt, dass der Anteil der Gene ohne ein homologes Protein oder Funktion mit 54,5% sehr hoch liegt. Diese 54,5% der regulierten Gene entsprechen 132 Genen, die sich auf die Funktionsklassen CHY (konserviert hypothetisches Protein), HY (hypothetisches

Protein) und NOF (noch keine Funktion) verteilen. Wie auch bei anderen mikrobiellen Genomen, kann bei *H. sal.* R1 die funktionelle Genomannotation aufgrund fehlender Sequenzhomologie zu bekannten Proteinen, nicht optimal ausgenutzt werden.

Wenn man bei der Verteilung der regulierten Gene auf die Funktionsklassen, die Klassen CHY, HY und NOF ausklammert kann man sehen, dass beim Vergleich aerobes und phototrophes Wachstum, fast alle Funktionsklassen differentiell exprimierte Gene enthalten, was darauf hindeutet, dass es zu einer breiten Neuorganisation des Proteinrepertoires oder des Regulationsnetzwerkes gekommen ist.

Betrachtet man die Anzahl der regulierten Gene pro Funktionsklasse, so bilden die Klassen, MIS (Verschiedenes), AA (Aminosäurenstoffwechsel) und TP (Transport) die Gruppe mit der größte Anzahl an Genen. Dies ist nicht verwunderlich, denn diese Funktionsklassen decken ein breites Spektrum an zellulären Prozessen ab, wie z.B. solche die für den Energiehaushalt der Zelle verantwortlich sind.

Für dieses Transkriptionsexperiment darf man nicht außer Acht lassen, dass für die Hälfte der Gene keine Aussage zu ihrer Funktion getroffen werden kann. Andererseits bedeutet für hypothetische ORFs das Auffinden von Transkripten, dass sie es sich um kodierende Sequenzen handelt und die in silico Vorhersage zutrifft. Wobei erwähnt werden muss, dass erst die Detektion des Proteins seine Existenz beweisen würde.

Man kann also sagen, dass mit dem Ende des Transkriptomexperiments die Herausforderung beginnt, die Funktionen dieser Gene aufzuklären. Es ist eine Sache die Expressionsrate von Genen zu bestimmen, aber eine andere deren Funktion aufzudecken.

In diesem Zusammenhang wäre ein denkbarer Ansatz eine vertiefende bioinformatische Analyse der als reguliert gefundenen „*conserved hypothetical genes*“ durchzuführen.

Ansätze für praktische Versuche, wären der Vergleich von Stämmen unterschiedlichen Phäno- und Genotyp bzw. die Herstellung von Deletionsmutanten. Für dieses Projekt wurde der praktische Ansatz der Herstellung von Deletionsmutanten gewählt. Wie in Kapitel 5.4.2 beschrieben,

wurden auf der Basis der Ergebnisse des Zweigruppenvergleichs die zwei ORFs OE3107F und OE3136F deletiert und die Mutanten beschrieben.

Für weitere Experimente sollte man den Aspekt, dass beim phototropen Wachstum zwei äußere Signale, wie geringe Sauerstoffkonzentration und starke Lichtintensität die Expression von BR induzieren, nicht außer Acht lassen. Denn die Adaption an das phototrophe Wachstum, ist eine Antwort auf zwei Signale.

In Kapitel 5.4.3, der Untersuchung zweier Stämme S9H und S9D im Vergleich zum WT R1, ist ein Versuchsaufbau zur möglichen Entkoppelung der Signale gewählt worden (Abb. 79).

Parallele zur Charakterisierung von Mutanten, wurden weitere DNA-Mikroarray-Experimente durchgeführt. Dazu wurden Zeitreihenexperimente gewählt, um den Prozess der Adaption an das phototrophe Wachstum zeitaufgelöst zu betrachten.

Die Zeitreihenexperimente sind Bestandteil des nächsten Kapitels.

### **5.3      Differentiell      exprimierte      Gene      der Zeitreihenexperimente**

Die statistische Datenanalyse der Transkriptom-Daten der zwei Zeitreihenexperimente erfolgt analog zu der des Zweigruppenvergleichs, mit dem Unterschied, dass hier die verschiedenen Zeitpunkte mit in das ANOVA-Modell einbezogen werden. Als Ergebnis der Datenauswertung erhält man eine Liste von differentiell exprimierten Genen, die in Bezug auf alle Zeitpunkte des Experiments, eine signifikante Veränderung zeigen.

Ziel dieser Zeitreihenexperimente ist es, für die verschiedenen Zeitintervalle der Experimente einen Expressionsverlauf der Gene zu erhalten, um so mehr Informationen über den Verlauf der Anpassung auf Transkriptomebene beim Wechsel von aerobem auf phototropes Wachstum zu bekommen.

Entsprechend den Ergebnissen des Zweigruppenvergleichs, ist auch für das Zeitreihenexperiment I, dem Wechsel von aerobem Wachstum auf phototropes Wachstum mit dem Zeitintervall bis 120 min, die vollständige Liste mit den differentiell exprimierten Genen in der HaloLex-Datenbank hinterlegt. Als Beispiel für diese Ergebnisliste sind in der folgenden Tabelle die ersten 50 signifikantesten Gene enthalten, deren statistische Datenanalyse auf dem Hypothesentest der FS-Statistik basiert.

Auch durch das zweite Transkriptom-Experiment von *H. sal.* R1 konnte mittels der Auswertung der DNA-Mikroarraydaten eindeutig bewiesen werden, dass die Hybridisierungsergebnisse im Konsens mit beschriebenen Ergebnissen stehen. So sind auch hier das *bop*, *bat* und *crtB1* Gen als induziert unter phototrophen Bedingungen gefunden worden. Ein weiteres Indiz also für die Zuverlässigkeit der etablierten Methode und deren Auswertung.

OE-Nummer	Stat. Rang	$t_1/t_0$	$t_2/t_0$	$t_3/t_0$	$t_4/t_0$	Funktionsklasse	Genname	Protein-Identifizierung	Proteinname
OE1821R	1	1.08	1.02	0.92	0.72	TL	rpl36aR	-	ribosomal protein L36a.eR (HLA)
OE3099R	2	1.07	1.16	1.29	1.09	LIP	crtB1	-	geranylgeranyl-diphosphate geranylgeranyltransferase (EC 2.5.1.32) (phytoene synthase)
OE3106F	3	0.91	1.06	1.19	1.59	EM	bop	Trusted	bacteriorhodopsin precursor
OE3925R	4	1.07	0.78	1.07	1.72	CP	cctB	Trusted	thermosome beta chain
OE3992R	5	0.94	1.1	0.9	0.66	EM	atpH	Insecure	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain H.a (A-type ATP synthase)
OE7063R	6	0.98	1.01	1.3	1.12	-	-	-	spurious ORF
OE3097R	7	1.11	1.1	1.2	1.02	HY	-	Questionable	hypothetical protein
OE4561F	8	1.08	1.13	1.15	0.91	CHY	-	Questionable	conserved hypothetical protein
OE5296F	9	0.98	1	1.1	1.07	HY	-	-	hypothetical protein
OE4122R	10	1.02	0.87	1.1	1.47	CP	cctA	Trusted	thermosome alpha chain
OE4427R	11	1.02	0.98	1.15	1.32	MIS	dpsA	Trusted	DNA-binding protein dpsA
OE1187R	12	1.01	1.01	0.96	0.76	-	-	-	spurious ORF
OE1584R	13	0.88	1.1	1.05	0.63	NOF	-	Trusted	conserved protein
OE3141R	14	1.05	1	0.96	0.75	TL	rpl37R	Questionable	ribosomal protein L37.eR
OE4370R	15	0.84	1.08	1.15	1.11	MIS	dip1	Trusted	drug export protein homolog
OE2864F	16	1.24	1.05	0.84	0.87	NUM	purD	Trusted	phosphoribosylamine--glycine ligase (EC 6.3.4.13)
OE4460R	17	1.05	1.1	0.92	0.69	TL	rpl39R	-	ribosomal protein L39.eR
OE2053F	18	1.01	1.04	0.91	0.87	HY	-	-	hypothetical protein
OE3468R	19	1.2	0.98	1.27	1.02	LIP	crtI2	Trusted	phytoene dehydrogenase (EC 1.14.99.-) (phytoene desaturase) 2
OE1371R	20	1.02	1.08	0.95	0.84	MIS	pfdB	Trusted	prefoldin beta chain
OE3487R	21	0.88	1.05	1	0.77	TL	eif5a	-	translation initiation factor eIF-5A
OE3759R	22	1	1.03	0.75	0.87	HY	-	-	hypothetical protein
OE2602R	23	1.08	0.95	0.98	0.85	TL	rpl1	Trusted	ribosomal protein L1
OE3634F	24	0.97	1.04	0.96	0.87	CIM	icd	Trusted	isocitrate dehydrogenase (NADP+) (EC 1.1.1.42)
OE7155R	25	0.9	1.01	1.12	1.18	CHY	-	-	conserved hypothetical protein
OE1227F	26	0.95	0.96	1.15	1.01	CHY	-	-	conserved hypothetical protein
OE1442R	27	0.97	1.06	0.86	0.89	CHY	-	Trusted	conserved hypothetical protein
OE1241R	28	1.01	1	1.14	0.94	CHY	-	-	conserved hypothetical protein
OE1858F	29	1.08	0.97	0.98	0.95	HY	-	-	hypothetical protein
OE2603R	30	1.08	0.94	0.88	0.81	TL	rpl11	Trusted	ribosomal protein L11
OE2648F	31	1.06	0.99	0.8	0.9	NOF	-	Trusted	conserved protein
OE3346R	32	0.97	1.01	1.11	0.86	NOF	-	Trusted	conserved protein
OE3400F	33	0.86	0.93	0.64	2.03	TL	rps17	Insecure	ribosomal protein S17
OE5083R	34	1.02	0.92	1.11	1.14	NOF	hsp5	Believed	heat shock protein homolog
OE4511R	35	1.04	1.05	0.8	0.77	HY	-	-	hypothetical protein
OE3216F	36	0.97	0.98	0.99	0.88	COM	cbiH2, cebJ2	Insecure	precorrin-3B C17-methyltransferase (EC 2.1.1.131) 2
OE3410F	37	0.9	1	0.95	0.99	TL	rps8	Trusted	ribosomal protein S8
OE3107F	38	0.94	1.14	1.15	1.26	HY	-	-	hypothetical protein
OE3136F	39	1.19	1.1	1.08	0.86	CHY	-	-	conserved hypothetical protein
OE3284R	40	1.05	0.99	0.96	0.86	NOF	-	Trusted	conserved protein
OE1818R	41	0.98	1.07	0.9	0.84	TL	eif2a	Trusted	translation initiation factor eIF-2 alpha chain
OE3728R	42	1.01	1.03	1	0.95	HY	-	-	hypothetical protein
OE5413F	43	0.95	1.03	1.15	0.99	CHY	-	-	conserved hypothetical protein
OE1270F	44	0.98	0.93	0.95	0.9	AA	gdhA3	Trusted	glutamate dehydrogenase (EC 1.4.1.2)
OE4735R	45	0.97	1.03	0.93	0.83	TL	rps7	Trusted	ribosomal protein S7
OE1860F	46	0.93	1.09	1.07	0.71	CHY	-	Questionable	conserved hypothetical protein
OE4674F	47	1.03	1.04	1.05	1.1	CIM	adh2	Trusted	alcohol dehydrogenase (EC 1.1.1.1)
OE24321R	48	1.17	1	0.96	0.87	HY	-	-	hypothetical protein
OE3942R	49	0.9	1.03	0.82	1.46	HY	-	-	hypothetical protein
OE7192F	50	0.9	1.05	1.29	1.01	NOF	-	Insecure	conserved protein
OE2304F	51	0.85	0.84	1.12	1.14	RRR	gyrA	Trusted	DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) chain A

Tab. 8: Liste mit den ersten 50 signifikantesten Genen des ersten Zeitreihenexperiments.

Abgesehen von einer anderen FDR im Vergleich zum ersten Transkriptom-Experiment, umfasst die Liste der differentiell exprimierten Gene auch 242.

Die Verteilung dieser 242 Gene auf die 24 Funktionsklassen von *H. sal.* R1 ist in Abb. 47 graphisch als Tortendiagramm dargestellt.

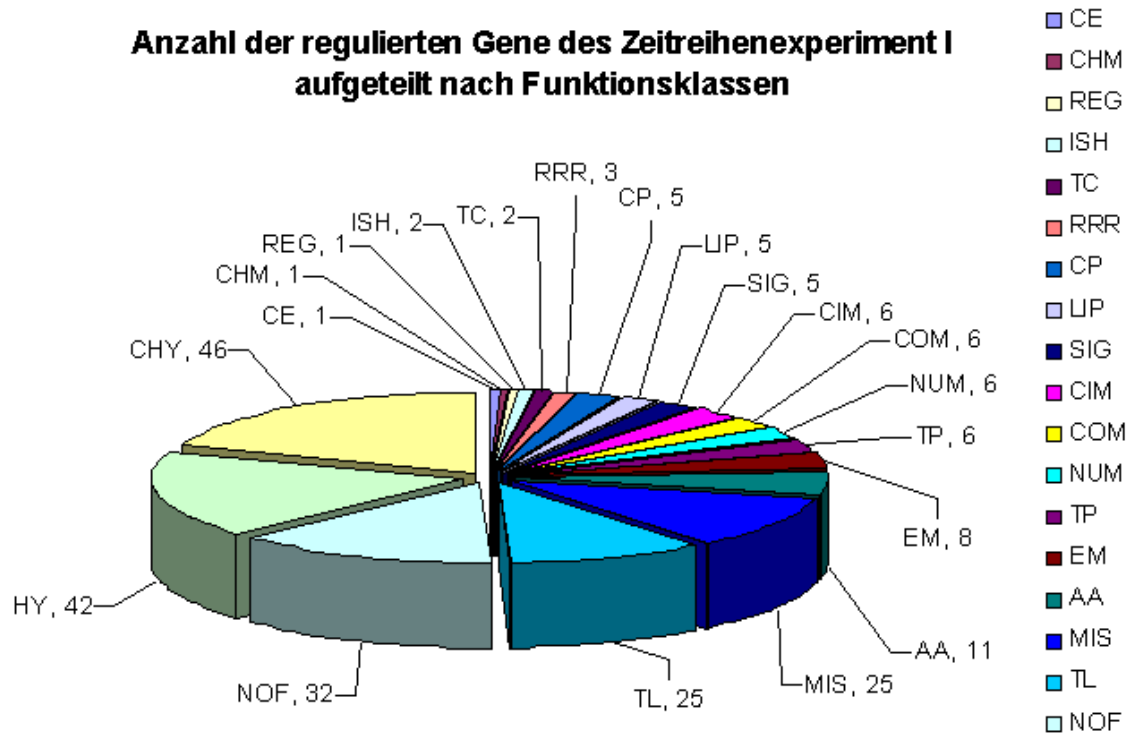


Abb. 47: Tortendiagramm der regulierten Gene des Transkriptom-Experiments der ersten Zeitreihe. Dargestellt sind die 242 signifikantesten Gene, berechnet auf der Basis der FS-Statistik. Die Anzahl der Gene in den jeweiligen Funktionsklassen ist neben den Abkürzungen angegeben. Die Abkürzungen bedeuten: AA = Aminosäurestoffwechsel, CE= Zellhülle, CHM = Kohlenhydratstoffwechsel, CHY = konserviert hypothetisches Protein, CIM = zentraler Intermediatstoffwechsel, COM = Koenzym Stoffwechsel, CP = zelluläre Prozesse, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, ISH = ISH-kodierte Transposasen, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes, NOF = noch keine Funktion, NUM = Nukleotid-Stoffwechsel, REG = Genregulation, RRR = Reproduktion Reparatur Rekombination, SIG = Signalübertragung, TC = Transkription, TL = Translation, TP = Transport.

Wenn man die Funktionsklassen analog der Anzahl der gefundenen Gene einteilt, erhält man sechs verschiedene Gruppen.

Die größte Gruppe mit den Funktionsklassen CHY, HY und NOF enthält 120 Gene und weist damit einen Anteil von fast 50% der regulierten Gene auf. Im Vergleich zum ersten Transkriptomexperiment ist dieser Anteil von Genen um knapp 5% geringer, jedoch kann diesen Genen bis zum jetzigen Zeitpunkt noch keine Funktion zugewiesen werden und somit keine Aussage über ihre Rolle und Beteiligung an der Anpassung an phototrophe Wachstumsbedingungen getroffen werden.

In der zweiten Gruppe sind die Funktionsklassen Translation (TL) und Verschiedenes (MIS) mit je 25 der differentiell exprimierten Genen was einem Anteil von 10,3% entspricht. Hier zeigt sich im Vergleich zum ersten

Transkriptom-Experiment ein auffälliger Unterschied. So waren im ersten Expressionsexperiment 4 Gene der Translationsfunktionsklasse und in diesem Zeitreihenexperiment sind es 25 Kandidatengene, die unter den differentiell exprimierten Genen enthalten sind. Unter diesen 25 Genen sind 18 ribosomale Proteine zu finden, die einem Anteil von 31,6% aller ribosomaler Proteine von *H. sal.* R1 entsprechen. Desweiteren schließt diese Gruppe von Genen zwei t-RNA Ligasen (Lysin/Tryptophan), drei Translationsinitiationsfaktoren (*elif2a*, *elif5a* und *infB*) sowie den Translationselongationsfaktor *aef1a* mit ein.

Die dritte Gruppe der Funktionsklassen weist einen Anteil von 7,8% auf und besteht aus dem Aminosäurenstoffwechsel (AA) mit 11 Kandidaten und dem Energiemetabolismus (EM) mit 8 Kandidaten. In dieser Gruppe ist auffällig, dass unter den Kandidatengenen des Energiemetabolismus drei der insgesamt neun Untereinheiten der ATP-Synthase von *H. sal.* R1 zu finden sind. Dies deutet offensichtlich daraufhin, dass sich die halobakteriellen Zellen dem stark veränderten Energiezustand anpassen.

Mit einem Anteil von zusammen 10% an den differentiell exprimierten Genen, besteht die vierte Gruppe aus den vier Funktionsklassen Transport (TP), Coenzymmetabolismus (COM), Nukleotid (NUM)- und zentraler Intermediatstoffwechsel (CIM) mit je 6 Kandidatengenen. Ebenso wie beim Zweigruppenvergleich sind unter den Kandidatengenen der Transportfunktionsklasse hauptsächlich ABC-Transportproteine und unter den Mitgliedern des Coenzymmetabolismus (COM) mit *thiC* und *thiD* auch wieder Gene der Thiaminbiosynthese zu finden. Zu den Kandidatengenen des Nukleotidstoffwechsels zählt das *cbiH2*, das an der Biosynthese von Percorin beteiligt ist. Die Gruppe des zentralen Intermediatstoffwechsels (CIM) setzt sich z.B. aus den Genen der Pyruvat-Carboxylase (*pyc*), der Pyruvat-Dehydrogenase (*porB*) und der Pyruvat-Kinase (*pykA*) zusammen.

Als vorletzte Gruppe kann man die Funktionsklassen der Signaltransduktion (SIG), des Lipidstoffwechsels (LIP) und der zellulären Prozesse (CP) mit je 5 Genen und einem Anteil von 6,3% der differentiell exprimierten Gene ansehen. Im Hinblick auf eine Neuorganisation der Proteinausstattung der Zellen von *H. sal.* R1 während des Anpassungsprozesses an die phototrophen Bedingungen ist die Tatsache, dass die beiden Untereinheiten des Thermosoms *cctA* und *cctB*, als Gruppe II Chaperone, die zu den Kandidatengenen der Funktionsklasse der zellulären Prozesse (CP) gehören, ein wichtiges Resultat. Dies ist deshalb von Bedeutung, da Chaperone an der nativen Faltung von Proteinen beteiligt sind.



Neben den Untereinheiten des Thermosoms, die Funktions- und Sequenzhomolog zu den TRIC (TCP-I) Chaperonen sind, ist ein weiteres Chaperon, das *cspD1*, ein Kälteschock-Chaperon, unter den differentiell exprimierten Genen. Das Vorhandensein der Carotinoidbiosynthesegene *crtB1* (Phytoensynthase), *crtI2* (Phytoendehydrogenase) und *crtY* (Lycopencyclase) unter den Kandidaten des Lipidstoffwechsels (LIP) zeigt auch hier die Bedeutung der Carotenoidbiosynthese im Zusammenhang mit der Adaptation an phototrophe Bedingungen. Die Kandidaten der Gruppe der Signaltransduktion setzen sich aus *cheA*, einer Sensorhistidinkinase, zwei Transducerproteinen *htr12* und *htr15* sowie zwei wahrscheinlichen Responsregulatoren *hrg* und *hlx2* zusammen.

Mit einem bis drei Kandidatengenen pro Funktionsklasse besteht die sechste Gruppe aus den Funktionsklassen Reproduktion, Reparatur, Rekombination (RRR), Transkription (TC), ISH-kodierende Transposasen (ISH), Genregulation (REG), Kohlenhydratstoffwechsel (CHM) und Zellhülle (CE).

Ähnlich wie beim ersten Transkriptom-Experiment teilen sich die 242 differentiell exprimierten Gene auf 20 verschiedene Funktionsklassen auf. Zu den fehlenden Funktionsklassen gehören die RNA-Reifung (RMT), die Proteinsekretion (SEC), die Motilität (MOT) und die Chaperone (CHP).

Die Liste der differentiell exprimierten Gene in Bezug auf das Zeitintervall von 24h, umfasst unabhängig von den anderen FDR-Werten ebenfalls 242 Gene nach der FS-Statistik basierenden Datenauswertung. Analog zu den vorher besprochenen Transkriptom-Experimenten, werden auch für das Zeitreihenexperiment II die ersten 50 signifikantesten Gene in der Tabelle 9 aufgelistet.

OE-Nummer	Stat. Rang	t <sub>0</sub> /t <sub>1</sub>	t <sub>1</sub> /t <sub>2</sub>	t <sub>2</sub> /t <sub>3</sub>	t <sub>3</sub> /t <sub>4</sub>	t <sub>4</sub> /t <sub>5</sub>	Funktionsklasse	Genname	Protein-Identifizierung	Proteinname
OE1162R	1	1.35	1.09	1.03	0.81	0.66	CP	ospD1	-	probable cold shock protein
OE1249R	2	0.84	1.05	1.11	1.09	1.27	CHY	-	-	conserved hypothetical protein
OE1797R	3	1.57	1.13	0.83	0.89	1.66	REG	sirR	Questionable	transcription regulator sirR
OE1821R	4	1.19	1	0.92	0.89	0.67	TL	rpl36aR	-	ribosomal protein L36a.eR (HLA)
OE1981R	5	1.38	1.04	0.92	0.89	0.71	HY	-	-	hypothetical protein
OE2013R	6	0.93	1.06	1.24	1.12	1.13	LIP	acd4	Trusted	probable acyl/butyryl-CoA dehydrogenase (EC 1.3.99.-)
OE2159R	7	1.4	0.89	1	0.92	0.7	TL	rps3aR	Trusted	ribosomal protein S3a.eR
OE2359F	8	1	0.97	0.98	1.16	1.33	NOF	-	Trusted	conserved protein
OE2515F	9	0.85	1.11	1.16	1.22	1.24	NOF	-	Trusted	conserved protein
OE2603R	10	1.62	1.08	0.86	0.77	0.56	TL	rpl11	Trusted	ribosomal protein L11
OE2628F	11	1.42	1	0.96	0.86	0.59	TL	rps4	Trusted	ribosomal protein S4
OE2629F	12	1.42	1.12	0.83	0.88	0.59	TL	rps11	Insecure	ribosomal protein S11
OE2638F	13	1.52	0.94	0.98	0.87	0.65	TC	rps9K	-	DNA-directed RNA polymerase (EC 2.7.7.6) chain K
OE2644F	14	1.33	1.03	0.95	0.97	0.66	TL	rps28R	Questionable	ribosomal protein S28.eR
OE2750R	15	1.29	1.02	0.87	0.85	0.58	NOF	-	Trusted	conserved protein
OE2906R	16	0.64	0.99	1.18	1.25	0.96	MIS	sod2	Insecure	superoxide dismutase (EC 1.15.1.1) 2
OE3062F	17	1.39	1.08	0.93	0.87	0.59	TL	rps17R	-	ribosomal protein S17.eR
OE3107F	18	1.32	1.24	1.15	0.93	1.14	HY	-	-	hypothetical protein
OE3141R	19	1.36	0.93	0.98	0.89	0.6	TL	rpl37R	Questionable	ribosomal protein L37.eR
OE3142R	20	1.3	0.89	1.05	0.86	0.58	MIS	snp	Questionable	S10 protein
OE3390F	21	1.6	0.92	0.95	0.94	0.63	TL	rpl23	Trusted	ribosomal protein L23
OE3395F	22	1.55	0.96	0.98	0.93	0.55	TL	rps3	Trusted	ribosomal protein S3
OE3398F	23	1.74	0.9	1.04	0.67	0.48	CHY	-	-	conserved hypothetical protein
OE3408F	24	1.48	1.05	0.92	0.84	0.53	TL	rps14	-	ribosomal protein S14
OE3410F	25	1.54	0.93	0.93	0.84	0.59	TL	rps8	Trusted	ribosomal protein S8
OE3416F	26	1.67	1.16	0.75	0.88	0.55	TL	rpl30	Insecure	ribosomal protein L30
OE3487R	27	1.14	1.11	0.85	0.93	0.71	TL	eif5a	-	translation initiation factor eIF-5A
OE3545F	28	0.86	0.94	1.12	1.07	2.17	HY	-	Questionable	hypothetical protein
OE3762R	29	1.3	1.01	0.85	0.88	0.83	CIM	glpK	Insecure	glycerol kinase (EC 2.7.1.30)
OE3798R	30	0.86	0.96	0.98	0.89	0.72	CHY	-	Questionable	conserved hypothetical protein
OE3822R	31	0.9	1.15	1.26	1.41	2.55	CHY	-	-	conserved hypothetical protein
OE3949R	32	0.88	1.05	1.29	1.16	1.37	CHY	-	Questionable	glutaredoxin homolog
OE3989R	33	1.41	0.97	0.83	0.98	0.62	EM	atpK	Trusted	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain K.a (A-type ATP synthase)
OE3992R	34	1.38	1	0.87	0.94	0.71	EM	atpH	Insecure	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain H.a (A-type ATP synthase)
OE4347F	35	0.72	0.98	1.04	1.07	1.35	HY	-	Questionable	hypothetical protein
OE4459R	36	1.16	1.07	0.89	0.98	0.7	TL	rpl31R	Trusted	ribosomal protein L31.eR
OE4460R	37	1.2	1.01	0.97	0.9	0.55	TL	rpl39R	-	ribosomal protein L39.eR
OE4511R	38	1.15	1	0.89	0.85	0.5	HY	-	-	hypothetical protein
OE4673F	39	0.97	1.13	1.25	1.21	1.56	MIS	cpx	Trusted	carboxypeptidase (EC 3.4.-.-)
OE4676F	40	0.97	0.94	1.22	1.08	2.15	HY	-	-	hypothetical protein
OE4720R	41	1.43	1.12	0.92	0.86	0.52	TL	rps10	Insecure	ribosomal protein S10
OE4735R	42	1.22	1.12	0.91	0.91	0.72	TL	rps7	Trusted	ribosomal protein S7
OE4736R	43	1.18	1.06	0.95	0.89	0.7	TL	rps12	Trusted	ribosomal protein S12
OE5125F	44	0.78	1.18	0.96	1.09	1.68	CP	gvpA2	Insecure	gas-vesicle protein gvpA2
OE5163R	45	0.71	1.06	1.16	1.13	1.78	HY	-	-	spurious ORF
OE6118F	46	0.98	1.08	1.07	1.08	1.35	HY	-	-	hypothetical protein
OE1160R	47	1.09	1.01	0.97	0.9	0.65	TL	rpl10R	Trusted	ribosomal protein L10.eR
OE1187R	48	0.9	1.03	0.87	0.95	0.7	-	-	-	spurious ORF
OE1271F	49	1.51	0.97	0.8	0.97	0.85	LIP	alkK1	Trusted	probable fatty-acid-CoA ligase (EC 6.2.1.-)
OE1756F	50	1.03	1.01	1.09	1.11	0.97	CHY	-	-	conserved hypothetical protein

Tab. 9: Liste mit den ersten 50 signifikantesten Genen des zweiten Zeitreihenexperiments.

Die sich ergebenden Auffälligkeiten bei der Verteilung der 242 differentiell exprimierten Gene des zweiten Zeitreihenexperiments auf die Funktionsklassen von *H. sal.* R1 sind in Abb. 48 dargestellt und werden in den folgenden Abschnitten näher beschrieben.

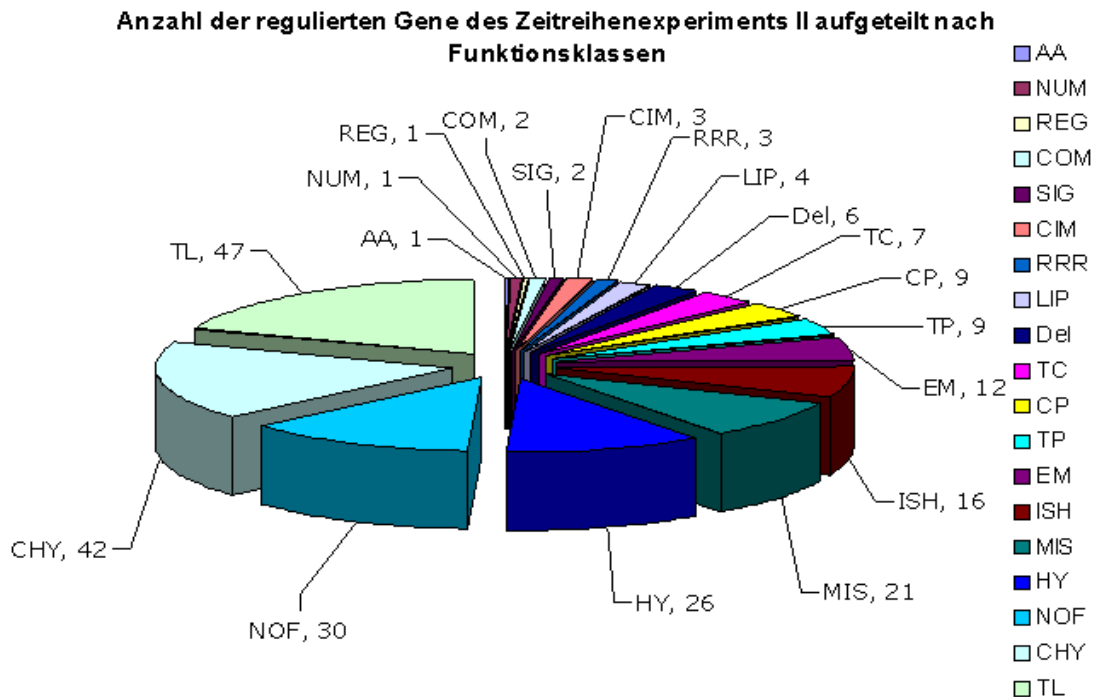


Abb. 48: Tortendiagramm der regulierten Gene des Transkriptom-Experiments der zweiten Zeitreihe. Dargestellt sind die 242 signifikantesten Gene, berechnet auf der Basis der FS-Statistik. Die Anzahl der Gene in den jeweiligen Funktionsklassen ist neben den Abkürzungen angegeben. Die Abkürzungen bedeuten: AA = Aminosäurestoffwechsel, CHM = Kohlenhydratstoffwechsel, CHY = konserviert hypothetisches Protein, CIM = zentraler Intermediatstoffwechsel, COM = Koenzym Stoffwechsel, CP = zelluläre Prozesse, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, ISH = ISH-kodierte Transposasen, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes, NOF = noch keine Funktion, NUM = Nukleotid-Stoffwechsel, REG = Genregulation, RRR = Reproduktion Reparatur Rekombination, SIG = Signalübertragung, TC = Transkription, TL = Translation, TP = Transport.

Auch für das zweite Zeitreihenexperiment kann man nach der Verteilung der signifikant differentiell exprimierten Gene auf die Funktionsklassen eine Einteilung in sechs verschiedene Gruppen durchführen.

Wiederum bilden die drei Funktionsklassen CHY, HY und NOF mit 98 Genen und einem Anteil von 40,5% die größte Gruppe. Im Vergleich zu den anderen Transkriptom-Experimenten, bei denen dieser Anteil bei 54 und 50% liegt, weist die zweite Zeitreihe damit einen geringeren Anteil an unbekanntem Proteinen auf.

Die zweite Gruppe besitzt einen Anteil von 19,4% an den signifikanten Genen. Hierbei ist allerdings besonders auffällig dass die zweite Gruppe nur aus der Funktionsklasse der Translation besteht. Ein Trend der sich vom ersten zum zweiten Zeitreihenexperiment fortgesetzt hat. Bemerkenswert ist ebenfalls der Anteil von 40 ribosomalen Proteinen. Damit zeigen 70% aller ribosomaler Proteine, die in *H. sal.* R1 die Ribosomen bilden, während des Zeitintervalls von

24h eine signifikante Änderung. Desweiteren zählen zu den 47 Kandidatengenen der Funktionsklasse TL drei Translationsinitiationsfaktoren (*eif2a*, *eif5a* und *sui1*), zwei Translationselongationsfaktoren (*aef1a* und *aef2*), eine t-RNA Ligase (Valin) und das Gen *erf1*, welches für einen möglichen Peptidkettenlösungsfaktor kodiert.

Einen Anteil von 15,3% weist die dritte Gruppe auf, die sich aus den zwei Funktionsklassen MIS mit 21 Kandidatengenen und ISH mit 16 Kandidatengenen zusammensetzt. Beispiele für Kandidatengene der MIS Funktionsklasse sind die Gene *sod2* (Superoxiddismutase 2), *blp* (Bakterioopsin-verbundenes Protein *blp*), *dpsA* (DNA-bindendes Protein), *grpE* (*dabJ/dnaK* ATPase Stimulator) oder das *flaCE* (flagella accessory protein).

Mit 17,8% an den differentiell exprimierten Genen besitzt die vierte Gruppe der Funktionsklassen nur einen unwesentlich größeren Anteil als die dritte Gruppe. Unter den vier Funktionsklassen der vierten Gruppe haben die Funktionsklasse des Energiemetabolismus mit 12 Kandidatengenen den größten Anteil von 5%, während die beiden anderen Funktionsklassen TP und CP mit je 9 Kandidatengenen einen Anteil von 3,7% und TC mit 7 Genen einen Anteil von 2,9% besitzen. Zu den Kandidatengenen des Energiemetabolismus zählt unter anderem das *bop* (Bakteriorhodopsin-Precursor) Gen, sechs Untereinheiten der ATP-Synthase (*atpB,D,E,F,H,K*), die Ketten I und II der Cytochrom-C-Oxidase (*cox1B* und *cox2A*), die Ketten CD und M der NADH-Dehydrogenase (*nuoCD* und *nuoM*) und das Ferredoxin *fer 5*. Neben dem *hop* (Halorhodopsin) Gen, findet man hauptsächlich ABC-Transportproteine unter den Kandidatengenen der Transportfunktionsklasse. Ein weiterer Beweis für die Zuverlässigkeit der Transkriptom-Experimente zeigt sich, wenn man sich die Mitglieder der Funktionsklasse CP näher ansieht. Denn dort findet man analog zum ersten Zeitreihenexperiment das Chaperon *cspD1*, ein Kälteschock-Protein und die zwei Untereinheiten *cctA* und *cctB* des Thermosoms. Zusätzlich dazu befindet sich auf der Liste der differentiell exprimierten Gene der zweiten Zeitreihe das *cspD2*, ein weiteres Kälteschock-Protein, das *hsp1* Gen, ein Hitzeschock-Protein, und mit dem *psmB* Gen eine Untereinheit des Proteasoms. Weiterhin beachtlich ist, dass sieben Untereinheiten (*rpoA1,D,E1,F,K,M1* und *N*) der RNA-Polymerase zu den differentiell exprimierten Genen der Funktionsklasse TC gehören, was bei 15 Untereinheiten der RNA-Polymerase in *H. sal.* R1 einem Anteil von 47% entspricht (HaloLex-Datenbank; Werner F, et al., 2000).

In der fünften Gruppe mit zwei bis vier Kandidatengenen, werden folgende fünf Funktionsklassen zusammengefasst, der Lipidstoffwechsel (LIP), die Reproduktion, Reparatur und Rekombination (RRR), der zentrale Intermediatmetabolismus (CIM), die Signaltransduktion (SIG) und der Coenzymstoffwechsel (COM). Zu den Kandidaten der Signaltransduktion gehört der Antwortregulator *cheY* und der mögliche Antwortregulator *hrg*. Für die Funktionsklasse RRR ist mit *nthA1* eine DNA-Ligase, mit *mutS* ein DNA „mismatch repair protein“ und mit *polB2* eine Typ-II DNA-Polymerase zu finden. Die Gruppe des Intermediatstoffwechsels wird gebildet von der Glycerin Kinase (*glpK*), der Phosphoenolpyruvat Kinase (*ppsA*) und der alpha Kette der Oxoglutarat-Ferridoxin-Oxidoreduktase (*korA*). Repräsentanten für den Lipidstoffwechsel sind die Phytoendehydrogenase *crtI2* und zwei wahrscheinliche Acyl/butyryl-CoA Dehydrogenasen *alkK1* und *acd3*. Die Kandidaten des Coenzymstoffwechsels sind mit dem *cbiF* ein Biosynthesegen des Pterocorrins und mit dem *hmcA* und dem *cobS* zwei Cobalamin-Biosynthesegene.

Die letzte und sechste Gruppe mit je einem Vertreter wird von den Funktionsklassen Aminosäurenstoffwechsel (AA), der Genregulation (REG) und dem Nukleotidstoffwechsel (NUM) gebildet.

Die differentiell exprimierten Gene des dritten Transkriptom-Experiments teilen sich damit auf 19 unterschiedliche Funktionsklassen auf, während die fünf Klassen des Kohlenhydratstoffwechsels (CHM), der RNA-Reifung (RMT), der Motilität (MOT), der Zellhülle (CE) und der Chaperone (CHP) fehlen.

### 5.3.1 Diskussion

Die Zielsetzung dieser Experimente ist es an Hand von verschiedenen Intervallen ein zeitaufgelöstes Bild des Expressionsverlaufs der Anpassung von *H. sal.* R1 beim Wechsel vom aeroben auf phototrophes Wachstum zu bekommen (siehe 5.3).

Für das erste Zeitreihenexperiment wurde ein Zeitintervall von 120 min und für das zweite ein Intervall von 24 h gewählt. Durch die unterschiedliche Dauer der Zeitintervalle könnte man sowohl schnelle als auch langsamere Anpassungsprozesse identifizieren (siehe 3.2.2.1.2).

Für die beiden Zeitreihenexperimente bzw. das zweite und dritte Transkriptomexperiment wurde ebenfalls das „common reference design“ gewählt

(siehe Tab. 6; Abb. 19 / 20). Die Referenz bestand also aus einer Mischung aller RNA-Proben des jeweiligen Transkriptomexperiments. Ein Vorteil dieses Experimentdesigns für eine Zeitreihe besteht darin, dass so alle Zeitpunkte untereinander vergleichbar sind und sich nicht z.B. auf den Vergleich mit dem Zeitpunkt  $t_0$  beziehen. Man kann so eine Aussage über alle Zeitpunkte des jeweiligen Experiments treffen. Außerdem ist man in der Lage auf der Basis der Datenanalyse weitere statistische Analysemethoden anzuwenden.

Entsprechend des ersten Transkriptomexperiments, dem Zweigruppenvergleich, beruhen die Listen der differentiell exprimierten Gene der Zeitreihenexperimente auch auf der statistischen Auswertung mittels des FS-Hypothesentests. Die Kandidatengenlisten fassen diejenigen Gene nach ihrem statistischen Rang zusammen, die über die Zeit signifikante Expressionsunterschiede aufweisen. Die Information über den zeitlichen Expressionsverlauf der Gene kommt durch die Angabe des Expressionsniveaus von einem Zeitpunkt zum anderen hinzu.

Auch die Zeitreihenexperimente wurden in einem „unsupervised“ Verfahren durchgeführt und bestätigen mit ihren Ergebnissen frühere Versuche. So beweisen alle drei unabhängig von einander durchgeführten Transkriptomexperimente die Zuverlässigkeit der für *H. sal.* R1 etablierten Methode.

Alle drei Transkriptomexperimente weisen einen hohen Anteil von regulierten unbekanntem Genen auf, jedoch nimmt dieser vom ersten mit 54,5% bis zum letzten Experiment mit 40,5% stetig ab. Neben dieser Auffälligkeit ist weiterhin festzustellen, dass eine deutliche Verschiebung der Anzahl der Gene bei der Aufteilung in die einzelnen Funktionsklassen zu beobachten ist.

Waren im ersten Transkriptomexperiment die Kandidatengene ungefähr gleich auf die Funktionsklassen verteilt, mit Ausnahme der Funktionsklassen CHY, HY und NOF, so steigt der prozentuale Anteil der Funktionsklasse TL (Translation) von 1,6% über 10,3% auf 19,4% für das dritte Transkriptomexperiment.

Was dieser hohe Anteil besagt, wird deutlich wenn man weiß, dass die 19,4% 40 ribosomalen Proteinen entsprechen, die ihrerseits 70% aller ribosomalen Proteine von *H. sal.* R1 repräsentieren.

Dieses Ergebnis steht auch beispielhaft dafür, dass durch die Kombination der Auswertung der drei Transkriptomexperimente die Zielsetzung der zeitaufgelösten Expressionsanalyse erfüllt worden ist. Denn allein durch den Zweigruppenvergleich aerob / phototroph konnten diese Gene nicht als regulierte

Gene gefunden werden, da der Hauptanteil der 40 regulierten ribosomalen Proteine die höchste Induktion am Zeitpunkt  $t_3$  der zweiten Zeitreihe aufweist, was 180 min entspricht. Dies ist auch die Erklärung dafür, dass der Anteil der regulierten ribosomalen Proteine für die erste Zeitreihe geringer ist, denn das gesamte Zeitintervall dieses Experiments betrug 120 min. Dadurch wird verständlich, dass es wichtig war zwei verschiedene Zeitintervalle für die Transkriptomexperimente durchzuführen.

Wie oben bereits erwähnt, erlaubt die gewählte Datenanalyse und das Experimentdesign der Transkriptomversuche eine weiterführende Analyse der Daten. Diese zusätzliche Auswertung könnte es ermöglichen eine genauere Deutung der Ergebnisse zu bekommen.

Als ein geeignetes Hilfsmittel zur weiteren Interpretation des zeitlichen Verlaufs der Expressionsanalyse des Zeitreihenexperiments ist die Durchführung einer Hauptkomponentenanalyse (**P**ro**u**nc**i**pal **C**omponent **A**nalysis, PCA) auf der Basis der statistisch ausgewerteten Hybridisierungsergebnisse. Die PCA stellt eine exploratorische Methode dar, die es ermöglicht in den komplexen Datensätzen der DNA-Mikroarray-Experimente vorhandene Muster aufzufinden. Am Ende einer PCA steht eine Dimensionsreduktion, die es erleichtert den Zusammenhang zwischen den Variablen überschaubarer und damit interpretierbarer zu machen. Hierbei sind die Variablen die Gene und Zeitpunkte des Genexpressionsexperiments. Hauptkomponentenanalysen wurden bereits erfolgreich zur Analyse von Expressionsdaten eingesetzt (Landgrebe J, et al., 2002; de Haan RJ, et al., 2006).

In einem DNA-Mikroarrayexperiment bedeutet also eine Dimensionsreduktion durch eine PCA das Aufspüren und Beseitigen der Redundanz in den Daten, um das Rauschen zu minimieren und Ausreisser zu identifizieren, die von Interesse sein könnten. Es werden die Gene herausgefiltert, deren Expressionswerte sich im Verlauf des Experiments nicht ändern, damit es erleichtert wird, solche Gene zu detektieren, deren Expressionswerte starke Unterschiede aufweisen.

Im Verlauf einer PCA werden die Variablen durch lineare Kombination zu den so genannten Hauptkomponenten zusammengefasst. Veranschaulicht dargestellt wird die Methode dadurch, dass im ersten Schritt der Ursprung eines Koordinatensystems in den Mittelpunkt der jeweiligen Datenwolke gelegt wird. Danach wird das Koordinatensystem so lange rotiert, bis auf der 1. Achse (oder 1. Hauptkomponente) die maximal mögliche Varianz in den Daten erfasst wird. In den nachfolgenden Schritten wird dieses Verfahren für die weiteren Achsen bzw.

Hauptkomponenten in analoger Weise durchgeführt. Es ist dabei darauf zu achten, dass die Achsen jeweils orthogonal aufeinander stehen, um der Voraussetzung der Unabhängigkeit der Variablen Rechnung zu tragen. Jede Achse (Hauptkomponente) repräsentiert sukzessiv weniger von der Varianz in den Daten, so dass durch die beiden ersten Hauptkomponenten der Hauptanteil der Varianz abgedeckt werden.

Der Startpunkt einer PCA ist eine Datenmatrix aus  $m$  Expressionswerten (Beobachtungen) und  $n$  Zeitpunkten (Variablen). Das Ziel ist es durch Finden neuer Variablen  $r$  ( $r < n$ ) die Dimensionalität der Datenmatrix zu reduzieren, wobei gilt dass  $r$  die maximale Varianz der orthogonalen Datenmatrix widerspiegelt. Vor Beginn einer PCA sind zwei Datentransformationen wichtig. Das Zentrieren der Daten, wobei von jedem Element der Mittelwert abgezogen wird, so dass die Mittelwerte eines Objekts Null werden. Das Standardisieren der Daten, worunter man die Division jedes Elements mit der Standardabweichung versteht und folglich die Standardabweichung gleich eins wird.

Im Rahmen dieser Arbeit sind die statistisch ausgewerteten DNA-Mikroarraydaten der zweiten Zeitreihe zur Interpretation des Expressionsverlaufs einer PCA unterzogen worden.

Als Endprodukt der PCA erhält man eine zwei- oder dreidimensionale grafische Darstellung der linearen Projektion der ursprünglichen Datenmatrix auf den niedriger dimensionierten Raum. Sind zwei Punkte bzw. Gene in dieser grafischen Darstellung benachbart, so zeigen sie einen ähnlichen Expressionsverlauf. Umgekehrt weist ein großer Abstand zweier Gene auf einen größeren Unterschied im Expressionsverhalten hin. Der große Vorteil einer solchen PCA und der anschließenden grafischen Darstellung in einem so genannten Biplot ist, dass eine beliebige Anzahl von Experimenten gemeinsam dargestellt werden können. So stehen Vektoren bzw. Pfeile in einem Biplot für z.B. verschiedene Versuche, oder wie für dieses Projekt, für verschiedene Zeitpunkte. Die Korrelation zwischen den einzelnen Vektoren ist durch den Kosinus des Winkels zwischen den Vektoren gegeben und die Länge der Vektoren repräsentiert die Standardabweichung. Im Idealfall würden sich die Vektoren einzelner Wiederholungen des gleichen Versuchs überlagern. Der Biplot mit den Ergebnissen für die zweite Zeitreihe ist in Abb. 49 gezeigt. Analog zum großen Potential dieser Analyseverfahren sind in dem Biplot der zweiten Zeitreihe alle 30 durchgeführten Hybridisierungen dargestellt, entsprechend der fünf Wiederholungen für jeden der sechs Zeitpunkte. Ein sehr beachtliches Ergebnis





Ein so genannter Matrixplot ist eine andere Art die Ergebnisse einer PCA grafisch darzustellen. In einem solchen Matrixplot werden den Zahlenwerten ein Farbkode zugeordnet und in dem Graph dargestellt. Der Vorteil dieser exploratorischen Darstellung liegt darin, dass man so sehr gut zeitliche Verläufe aufzeigen kann. Aus diesem Grund wurde das Ergebnis der PCA der zweiten Zeitreihe als Matrixplot dargestellt, in dem die Zeilen des Plots den Genen entsprechen und die Spalten den Zeitpunkten. Zur vereinfachten Darstellung wurden die Ergebnisse der fünf Wiederholungen jedes Zeitpunktes zu einem Wert zusammengefasst, so dass in dem folgenden Matrixplot sechs statt 30 Spalten zu sehen sind. Beim Farbkode entspricht grün einer Repression und rot einer Induktion des Gens. Die Werte dazwischen werden durch farbliche Abstufungen angegeben.

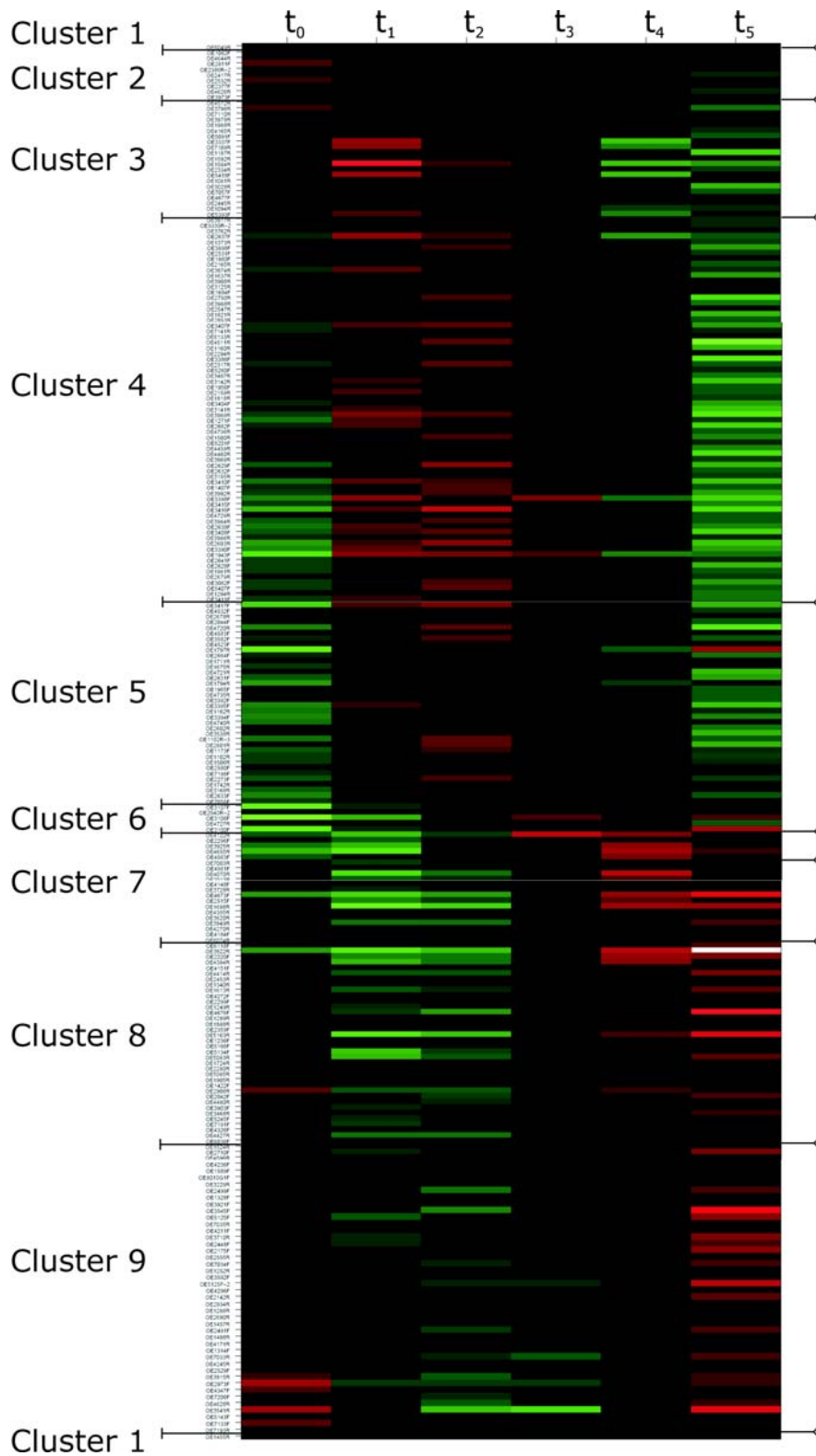


Abb. 50: Matrixplot nach einer PCA auf der Basis des Datensatzes des dritten Transkriptom-Experiments. Links sind die Angaben der Einteilung in verschiedene Cluster und die Gennamen (Zeilen). Oben sind die verschiedenen Zeitpunkte des Versuchs angegeben (Spalten), wobei  $t_0$  der Ausgangspunkt ist und  $t_1= 45\text{min}$ ,  $t_2= 90\text{min}$ ,  $t_3= 180\text{min}$ ,  $t_4= 360\text{min}$  und  $t_5= 1440\text{min}$  die Zeitpunkte der Probennahme unter phototrophen Wachstumsbedingungen markieren. Eine Induktion des Gens wird durch rot repräsentiert und eine Repression durch grün.

Diese Form der Ergebnisdarstellung erlaubt es nun, anhand des Farbverlaufs das Expressionsverhalten eines Gens während des untersuchten Zeitintervalls sichtbar zu machen. Mit der Information der jeweiligen Expressionsverläufe können die Gene mit einem ähnlichen Verlauf in Gruppen, so genannte Cluster, eingeteilt werden. Die sich ergebenden Cluster nach der PCA für die zweite Zeitreihe sind in der Abb. 50 eingezeichnet.

Die eingeteilten neun Cluster werden in den folgenden Abschnitten näher beschrieben.

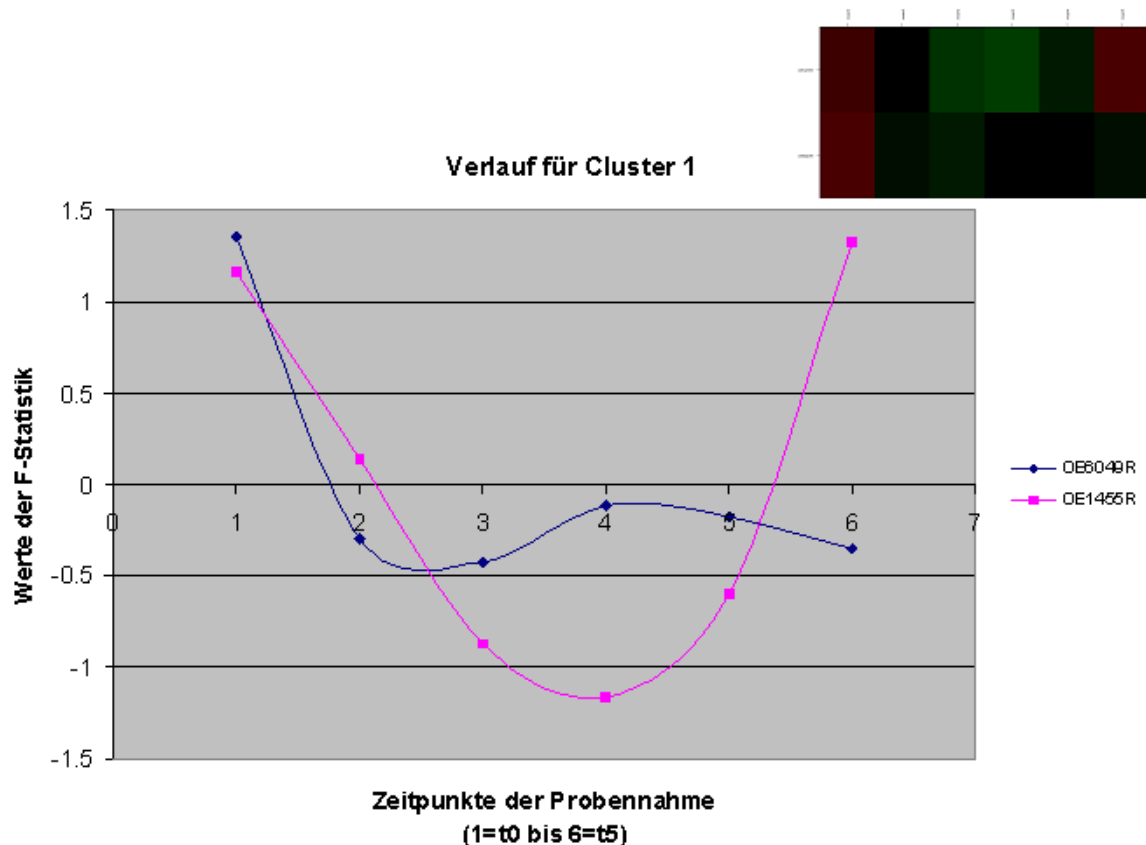


Abb. 51: Gezeigt wird der Expressionsverlauf der Gene des Clusters 1 anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0= 0\text{min}$ ,  $t_1= 45\text{min}$ ,  $t_2= 90\text{min}$ ,  $t_3= 180\text{min}$ ,  $t_4= 360\text{min}$ ,  $t_5= 1440\text{min}$ ). Weitere Informationen zu den Genen siehe folgende Tabelle.

OE-Nummer	Funktionsklas:	Genname	Proteinname	Protein-Identifizierung
OE6049R	CHY	-	conserved hypothetical protein	Trusted
OE1455R	NOF	-	conserved protein	Trusted

Tab. 8: Die im Cluster 1 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Aus der Abb. 51 ist der genaue Verlauf des Expressionsniveaus und die Mitglieder des Clusters eins zu entnehmen. Die beiden Mitglieder des ersten Clusters sind mit OE6049R ein „conserved hypothetical protein“ und mit OE1455R ein „conserved protein“ der Funktionsklasse NOF. Der Expressionsverlauf von

OE1455R ist durch einen starken Abfall bis zum Zeitpunkt  $t_3 = 180\text{min}$  und einen Anstieg auf das Ausgangsniveau gekennzeichnet. Das Gen OE6049R könnte aufgrund seines Verlaufs auch in den Cluster 2 gehören, der einen etwas wellenförmigen Verlauf zeigt.

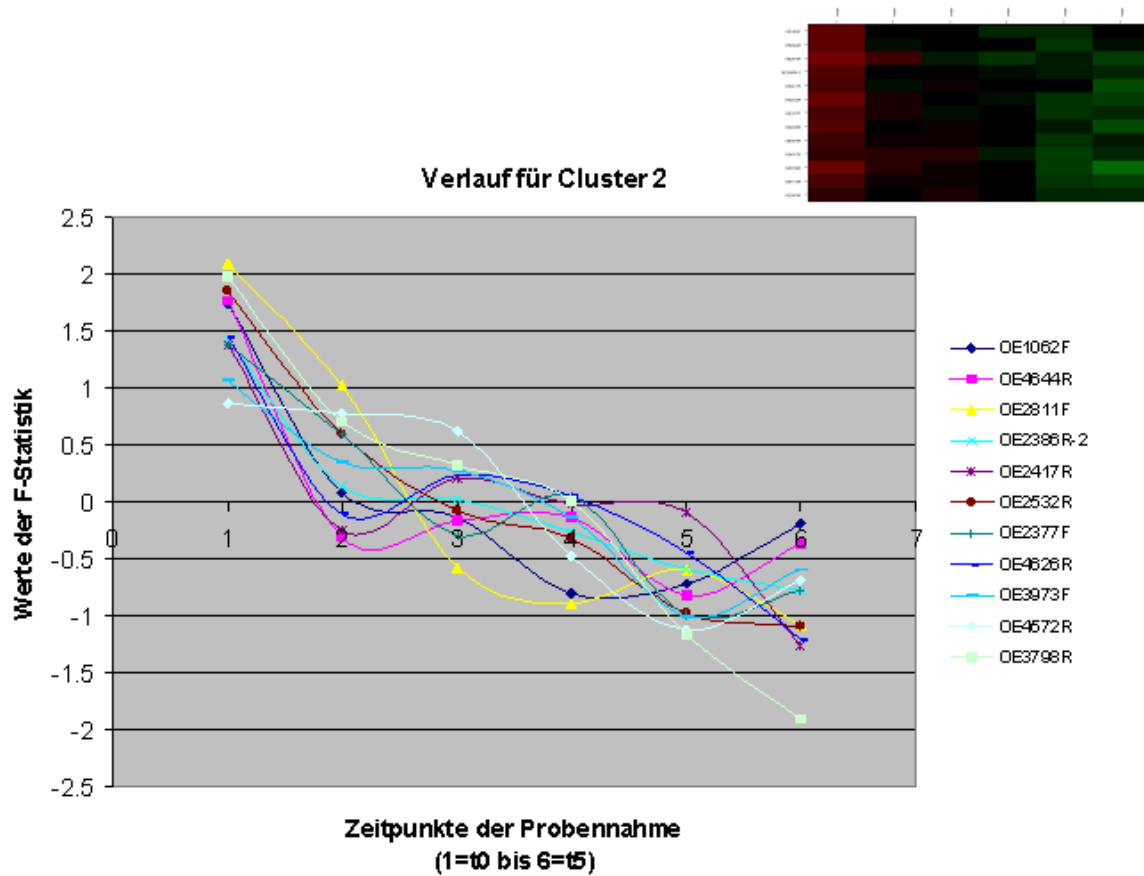


Abb. 52: Gezeigt wird der Expressionsverlauf der Gene des Clusters zwei anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0 = 0\text{min}$ ,  $t_1 = 45\text{min}$ ,  $t_2 = 90\text{min}$ ,  $t_3 = 180\text{min}$ ,  $t_4 = 360\text{min}$ ,  $t_5 = 1440\text{min}$ ). Weitere Informationen zu den Genen siehe folgende Tabelle.

OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE3798R	CHY	-	conserved hypothetical protein	Trusted
OE2377F	CHY	-	conserved hypothetical protein	-
OE2811F	CHY	-	conserved hypothetical protein	-
OE1062F	HY	-	hypothetical protein	-
OE2386R	MIS	flaCE	flagella accessory protein	Questionable
OE4644R	NOF	-	conserved protein	Trusted
OE2532R	NOF	-	conserved protein	Trusted
OE2417R	SIG	cheY	response regulator cheY	Trusted
OE4626R	TL	sui1	probable translation initiation factor SUI1	Trusted
OE4572R	TL	valS	valine--tRNA ligase (EC 6.1.1.9)	Trusted
OE3973F	TL	erf1	peptide chain release factor eRF-1	Questionable

Tab. 9: Die im Cluster 2 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Die Abbildung für den Cluster 2 zeigt deutlich, dass die aufgeführten Mitglieder einen periodischen Verlauf ihrer Expression aufweisen. Nach einem Abfall bis  $t_1 =$

15min folgt ein Anstieg bis  $t_3= 180\text{min}$  und ein erneuter Abfall zu  $t_4= 360\text{min}$  mit anschließendem Anstieg zu  $t_5= 1440\text{min}$ . Der Anstieg bis zum 24h Zeitpunkt erreicht bei allen Genen nicht den Ausgangswert, so dass netto eine Repression der Gene vom Wechsel des aeroben auf das phototrophe Wachstum stattfindet. Zu den Mitgliedern dieses Cluster gehören *cheY*, ein Response Regulator, *sui1*, ein möglicher Translationsinitiationsfaktor und *flaCE*, ein „flagella accessory protein“.

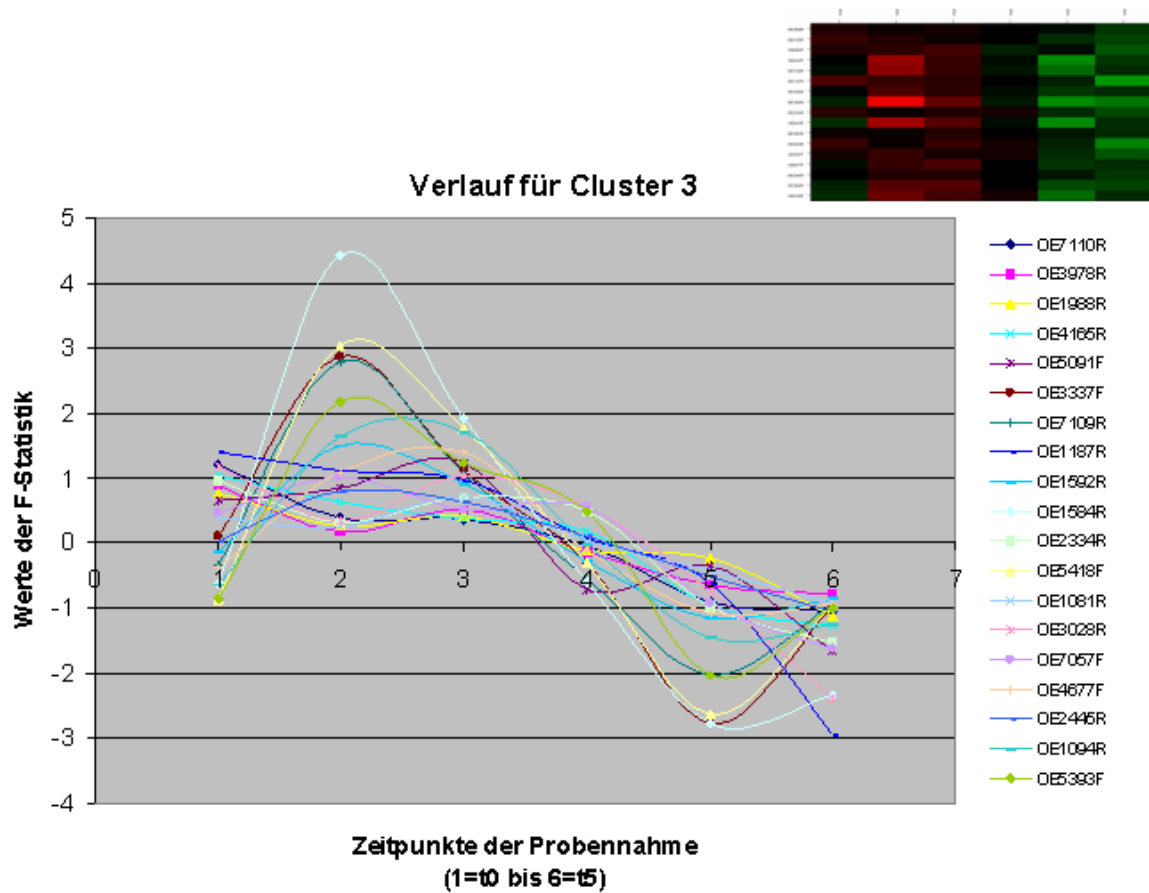


Abb. 53: Gezeigt wird der Expressionsverlauf der Gene des Clusters drei anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0= 0\text{min}$ ,  $t_1= 45\text{min}$ ,  $t_2= 90\text{min}$ ,  $t_3= 180\text{min}$ ,  $t_4= 360\text{min}$ ,  $t_5= 1440\text{min}$ ). Weitere Informationen zu den Genen siehe folgende Tabelle.

## Ergebnisse und Diskussion

OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE1187R	-	-	[del] spurious ORF	-
OE7057F	CHY	-	conserved hypothetical protein	Questionable
OE2445R	CHY	-	phage PhiH1 repressor protein homolog	-
OE3978R	EM	atpD	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain D.a (A-type ATP synthase)	Trusted
OE1988R	EM	cox2A	cytochrome-c oxidase (EC 1.9.3.1) chain II	-
OE5418F	ISH	-	insertion element protein (ISH2)	-
OE1094R	ISH	-	probable transposase (ISH10)	Questionable
OE3337F	ISH	-	insertion element protein (ISH2)	-
OE7109R	ISH	-	insertion element protein (ISH2)	-
OE4677F	MIS	uvrD	DNA helicase II	Trusted
OE1081R	MIS	rfbU1	probable glycosyltransferase (EC 2.-.-)	Trusted
OE5091F	NOF	-	conserved protein	Trusted
OE4165R	NOF	pcn	DNA-directed DNA polymerase sliding clamp homolog	Trusted
OE7110R	NOF	-	conserved protein (nonfunctional, N-terminal part, interrupted by ISH2)	Trusted
OE1592R	NOF	epf2	mRNA 3'-end processing factor homolog	Trusted
OE1584R	NOF	-	conserved protein	Trusted
OE3028R	NOF	-	conserved protein	Trusted
OE2334R	SIG	hrg	probable response regulator / probable transcription regulator	-
OE5393F	ISH	-	insertion element protein (ISH2)	-

Tab. 10: Die im Cluster 3 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Wie die Mitglieder des Cluster 2, zeigen die Gene des Cluster 3 ebenfalls einen periodischen Verlauf, mit dem Unterschied, dass erst mit einem Anstieg begonnen wird. Man sieht deutlich in Abb. 54 für  $t_1 = 15\text{min}$  ein Maximum, den Wendepunkt zwischen  $t_2 = 90\text{min}$  und  $t_3 = 180\text{min}$  und das Minimum an  $t_4 = 360\text{min}$  gefolgt von einem Anstieg zum Zeitpunkt  $t_5 = 1440\text{min}$ . Analog zum Cluster 2 zeigt sich hier auch ein netto Abfall der Expression. Beispiele für Mitglieder dieses Clusters sind das Gen *atpD*, die Kette D der Typ A ATP-Synthase, *cox2A*, die Kette II der Cytochrome-C-Oxidase und *hrg*, der wahrscheinliche Response Regulator.

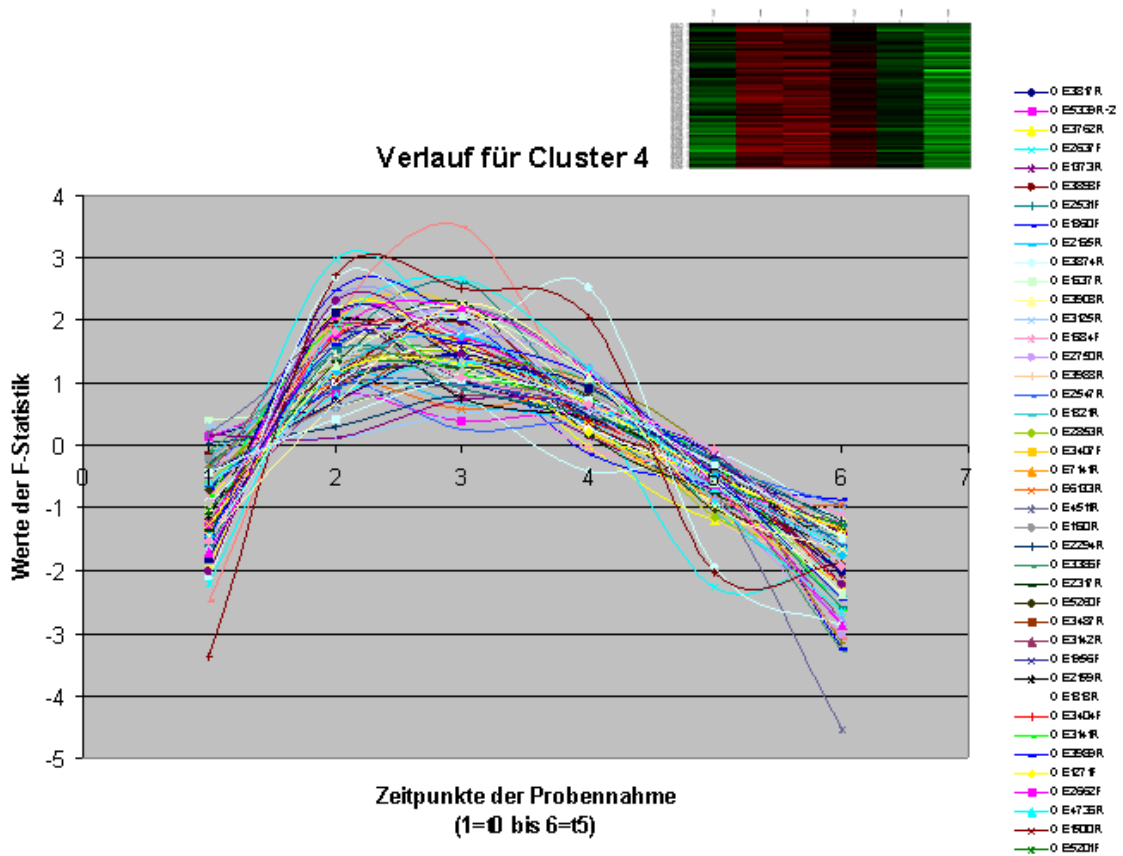


Abb. 54: Gezeigt wird der Expressionsverlauf der Gene des Clusters vier anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0= 0\text{min}$ ,  $t_1= 45\text{min}$ ,  $t_2=0 90\text{min}$ ,  $t_3= 180\text{min}$ ,  $t_4= 360\text{min}$ ,  $t_5= 1440\text{min}$ ). Weitere Informationen zu den Genen siehe Tabelle 11.

Der Expressionsverlauf der Mehrzahl der Gene des Clusters 4 lässt ein Maximum zum Zeitpunkt  $t_2= 90\text{min}$  erkennen. Ein eindeutiger Trend der netto Expression ist nicht auszumachen, aber man kann eine leicht geringere Expressionsrate für den den 24h-Wert annehmen.

Eine große Auffälligkeit in der Zusammensetzung der Mitglieder des Clusters 4 ist darin zu sehen, dass hier 29 ribosomale Proteine enthalten sind.



## Ergebnisse und Diskussion

OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE1943F	AA	gdhA1	glutamate dehydrogenase (NADP+) (EC 1.4.1.4)	Trusted
OE1860F	CHY	-	conserved hypothetical protein	Questionable
OE2531F	CHY	-	conserved hypothetical protein	-
OE3125R	CHY	-	conserved hypothetical protein	Questionable
OE3398F	CHY	-	conserved hypothetical protein	-
OE2294R	CHY	-	conserved hypothetical protein	Trusted
OE7141R	CHY	-	multidrug resistance transport protein homolog	-
OE1500R	CIM	ppsA	pyruvate, water dikinase (EC 2.7.9.2) (phosphoenolpyruvate synthase)	Trusted
OE3762R	CIM	glpK	glycerol kinase (EC 2.7.1.30)	Insecure
OE3986R	EM	atpF	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain F.a (A-type ATP synthase)	Trusted
OE3989R	EM	atpK	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain K.a (A-type ATP synthase)	Trusted
OE3988R	EM	atpE	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain E.a (A-type ATP synthase)	Trusted
OE1956F	EM	nuoCD	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CD	Trusted
OE3984R	EM	atpB	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain B.a (A-type ATP synthase)	Trusted
OE3992R	EM	atpH	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain H.a (A-type ATP synthase)	Insecure
OE1981R	HY	-	hypothetical protein	-
OE4511R	HY	-	hypothetical protein	-
OE5260F	ISH	-	probable transposase (ISH10)	-
OE6133R	ISH	-	probable transposase (ISH10)	-
OE5339R	ISH	-	probable transposase (ISH1)	-
OE5407F	ISH	-	probable transposase (ISH8/ISH26)	-
OE1271F	LIP	alkK1	probable fatty-acid--CoA ligase (EC 6.2.1.-)	Trusted
OE1684F	MIS	nodP	probable sulfate adenyltransferase (EC 2.7.7.4) small chain	Trusted
OE3142R	MIS	snp	SM protein	Questionable
OE2853R	MIS	slyD	probable peptidylprolyl isomerase (EC 5.2.1.8)	Insecure
OE1407F	MIS	ipp	inorganic pyrophosphatase (EC 3.6.1.1)	Insecure
OE2547R	MIS	exoM	probable glycosyltransferase (EC 2.-.-.-)	Questionable
OE3155R	NOF	-	conserved protein	Trusted
OE2750R	NOF	-	conserved protein	Trusted
OE3386F	NOF	-	conserved protein	Insecure
OE5201F	NUM	pyrB	aspartate carbamoyltransferase (EC 2.1.3.2) catalytic chain	Trusted
OE2638F	TC	rpoK	DNA-directed RNA polymerase (EC 2.7.7.6) chain K	-
OE1637R	TC	rpoM1	probable DNA-directed RNA polymerase (EC 2.7.7.6) chain M	-
OE3874R	TC	rpoE1	DNA-directed RNA polymerase (EC 2.7.7.6) chain E'	Questionable
OE2637F	TC	rpoN	DNA-directed RNA polymerase (EC 2.7.7.6) chain N	Trusted
OE1294R	TL	rpl15R	ribosomal protein L15.eR	Trusted
OE3411F	TL	rpl6	ribosomal protein L6	Trusted
OE3415F	TL	rps5	ribosomal protein S5	Trusted
OE3062F	TL	rps17R	ribosomal protein S17.eR	Trusted
OE3416F	TL	rpl30	ribosomal protein L30	Insecure
OE4729R	TL	aef2	translation elongation factor aEF-2	Trusted
OE2603R	TL	rpl11	ribosomal protein L11	Trusted
OE2641F	TL	rps2	ribosomal protein S2	Trusted
OE3390F	TL	rpl23	ribosomal protein L23	Trusted
OE2679R	TL	rpl21R	ribosomal protein L21.eR	Trusted
OE3408F	TL	rps14	ribosomal protein S14	-
OE2628F	TL	rps4	ribosomal protein S4	Trusted
OE3410F	TL	rps8	ribosomal protein S8	Trusted
OE1818R	TL	eif2a	translation initiation factor aIF-2 alpha chain	Trusted
OE1160R	TL	rpl10R	ribosomal protein L10.eR	Trusted
OE3407F	TL	rpl5	ribosomal protein L5	Believed
OE2159R	TL	rps3aR	ribosomal protein S3a.eR	Trusted
OE3487R	TL	eif5a	translation initiation factor aIF-5A	Trusted
OE1821R	TL	rpl36aR	ribosomal protein L36a.eR (HLA)	-
OE1373R	TL	rpl37aR	ribosomal protein L37a.eR	Trusted
OE3817R	TL	rps19R	ribosomal protein S19.eR	Trusted
OE2165R	TL	rps15	ribosomal protein S15	Trusted
OE4460R	TL	rpl39R	ribosomal protein L39.eR	-
OE4459R	TL	rpl31R	ribosomal protein L31.eR	Trusted
OE4736R	TL	rps12	ribosomal protein S12	Trusted
OE2632F	TL	rpl18R	ribosomal protein L18.eR	Trusted
OE2629F	TL	rps11	ribosomal protein S11	Trusted
OE3869R	TL	rps24R	ribosomal protein S24.eR	Insecure
OE3141R	TL	rpl37R	ribosomal protein L37.eR	Trusted
OE3404F	TL	rpl24	ribosomal protein L24	Trusted
OE3898F	TL	rpl40R	ribosomal protein L40.eR	Questionable
OE2662F	TL	rpl7aR	ribosomal protein L7a.eR/HS6 (isolated from the small ribosomal subunit)	Trusted
OE3908R	TP	phnC	ABC-type transport system ATP-binding protein	Insecure
OE2317R	TP	-	probable ABC-type transport system periplasmic substrate-binding protein	Trusted

Tab. 11: Die im Cluster 4 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Die 29 ribosomalen Proteine plus der beiden Translationsinitiationsfaktoren *eif2a* und *eif5a* und dem Translationelongationsfaktor *aef2* repräsentieren 68% der Gene die für die Funktionsklasse der Translation des zweiten Zeitreihenexperiments als differentiell exprimiert gefunden wurden. Neben den ribosomalen Proteinen weisen auch die fünf Untereinheiten *atpB,E,F,H* und *K* der

ATP-Synthase sowie die vier Untereinheiten *rpoE1, K, M1* und *N* der halobakteriellen RNA-Polymerase das gleiche Expressionsverhalten auf.

Mit insgesamt 69 Genen stellt der Cluster 4 die größte Gruppe von Genen mit einem ähnlichen Expressionsverlaufs dar.

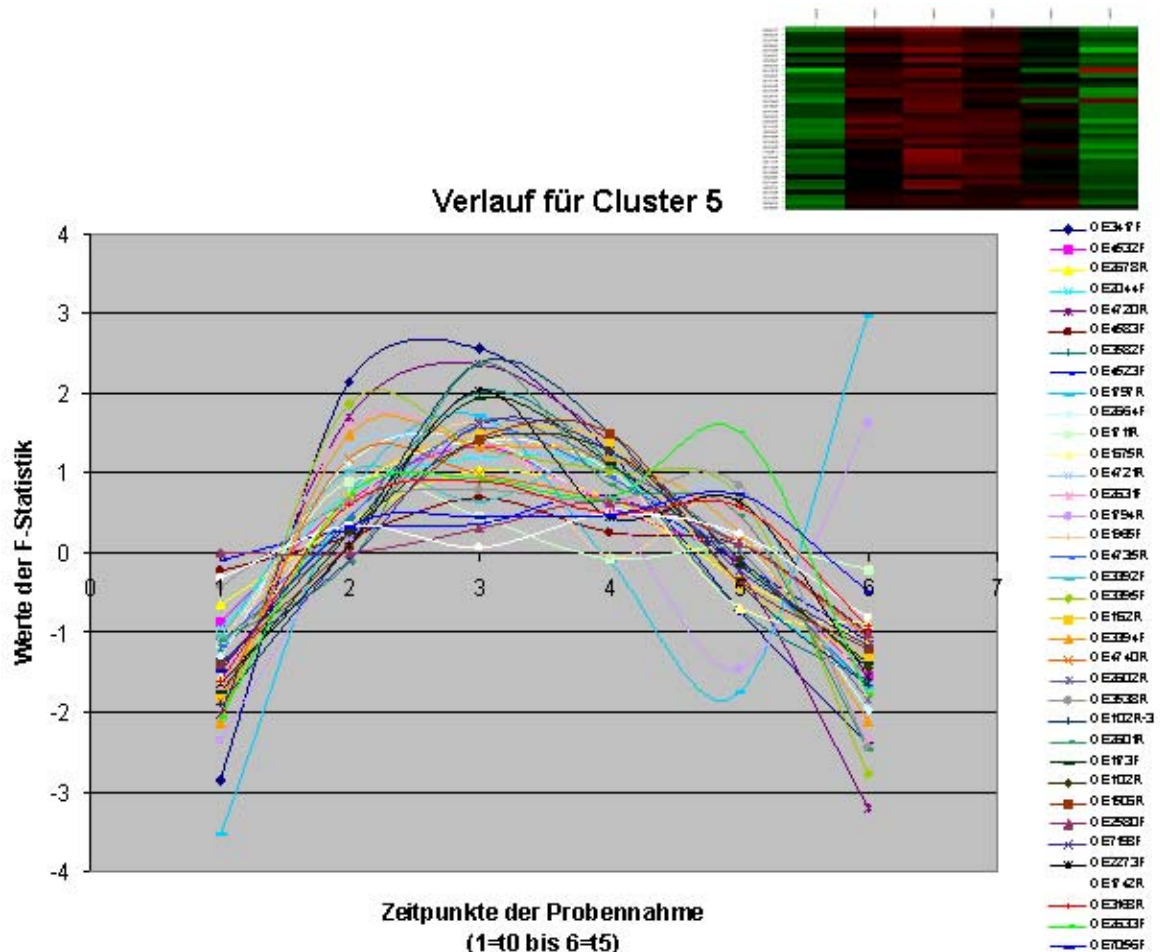


Abb. 55: Gezeigt wird der Expressionsverlauf der Gene des Clusters fünf anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0= 0\text{min}$ ,  $t_1= 45\text{min}$ ,  $t_2= 90\text{min}$ ,  $t_3= 180\text{min}$ ,  $t_4= 360\text{min}$ ,  $t_5= 1440\text{min}$ ). Weitere Informationen zu den Genen siehe Tabelle 12.

Bei dem Verlauf der Expression der Gene des Clusters 5 ist auffällig, dass er mit einem unterschiedlichen Maximum beim Zeitpunkt  $t_2= 90\text{min}$  einen fast identischen Verlauf zeigt, wie der Cluster 4. Zwar ist der Cluster 5 mit einer Gesamtzahl von 36 Genen offensichtlich kleiner, aber mit 11 Genen der Funktionsklasse der Translation ist dies auch hier die größte Gruppe innerhalb des Clusters 5. So zeigen die ribosomalen Proteine des Clusters 4 und 5 ein sehr ähnliches Expressionsverhalten und bilden zusammen 70% aller Proteine des Ribosoms von *H. sal.* R1. Fasst man die drei Untereinheiten *rpoA1, D* und *F* der

RNA-Polymerase mit den in Cluster 4 gefundenen vier Unterheiten des Komplexes zusammen, werden sieben der insgesamt 15 Untereinheiten mit einem ähnlichen Expressionsverhalten identifiziert. Dies kann als offensichtlicher Hinweis auf die Neuorganisation des Proteininventars im Hinblick auf die geänderten äußeren Umweltbedingungen der Zellen von *H. sal.* R1 verstanden werden. Die Zusammenstellung aller Mitglieder des Clusters 5 sind in nachfolgender Tabelle zu finden.

OE-Nummer	Funktionsklasse	Gencode	Proteinname	Protein-Identifizierung
OE7056F	-	-	[del] spurious ORF	-
OE1794R	CHY	-	conserved hypothetical protein	-
OE2044F	CHY	-	conserved hypothetical protein	Trusted
OE2580F	CHY	-	conserved hypothetical protein	-
OE1711R	CIM	korA	oxoglutarate--ferredoxin oxidoreductase (EC 1.2.7.3) alpha chain	Trusted
OE1162R	CP	cspD1	probable cold shock protein	Trusted
OE3582F	CP	cspD2	probable cold shock protein	-
OE1965F	EM	nuoM	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain M	Trusted
OE4583F	EM	fer5	ferredoxin (2Fe-2S)	Trusted
OE2273F	HY	-	hypothetical protein	-
OE3538R	HY	-	hypothetical protein	Questionable
OE4523F	HY	-	hypothetical protein	Questionable
OE1102R	ISH	-	probable transposase (ISH8/ISH26)	-
OE1102R	ISH	-	probable transposase (ISH8/ISH26)	-
OE1173F	ISH	-	probable transposase (ISH8/ISH26)	-
OE1506R	ISH	-	probable transposase (ISH8/ISH26)	-
OE7198F	ISH	-	probable transposase (ISH8/ISH26)	-
OE1742R	MIS	grpE	dnaJ/dnaK ATPase stimulator grpE	Insecure
OE3168R	MIS	-	pyridoxal phosphate-dependent aminotransferase	Trusted
OE1797R	REG	sirR	transcription regulator sirR	Questionable
OE2631F	TC	rpoD	DNA-directed RNA polymerase (EC 2.7.7.6) chain D	Trusted
OE2678R	TC	rpoF	DNA-directed RNA polymerase (EC 2.7.7.6) chain F	Trusted
OE4740R	TC	rpoA1	DNA-directed RNA polymerase (EC 2.7.7.6) chain A' (chain A)	Trusted
OE2601R	TL	rpl10	ribosomal protein L10	Trusted
OE2602R	TL	rpl1	ribosomal protein L1	Trusted
OE2633F	TL	rpl13	ribosomal protein L13	Trusted
OE2664F	TL	rps28R	ribosomal protein S28.eR	Trusted
OE3392F	TL	rpl2	ribosomal protein L2	Trusted
OE3394F	TL	rpl22	ribosomal protein L22	Trusted
OE3395F	TL	rps3	ribosomal protein S3	Trusted
OE3417F	TL	rpl15	ribosomal protein L15	Trusted
OE4532F	TL	rps6R	ribosomal protein S6.eR	Trusted
OE4720R	TL	rps10	ribosomal protein S10	Trusted
OE4721R	TL	aef1a	translation elongation factor aEF-1 alpha chain	Trusted
OE4735R	TL	rps7	ribosomal protein S7	Trusted
OE1675R	TP	pstB2	probable ABC-type phosphate transport system ATP-binding protein	Trusted

Tab. 12: Die im Cluster 5 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Einen erheblich unterschiedlichen Expressionsverlauf im Vergleich mit den anderen Clustern zeigen die Kandidaten des Clusters 6. In diesem Cluster sind die Gene zusammengefasst, die einen kontinuierlichen Anstieg ihrer Genexpression erkennen lassen. Welche Gene in diesem Cluster zu finden sind ist in der nachfolgenden Tabelle angegeben.

OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE3106F	EM	bop	bacteriorhodopsin precursor	Trusted
OE3107F	HY	-	hypothetical protein	-
OE2540R	ISH	-	probable transposase (ISH6/ISHS1)	-
OE4727R	ISH	-	probable IS1341-type transposase (ISHnew83)	-
OE3100F	MIS	blp	bacterioopsin-linked protein blp	Trusted

Tab. 13: Die im Cluster 6 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Unter den Kandidaten sind mit dem *bop* (Bakterioodopsin Vorläufer), dem *blp* (Bakterioopsin-verbundenes Protein) und dem ORF OE3107F drei Mitglieder des so genannten *bop*-Clusters des halobakteriellen Genoms von *H. sal.* R1 enthalten. Aufgrund von Ergebnissen früherer Experimente weisen diese Gene den erwarteten Expressionsverlauf auf (Abb. 56).

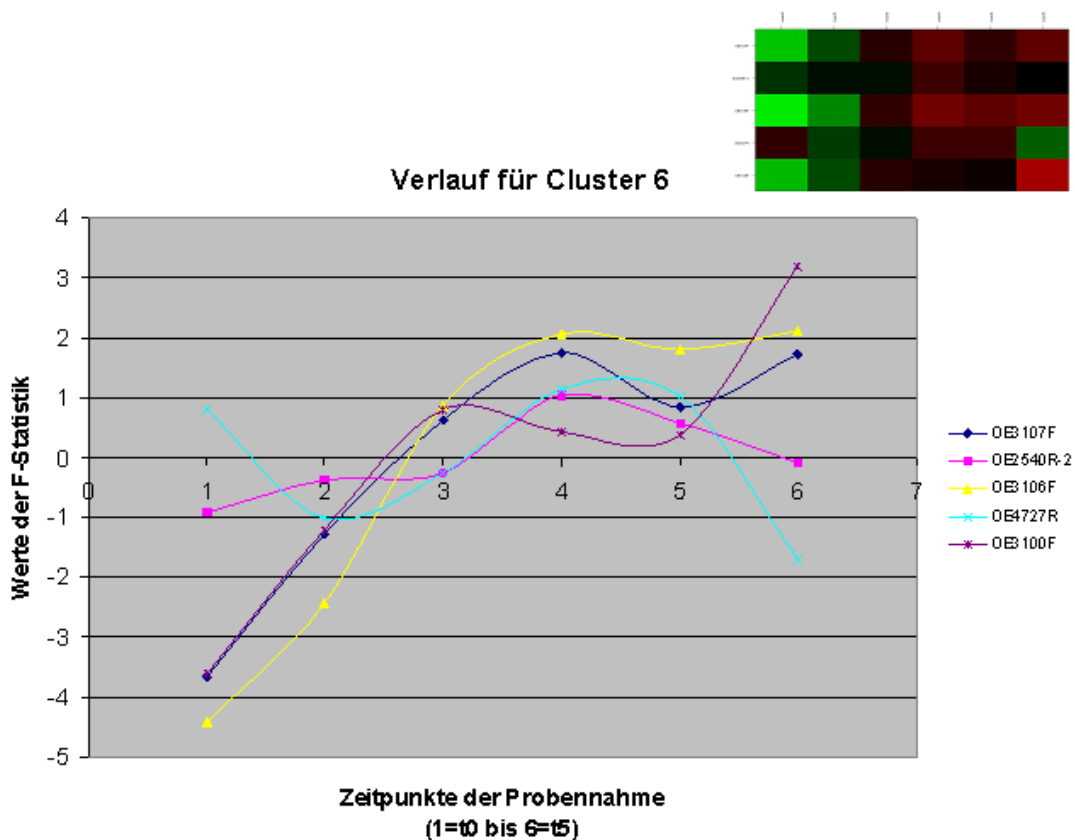


Abb. 56: Gezeigt wird der Expressionsverlauf der Gene des Clusters sechs anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0 = 0\text{min}$ ,  $t_1 = 45\text{min}$ ,  $t_2 = 90\text{min}$ ,  $t_3 = 180\text{min}$ ,  $t_4 = 360\text{min}$ ,  $t_5 = 1440\text{min}$ ). Weitere Informationen zu den Genen siehe Tabelle 13.

Ein wellenförmiger Verlauf kann für Cluster 7 beobachtet werden. In [Abb. 57](#) ist eine Expressionsabnahme bis zum Zeitpunkt  $t_1 = 15\text{min}$ , danach ein Anstieg der Expression durch den Wendepunkt zwischen  $t_2 = 90\text{min}$  und  $t_3 = 180\text{min}$  bis zum

Maximum bei  $t_4 = 360\text{min}$  mit abschließendem Abfall der Expression bis auf das Ausgangsniveau für den Zeitpunkt  $t_5 = 24\text{h}$  zu sehen.

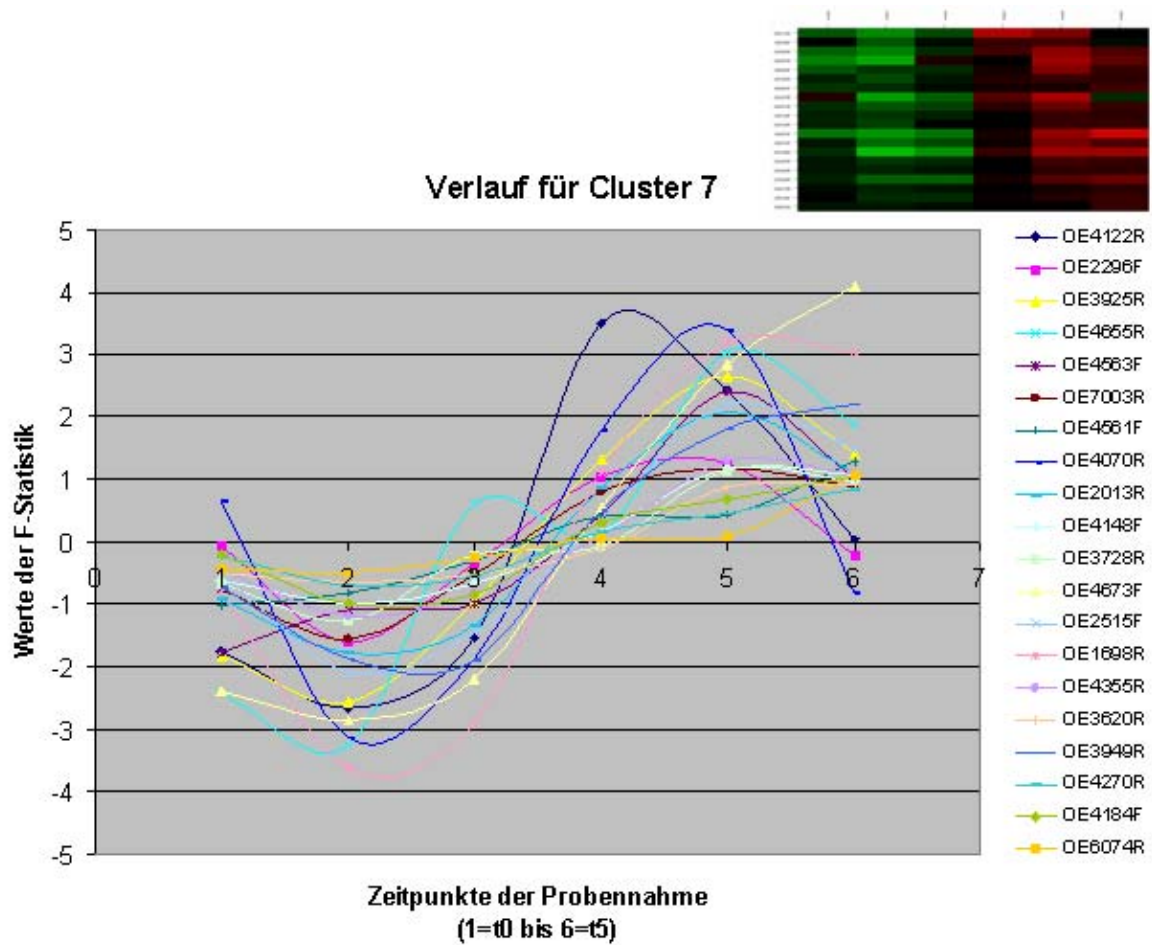


Abb. 57: Gezeigt wird der Expressionsverlauf der Gene des Clusters sieben anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0 = 0\text{min}$ ,  $t_1 = 45\text{min}$ ,  $t_2 = 90\text{min}$ ,  $t_3 = 180\text{min}$ ,  $t_4 = 360\text{min}$ ,  $t_5 = 1440\text{min}$ ). Weitere Informationen zu den Genen siehe folgende Tabelle.

Welche Gene in Cluster 7 zusammengefasst sind ist in der Tabelle 14 zu finden.

OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE4270R	-	-	[del] spurious ORF	-
OE3949R	CHY	-	glutaredoxin homolog	Questionable
OE4561F	CHY	-	conserved hypothetical protein	Questionable
OE7003R	CHY	-	conserved hypothetical protein	-
OE2296F	CP	psmB	proteasome beta chain	Questionable
OE3925R	CP	cctB	thermosome beta chain	Trusted
OE4122R	CP	cctA	thermosome alpha chain	Trusted
OE4070R	EM	cox1B	cytochrome-c oxidase (EC 1.9.3.1) chain I	Trusted
OE3728R	HY	-	hypothetical protein	-
OE4148F	HY	-	hypothetical protein	-
OE4563F	HY	-	hypothetical protein	Trusted
OE6074R	HY	-	hypothetical protein	-
OE2013R	LIP	acd4	probable acyl/butyryl-CoA dehydrogenase (EC 1.3.99.-)	Trusted
OE1698R	MIS	oxr4	probable oxidoreductase (EC 1.1.1.-) (aldehyde reductase homolog)	Trusted
OE3620R	MIS	hit2	histidine triad family protein	Insecure
OE4673F	MIS	cxp	carboxypeptidase (EC 3.4.-.-)	Trusted
OE2515F	NOF	-	conserved protein	Trusted
OE4184F	NOF	orc6	cell division control protein cdc6 homolog	Trusted
OE4355R	NOF	tssA	thiosulfate sulfurtransferase homolog	Trusted
OE4655R	NOF	-	conserved protein	Trusted

Tab. 14: Die im Cluster 7 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Besonders erwähnenswert unter den Mitgliedern des Clusters sieben sind die beiden Untereinheiten des Thermosoms (*cctA* und *cctB*) und die beta-Kette des Proteasoms (*psmB*) von *H. sal.* R1. So zeigen Untereinheiten zweier für die Proteinintegrität sehr wichtige Komplexe das gleiche Expressionsverhalten. Dies ist ein Indiz für eine verstärkte Proteinproduktion und damit für eine Umstrukturierung des vorhandenen Proteininventars.

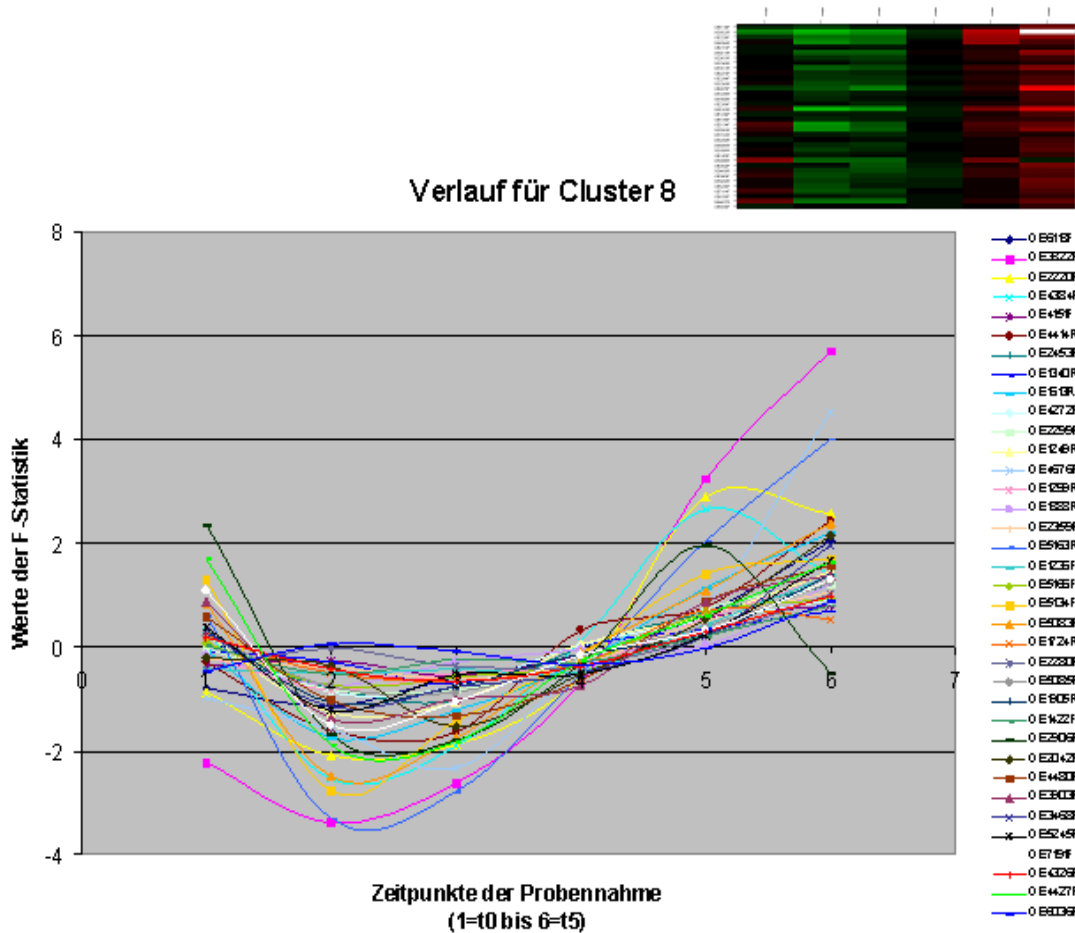


Abb. 58: Gezeigt wird der Expressionsverlauf der Gene des Clusters acht anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0=0\text{min}$ ,  $t_1=45\text{min}$ ,  $t_2=90\text{min}$ ,  $t_3=180\text{min}$ ,  $t_4=360\text{min}$ ,  $t_5=1440\text{min}$ ). Weitere Informationen zu den Genen siehe Tabelle 13.

Für Cluster 7 und Cluster 8 ist in den entsprechenden Abbildungen der Expressionsverläufe ein sehr ähnlicher Verlauf zu erkennen, mit dem Unterschied dass in Cluster acht ab dem Durchlaufen des Minimums beim Zeitpunkt  $t_1=15\text{min}$  ein kontinuierlicher Anstieg der Expression bis zum Zeitpunkt  $t_5=24\text{h}$  für die Mehrzahl der Kandidatengene zu beobachten ist.

Sieht man sich die Liste der Kandidatengene von Cluster acht in Tab. 15 an und addiert die Mitglieder der Funktionsklassen CHY, HY und NOF, stellt man fest dass hier für 24 Gene, die einen Anteil von 66,6% der Kandidatengene des Clusters ausmachen, zur Zeit keine näheren Informationen über ihre Funktionen zur Verfügung stehen.

OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE7191F	-	-	[del] spurious ORF	-
OE1236F	CHY	-	conserved hypothetical protein	-
OE1249R	CHY	-	conserved hypothetical protein	-
OE1724R	CHY	-	conserved hypothetical protein	-
OE1888R	CHY	-	conserved hypothetical protein	Trusted
OE1905R	CHY	-	probable hexosyltransferase (lipopolysaccharide core biosynthesis mannosyltransferase homolog)	-
OE2220F	CHY	-	conserved hypothetical protein	-
OE2299F	CHY	pho2	p-nitrophenylphosphatase homolog	-
OE2453R	CHY	-	conserved hypothetical protein	Trusted
OE3822R	CHY	-	conserved hypothetical protein	-
OE3903F	CHY	-	conserved hypothetical protein	-
OE4326F	CHY	-	conserved hypothetical protein	-
OE4414R	CHY	-	salinity-regulated protein homolog	-
OE1340R	HY	-	hypothetical protein	-
OE4151F	HY	-	hypothetical protein	-
OE4676F	HY	-	hypothetical protein	-
OE5085R	HY	-	hypothetical protein	-
OE5163R	HY	-	[del] spurious ORF	-
OE6118F	HY	-	hypothetical protein	-
OE3468R	LIP	crit2	phytoene dehydrogenase (EC 1.14.99.-) (phytoene desaturase) 2 (alternative EC 1.3.-.)	Trusted
OE1613R	MIS	-	probable acylaminoacyl-peptidase (EC 3.4.19.1)	Trusted
OE2906R	MIS	sod2	superoxide dismutase (EC 1.15.1.1) 2	Insecure
OE4272F	MIS	-	conserved hem operon protein	Trusted
OE4427R	MIS	dpsA	DNA-binding protein dpsA	Trusted
OE1422F	NOF	-	conserved protein	Trusted
OE2280R	NOF	-	protein OE2280R	Insecure
OE2359F	NOF	-	conserved protein	Trusted
OE4384R	NOF	-	conserved protein	Trusted
OE5083R	NOF	hsp5	heat shock protein homolog	Believed
OE5134F	NOF	-	conserved protein	Trusted
OE6036F	RRR	polB2	probable DNA-directed DNA polymerase (EC 2.7.7.7) type II	-
OE1299R	TP	hop	halorhodopsin	-
OE2042F	TP	yvgX	probable copper-transporting ATPase (EC 3.6.1.-)	Trusted
OE4480R	TP	pstA1	probable ABC-type phosphate transport system permease protein	Insecure
OE5166F	TP	ugpB	probable ABC-type transport system periplasmic substrate-binding protein	Trusted
OE5245F	TP	-	ABC-type transport system ATP-binding protein	-

Tab. 15: Die im Cluster 8 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Mit fünf Genen der Funktionsklasse TP ist hier der größte Anteil dieser Funktionsklasse in allen Clustern zu finden. Außer dem Chloridtransporter Halorhodopsin sind diese Gene Mitglieder von ABC-Transportsystemen.

Wie Cluster eins zeigt auch Cluster neun einen Abfall der Expression bis zum Minimum zwischen  $t_2 = 90\text{min}$  und  $t_3 = 180\text{min}$  und danach einen Anstieg auf das Ausgangsexpressionsniveaus (siehe Abb. 58).



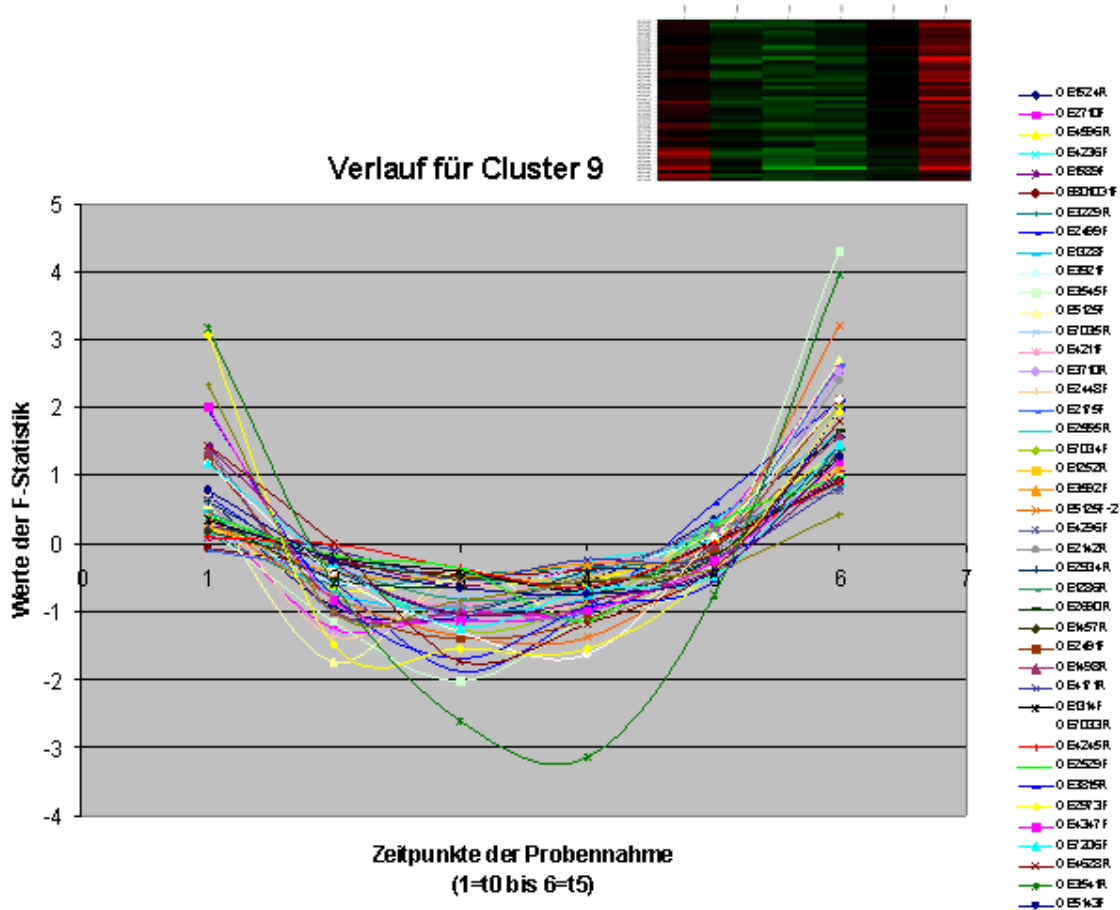


Abb. 59: Gezeigt wird der Expressionsverlauf der Gene des Clusters 9 anhand eines Graphen und des entsprechenden Ausschnitts aus dem Matrixplot aus Abb. 50. Die Skalierung des Zeitintervalls ist nicht linear ( $t_0 = 0\text{min}$ ,  $t_1 = 45\text{min}$ ,  $t_2 = 90\text{min}$ ,  $t_3 = 180\text{min}$ ,  $t_4 = 360\text{min}$ ,  $t_5 = 1440\text{min}$ ). Weitere Informationen zu den Genen siehe folgende Tabelle.

Wie bei Cluster 8 macht auch bei Cluster 9 die Summe der drei Funktionsklassen CHY, HY und NOF einen Anteil von 66% ihrer Kandidatengene aus. Aufgrund des sehr ähnlichen Expressionsverlaufs könnte man für diese Gene eine ähnliche Funktion vermuten, denn wie man am Beispiel von Cluster 4 und 5 gesehen hat, sind in diesen beiden Clustern mit einem ähnlichen Expressionsverlauf alle ribosomalen Proteine zu finden, die für das zweite Zeitreihenexperiment als differentiell exprimierte Gene anzusehen sind.


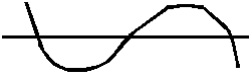



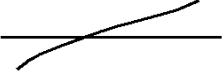
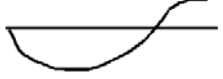


OE-Nummer	Funktionsklasse	Genname	Proteinname	Protein-Identifizierung
OE7035R	-	-	[del] spurious ORF	-
OE7193R	-	-	[del] spurious ORF	-
OE1252R	CHY	-	thioredoxin homolog	-
OE1314F	CHY	-	transport protein homolog	-
OE1457R	CHY	-	conserved hypothetical protein	-
OE1524R	CHY	-	conserved hypothetical protein	-
OE2142R	CHY	-	conserved hypothetical protein	-
OE2448F	CHY	boa4	transcription regulator bat homolog	-
OE2491F	CHY	-	conserved hypothetical protein	-
OE2529F	CHY	-	polysaccharide biosynthesis protein homolog	Insecure
OE3815R	CHY	-	conserved hypothetical protein	-
OE3921F	CHY	-	conserved hypothetical protein	-
OE4236F	CHY	-	phosphate transport protein homolog	-
OE4596R	CHY	-	conserved hypothetical protein	-
OE2555R	COM	menE	probable acid-CoA ligase (EC 6.2.1.-) menE	Trusted
OE3229R	COM	hmcA	probable chelatase (EC 4.99.1.-) (cobalamin cluster)	-
OE3541R	CP	hsp1	probable heat shock protein	Trusted
OE5125F	CP	gvpA2	gas-vesicle protein gvpA2	Insecure
OE7034F	CP	gvpA1	gas-vesicle protein gvpA1	-
OE1328F	HY	-	[del] spurious ORF	-
OE2499F	HY	-	hypothetical protein	-
OE2690R	HY	-	hypothetical protein	Questionable
OE3545F	HY	-	hypothetical protein	Questionable
OE4245R	HY	-	hypothetical protein	-
OE4296F	HY	-	hypothetical protein	-
OE4347F	HY	-	hypothetical protein	Questionable
OE7133F	HY	-	hypothetical protein	-
OE7206F	HY	-	hypothetical protein	-
OE2710F	LIP	acd3	probable acyl/butyryl-CoA dehydrogenase (EC 1.3.99.-)	Trusted
OE4211F	MIS	cat	probable succinal-CoA transferase (EC 2.8.3.-)	Questionable
OE7033R	MIS	gvpD1	gas-vesicle operon protein gvpD1 (probable repressor protein)	-
OE8010G1F	MIS	repI2	replication protein repI2	-
OE1498R	NOF	gadD	glutamate decarboxylase homolog	Trusted
OE2175F	NOF	-	conserved protein	Trusted
OE2934R	NOF	-	conserved protein	Trusted
OE2973F	NOF	-	conserved protein	Trusted
OE3592F	NOF	-	conserved protein	Trusted
OE3710R	NOF	trkA6	trkA domain protein	Trusted
OE4171R	NOF	-	conserved protein	Trusted
OE4628R	NOF	-	conserved protein	Insecure
OE1286R	RRR	mutS	DNA mismatch repair protein	-
OE1589F	RRR	nthA1	DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) endonuclease III	Trusted
OE5143F	TP	-	probable transport protein	-

Tab. 16: Die im Cluster 9 zusammengefassten Gene mit Informationen aus der HaloLex-Datenbank.

Aus der Tab. 16 ist zu entnehmen, dass in Cluster 9 alle gefundenen Gasvesikel-Proteine zu finden sind.

Für die gefundenen Chaperone zeigt sich eine Verteilung auf mehrere Cluster. So wird mit dem *hsp1*, einem Hitzeschockprotein, ein Chaperon im Cluster 9 gefunden, während die Gene *cspD1* und *cspD2*, beides Kälteschockproteine, in Cluster 5 zu finden sind.

Eine kleine Zusammenfassung der Zusammensetzung der Cluster in Bezug auf die Verteilung der Gene auf die Funktionsklassen ist in der nachfolgenden Tabelle gegeben. Als Zusatzinformation enthalten ist der Verlauf der Genexpression für die einzelnen Cluster schematisch dargestellt.

Nummer Cluster	Anzahl Gene pro Cluster	Verteilung Gene auf Funktionsklassen	Verlauf der Expression
Cluster 1	2	1 CHY, 1 NOF	
Cluster 2	11	3 CHY, 1 HY, 1 MIS, 2 NOF, 1 SIG, 3 TL	
Cluster 3	19	2 CHY, 2 EM, 5 ISH, 2 MIS, 6 NOF, 1 SIG, 1 Del	
Cluster 4	69	1 AA, 6 CHY, 2 CIM, 6 EM, 2 HY, 4 ISH, 1 LIP, 5 MIS, 3 NOF, 1 NUM, 4 TC, 32 TL, 2 TP	
Cluster 5	36	3 CHY, 1 CIM, 2 CP, 2 EM, 3 HY, 5 ISH, 2 MIS, 1 REG, 3 TC, 12 TL, 1 TP	
Cluster 6	5	1 EM, 1 HY, 2 ISH, 1 MIS	
Cluster 7	20	3 CHY, 3 CP, 1 EM, 4 HY, 1 LIP, 3 MIS, 4 NOF, 1 Del	
Cluster 8	36	12 CHY, 6 HY, 1 LIP, 4 MIS, 6 NOF, 1 RRR, 5 TP, 1 Del	
Cluster 9	44	12 CHY, 2 COM, 4 CP, 9 HY, 1 LIP, 3 MIS, 8 NOF, 2 RRR, 1 TP, 2 Del	

Tab. 17: Angaben über die Anzahl und Verteilung der differentiell exprimierten Gene des zweiten Zeitreihenexperiments auf die Cluster nach der Hauptkomponentenanalyse. Die Anzahl der Gene pro Funktionsklasse wird durch die Zahl vor der Abkürzung angegeben. Schematisch dargestellt ist der Expressionsverlauf der einzelnen Cluster.

Anhand der Tabelle 17 ist gut zu erkennen, dass eine Hauptkomponentenanalyse eine sehr hilfreiche Methode darstellt, um Tendenzen und Verläufe im Zusammenhang eines Zeitreihenexperimentes abzubilden.

Durch die Anwendung dieser exploratorischen Analysenmethode auf den Datensatz des zweiten Zeitreihenexperimentes von *H. sal* R1 konnte bewiesen

werden, dass eine PCA geeignet ist Gene mit einem ähnlichen Expressionsverlauf zu identifizieren und zu gruppieren. Diese Tatsache wird z.B. dadurch untermauert, dass 70% aller ribosomalen Proteine ein identisches Expressionsverhalten aufweisen und in den Clustern 4 und 5 zusammengefasst wurden.

Ein zusätzlicher Aspekt bei der Identifizierung von Genen mit einem gleichen Expressionsmuster ist, dass man diesen Kandidatengenen eine gleiche oder zumindest sehr ähnliche Funktion zu ordnen kann. Durch diesen Punkt wird eine neue Perspektive in der Interpretation der Transkriptomergebnisse eröffnet, denn so könnte auf der Basis dieser PCA-Analyse für die jetzigen Gene der Funktionsklassen CHY, HY und NOF versucht werden eine grobe Zuordnung ihrer Rolle durchzuführen. Diese Hypothesen könnten dann in weiteren Versuchen bestätigt werden. Somit stellt eine PCA ein zusätzliches Instrument dar, wodurch dem Experimentator gute Hinweise gegeben werden, Experimente zu planen.

Schaut man sich die Maxima in den Expressionsverläufen der einzelnen Cluster einmal an, wird deutlich, dass diese in der Mehrzahl zwischen  $t_2 = 90\text{min}$  und  $t_4 = 360\text{min}$  liegen. Man kann somit eine relativ späte Reaktion auf den Stimulus phototrophes Wachstum feststellen.

Eine ebenso nahe liegende Interpretation auf der Basis der PCA und der Tatsache der moderat ausfallenden Induktion der Gene ist, dass der Anpassungsprozess an das phototrophe Wachstum bei *H. sal.* R1 nicht nur auf der Gen- sondern auch auf der Proteinebene abläuft.

Da Proteomexperimente zum phototrophen Wachstum ebenfalls keine starke Veränderungen der Proteinmuster zeigt, liegt die Vermutung nahe, dass es einen sehr gut regulierten Mechanismus gibt auf der Basis von Proteinexpression, Stabilität und Abbau.

Im nächsten Kapitel werden zwei konkrete Ansätze zur Interpretation der Transkriptomanalyse vorgestellt.

## 5.4 Interpretation der Daten und weitere Experimente

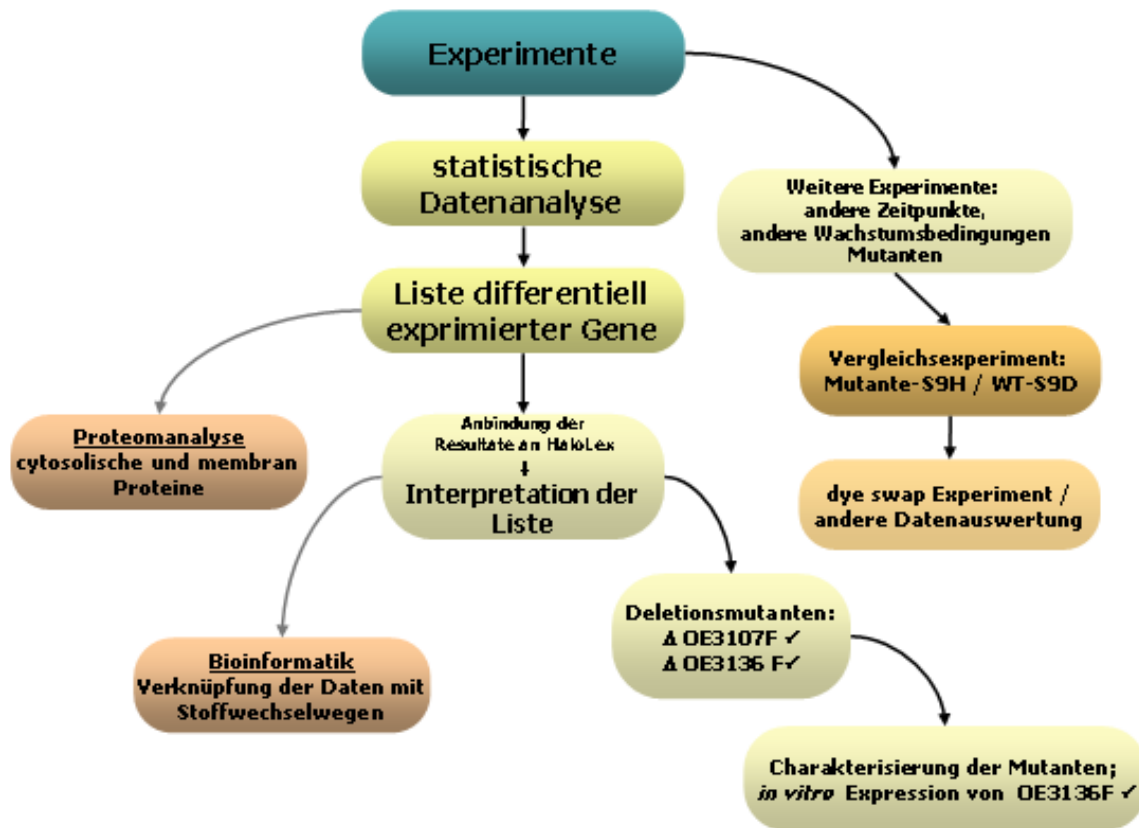


Abb. 60: Flussdiagramm der durchgeführten Experimente zur Interpretation der Ergebnisse der Transkriptom-Experimente.

In dem Flussdiagramm aus Abb. 60 ist zu sehen, dass ein Transkriptom-Experiment und die erhaltenen Daten in verschiedenen Ebenen bewertet und interpretiert werden können. Abb. 60 stellt bildhaft die Bedeutung eines Transkriptom-Experimentes vor. Mit der Erstellung einer Liste von differentiell exprimierten Genen mittels DNA-Mikroarray-Analyse ist zwar die Expressionsanalyse formal abgeschlossen, doch die erstellte Liste ist nur Ausgangspunkt für weiterführende Experimente auch anderer Disziplinen wie z.B. der Bioinformatik oder Proteomik.

In diesem Kapitel sollen einige weiterführende Experimente näher beschrieben werden.

#### 5.4.2 Deletion einzelner Gene

Die Experimente in diesem Kapitel basieren auf den Ergebnissen des Transkriptomexperiments des Zweigruppenvergleichs aerob / phototroph.

Wie in Abschnitt 5.2 bereits erwähnt, gehören 54,5% der regulierten Gene des Transkriptomexperiments der Funktionsklassen CHY, HY und NOF an. Es existieren mehrere Ansätze zu versuchen, diesen Kandidatengenen eine Funktion zu zuweisen.

Eine Möglichkeit der Interpretation und Validierung dieser regulierten Gene besteht darin, Deletionsmutanten von Kandidatengenen dieser Liste herzustellen. Mit Hilfe solcher Mutanten besteht die Chance, die Relevanz bzw. Funktion im Regulationsnetzwerk zu bestätigen und zu untersuchen.

Dieser Lösungsansatz zur Aufklärung der Bedeutung von bisher unbekanntem Genen wurde auch für dieses Projekt gewählt.

Die Auswahl der zu deletierenden Gene erfolgte zum einen auf Stärke der Regulation unter den gewählten Versuchsbedingungen und zum anderen auf der Nähe zu dem *bop*-Gen Cluster (siehe Abb. 45). Diese Strategie wurde für mehrere Gene verfolgt. Für zwei der möglichen Kandidatengene konnte eine Deletionsmutante hergestellt werden. Dabei handelt es sich um den ORF OE3107F und OE3136F. Der ORF 3107F befand sich wegen seiner direkten Nachbarschaft zum *bop* Gen auf der Liste und der ORF OE3136F zum einen wegen seiner relativen Nähe zum *bop*-Gen Cluster und zum anderen wegen einer möglichen DNA bindenden Eigenschaft seines Genproduktes. Im folgendem wird die Methode zur Herstellung der Mutanten für die ORFs OE3107F und OE3136F beschrieben.

Von den beiden Kandidatengenen OE3107F und OE3136F wurden die proteinkodierenden Sequenzen vollständig aus dem Genom von *H. sal.* R1 entfernt.

Dazu wurden Zellen von *H. sal.* R1 mit einem Plasmid das einen Selektionsmarker und die modifizierte Sequenz enthält transformiert, wobei die DNA in das Genom von *H. sal.* R1 integriert. Die Plasmide wurden so konstruiert, dass sie als Deletionskassette 300-500 Basenpaare flankierende Sequenzen stromauf- und stromabwärts des jeweils zu deletierenden Gens enthielten. Über einen zweiten homologen Rekombinationsschritt werden die Zielgene aus dem Genom entfernt und stattdessen die Deletionskassetten eingebaut.

Die Fragmente der Deletionskassetten wurden mittels PCR in zwei Schritten amplifiziert. Während des ersten Schritts werden die flankierenden Bereiche stromauf- und stromabwärts des jeweiligen Gens in zwei verschiedenen PCR-Reaktionen erzeugt. Dabei wurden die äußeren Primer der Fragmente so gewählt, dass sich die Schnittstellen für die Restriktionsenzyme zur Klonierung in das Plasmid einmal am 5'-Ende des stromaufwärts Fragmentes und am 3'-Ende des stromabwärts Fragmentes befinden. Die inneren Primer besitzen neben einer sequenzspezifischen Region, jeweils einen Überhang von ca. 14 Basenpaaren, der komplementär zum anderen Fragment ist. Für den zweiten Schritt der PCR wurden die Fragmente mit den komplementären Enden aufgereinigt und zusammen mit den äußeren Primern in der Fusions-PCR eingesetzt, wodurch das stromauf- mit dem stromabwärts Fragment verknüpft wurde und die Deletionskassette entstanden ist. In der folgenden Abbildung wird der Ablauf der beschriebenen Schritte schematisch dargestellt.

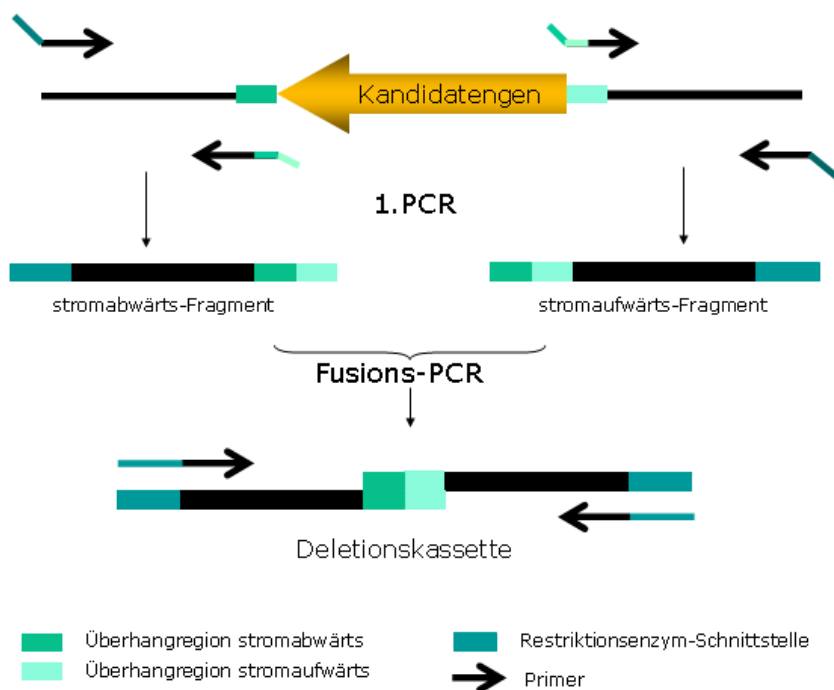


Abb. 61: Schematische Darstellung des Ablaufs einer zwei Stufen PCR zur Herstellung einer Deletionskassette.

Die entstandenen Plasmide mit den Deletionskassetten wurden zur Kontrolle sequenziert. Die bis hierher beschriebene Methode zur Erzeugung von Deletionsstämmen, wird im folgenden anhand der Ergebnisse der zwei Gene OE3107F und OE3136F illustriert.

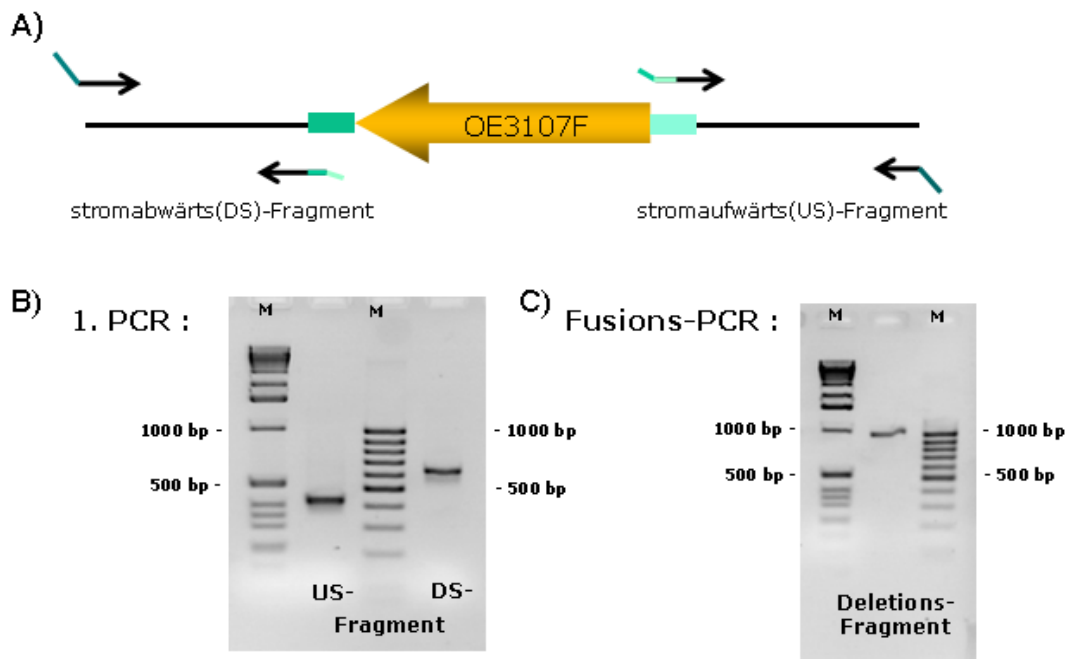


Abb. 62: Darstellung der Konstruktion der Deletionsmutante R1 $\Delta$ 3107F. A) Schema zur Herstellung der Deletionskassette. B) 1%iges Agarosegel mit den Produkten der 1. PCR zur Amplifizierung der flankierenden Sequenzen von OE3107F. Erwartete Bandengröße des US-Fragments ist 400 bp und die des DS-Fragments ist 600 bp. C) Agarosegel des Produkts der Fusions-PCR mit der erwarteten Größe von 1000 bp, welches zur Klonierung mit den Restriktionsenzymen BamHI/XbaI behandelt wurde. Als Marker (M) wurden ein 1 kb Marker (links) und ein 1000 bp Marker (rechts) eingesetzt. Abkürzungen: US = „up-stream“/stromaufwärts; DS = „down-stream“/stromabwärts; bp = „base pair“/Basenpaar; kb = „kilo base“/ Kilobasen.

In Abbildung 62 ist die Amplifizierung der Deletionskassette für die Deletionsmutante R1 $\Delta$ 3107 gezeigt. Analog zu diesem Ablauf wurde auch die Deletionsmutante R1 $\Delta$ 3136 hergestellt.

Zur Klonierung wurden die Deletions-Fragmente mit den Restriktionsenzymen BamHI und XbaI behandelt, aufgereinigt und in das ebenfalls mit BamHI/XbaI behandelte Ausgangsplasmid pM@ $\beta$  ligiert. Dieses Ausgangsplasmid hat den entscheidenden Vorteil, dass es neben dem Selektionsmarker (Mevinolinresistenzgen) noch ein  $\beta$ -Galactosidasegen enthält und somit nach der Transformation in *H. sal.* R1 Zellen und dem zweiten Rekombinationsschritt, eine Blau-Rot-Selektion zur Detektion der positiven Transformanten durchgeführt werden kann.



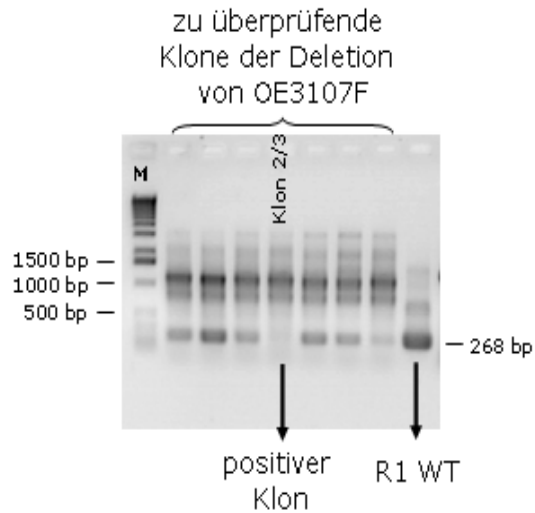


Abb. 63: Agarosegel mit den Produkten der Klon-PCR zur Detektion der Deletion des Gens OE3107F. Genspezifische Primer liefern ein Produkt mit einer Bandengröße von 268 bp. Abkürzungen: WT = Wildtyp; M = 1k DNA-Marker.

Eine schnelle Methode zur Detektion von positiven Deletionsmutanten, ist die Durchführung einer PCR z.B. mit genspezifischen Primern. Hierbei sollte mit DNA des WT-Stammes ein PCR-Produkt in der erwarteten Größe entstehen, während die Deletionsmutanten kein Produkt aufweisen. Wie am Beispiel der Deletion von OE3107F in Abbildung 63 gezeigt, erhält man mit genspezifischen Primern eine Bande der erwarteten Größe von 268 bp (Positivkontrolle), während man für den gekennzeichneten Klon kein Produkt dieser Größe beobachtet.

Der Test auf eine Deletion mittels PCR ist ein schnelles Verfahren, jedoch ist die Durchführung eines Southern-Blots zur eindeutigen Bestätigung einer Deletion, einer PCR vorzuziehen. Aus diesem Grund wurde für beide Deletionsstämme eine Southern-Blot-Analyse durchgeführt.

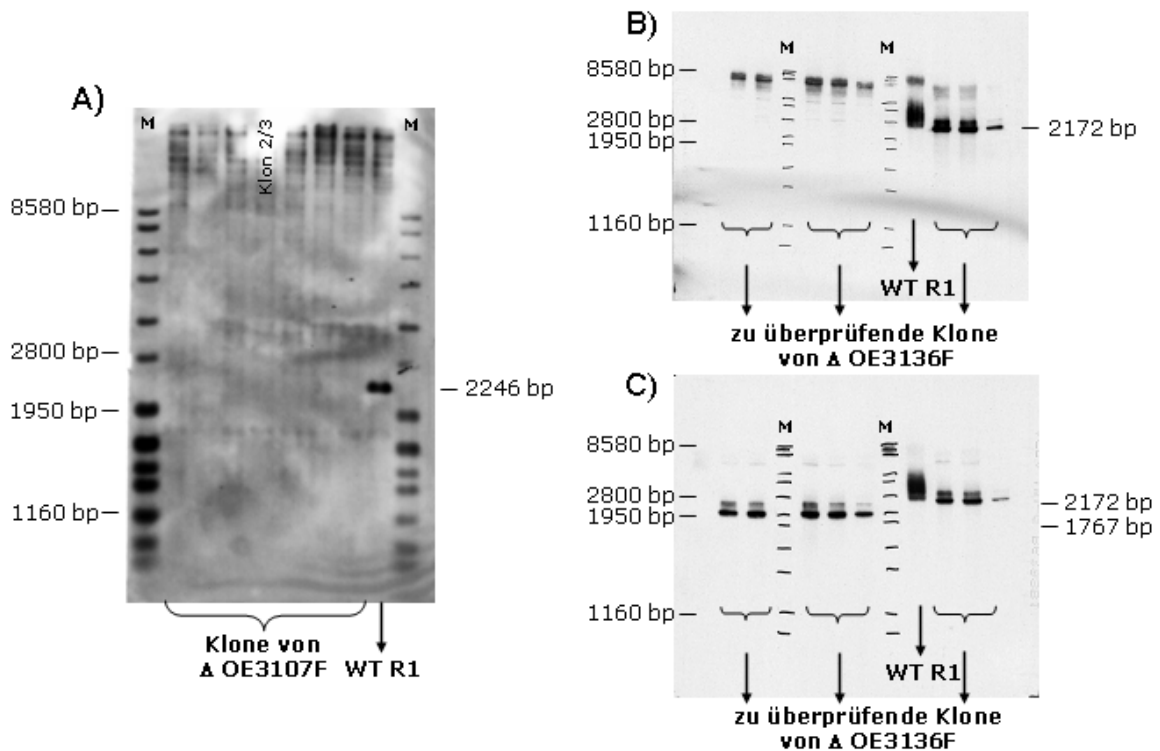


Abb. 64: Ergebnisse der Southern Blot-Analyse für die Stämme R1 $\Delta$ 3107 und R1 $\Delta$ 3136. A) Detektion von verschiedenen Klonen auf Deletion des Gens OE3107F mittels einer Digoxigenin(DIG)-markierten genspezifischen Sonde. Erwartete Größe der Bande nach einem KpnI-Verdau genomischer DNA ist 2246 bp. B) Überprüfung der Deletion des Gens OE3136F in verschiedenen Klonen mit Hilfe einer DIG-markierten genspezifischen Sonde. Nur der WT-Stamm und negative Klone zeigen nach BglII-Verdau eine Bande auf der Höhe von 2172 bp. C) Für die gleichen Klone wie unter B, Kontrolle nicht mit einer genspezifischen sondern mit einer flankierenden Sonde. Nach der Spaltung mit BglII weisen positive Klone eine Bande der Größe von 1767 bp auf.

Die Resultate der Southern-Blot-Analysen sind in (Abb. 64) gezeigt. Eindeutig ist in (Abb. 64 A) zu sehen, dass mit einer Digoxigenin(DIG)-markierten spezifischen Sonde für OE3107F keine Banden bei den untersuchten Klonen zu detektieren sind, während für den R1-Wildtyp-Stamm nach der Spaltung der genomischen DNA mit KpnI eine Bande der erwarteten Größe von 2246 bp mit einer OE3107F spezifischen Sonde zu beobachten ist. Somit zeigt der Genotyp aller untersuchten Klone der Deletionsmutante das Fehlen des Gens OE3107F. Da für den Klon 2/3 sowohl die Southern-Blot-Analyse als auch das Ergebnis der PCR eindeutig die Deletion des Gens bestätigen, wurde dieser Klon für alle weiteren Experimente eingesetzt.

Für die Deletionsmutante R1 $\Delta$ 3136F werden in (Abb. 64 B) und C) die Ergebnisse der Southern-Blot-Analysen dargestellt. Zur Bestätigung der Mutation wurden hier zwei Sonden eingesetzt. Wie in (Abb. 64 B) zu sehen, sind bei Deletion des Gens OE3136F nach einer Spaltung der genomischen DNA mit BglII keine Banden

für die genspezifische Sonde auf Höhe von 2172 bp vorhanden. Diese Ergebnisse werden durch die Verwendung einer flankierenden Sonde bestätigt, denn wie in (Abb. 64 C) dargestellt, weisen die Klone mit der Deletion des Gens im Gegensatz zum Wildtyp eine um 405 bp kleinere Bande auf, was genau der Länge des Gens entspricht. Die Ergebnisse der Deletion von OE3136F in den untersuchten Klonen wird auch durch die hier nicht gezeigten Resultate der PCR-Analyse bestätigt. Die weiteren Experimente zur Charakterisierung dieser Deletionsmutante wurde alle mit dem Klon 2/5 durchgeführt.

In der folgenden Tabelle ist eine Zusammenstellung von Proteininformationen aus der HaloLex-Datenbank für die Gene OE3107F und OE3136F gegeben.

	<b>OE3107F</b>	<b>OE3136F</b>
Eintrag in PIR-Datenbank	A84301	C84303
Proteinname	hypothetical protein	conserved hypothetical protein
Funktionsklasse	HY	CHY
Genlokus	CHR1083080-1083349F	CHR1100273-1100677F
Anzahl der Basen	270 bp	405 bp
Anzahl der AA	89	134
Theoretisches Molekulargewicht	9170 Da	14233 Da
Lokalisierung	Membran	Cytosol

Tab. 18: Auszug von proteinspezifischen Information aus der HaloLex-Datenbank für die Gene OE3107F und OE3136F (mit PIR = Protein Information Resource -Datenbank, HY = hypothetische Proteine, CHY = konserviert hypothetische Proteine, bp = Basenpaar, Da = Dalton).

Von den verifizierten Deletionsmutanten R1 $\Delta$ 3107F und R1 $\Delta$ 3136F wurden zur Charakterisierung Wachstumskurven im Vergleich zu dem Mutterstamm *H. sal.* R1 aufgenommen. Dazu wurden die Stämme im Standardvollmedium unter aeroben und phototrophen Wachstumsbedingungen kultiviert. Die aufgezeichneten Wachstumskurven sind in Abb. 65 gezeigt.

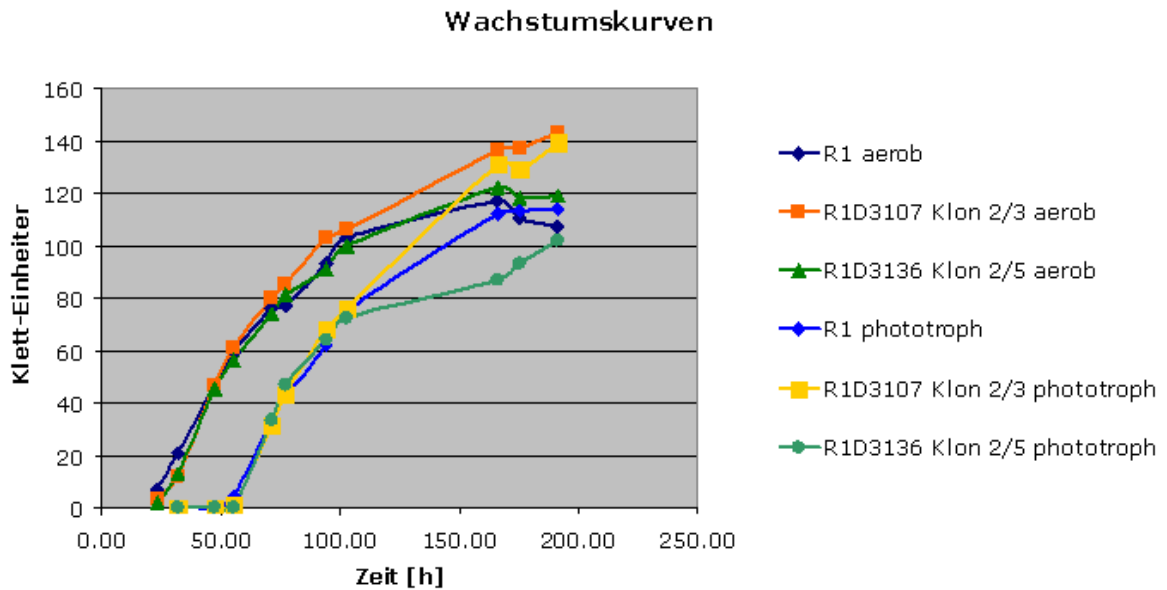


Abb. 65: Wachstumskurven der Deletionsstämme R1 $\Delta$ 3107F und R1 $\Delta$ 3136F im Vergleich zum Wildtyp-Stamm R1 unter aeroben und phototrophen Wachstumsbedingungen.

Als Fazit aus den Wachstumskurven der Mutanten kann geschlossen werden, dass die Deletionsstämme R1 $\Delta$ 3107F und R1 $\Delta$ 3136F bis zur stationären Wachstumsphase genauso wie der Wildtyp *H. sal.* R1 wachsen. Man kann für den Stamm R1 $\Delta$ 3107F ab ca. 120 h Kultivierung ein gesteigertes Wachstum im Vergleich zu den anderen Stämmen feststellen (aerob wie phototroph). Weiterführende Experimente müssten dies verifizieren.

Zum Vergleich der Färbungen der aerob und phototroph gewachsenen Deletionsstämme und des Wildtyps, sind in Abb. 66 Photographien der Kulturen gezeigt.

Anhand der Bilder kann man deutlich sehen, dass alle drei Stämme unter phototrophen Wachstumsbedingungen im Vergleich zu den aeroben Kulturen eine violetttere Färbung besitzen. Die aerob gewachsene Kultur der Deletion von OE3136F zeigt im Vergleich zum Wildtyp *H. sal.* R1 eine heller orange Färbung (Abb. 66 B). Allerdings ist mit dem Auge eine nähere Differenzierung der Färbungen der Stämme nicht durchzuführen, so dass man feststellen kann, dass ein offensichtlicher Phänotyp aufgrund der Deletionen nur sehr schwer zu erkennen ist.

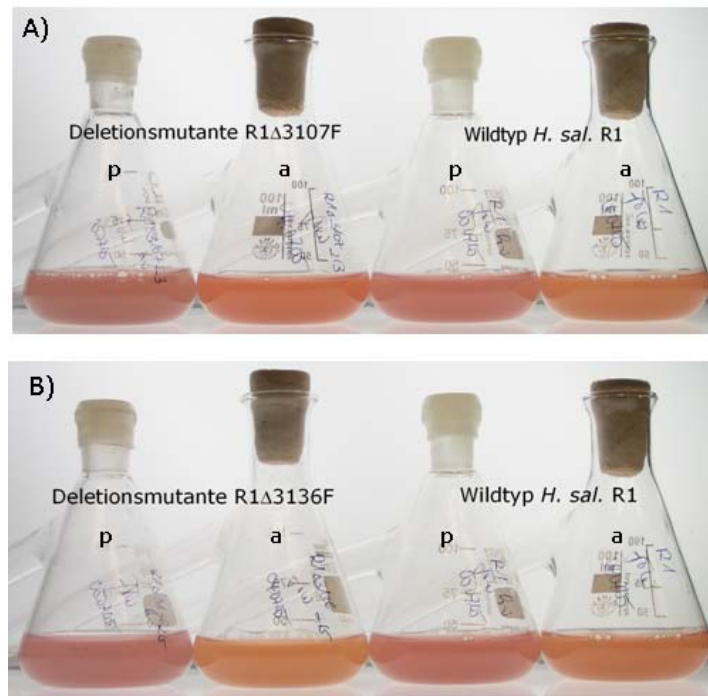


Abb. 66: Photographien von Kulturen der Deletionsstämmen R1Δ3107F (A) und R1Δ3136F (B) und des Wildtyps *H. sal.* R1. Es werden im Vergleich aerobe (Kolben mit a) und phototrophe (Kolben mit p) Kulturen gezeigt.

Zur Charakterisierung der Stämme wurden die Zelllysate von aerob gewachsenen Kulturen zur Analyse der Pigmentzusammensetzung photometrisch vermessen. Dazu wurde ein so genannter Schnelltest des BR-Gehalts durchgeführt, bei dem die jeweiligen Lysate halbiert und einer Hälfte CTAB (Cetyltrimethylammoniumbromid) zugesetzt wurde. Die Behandlung der Lysate mit diesem milden Detergenz war notwendig, um die Purpurchromophore zu entfärben, damit im Differenzspektrum die Absorptionsbanden des Bakteriorhodopsions bei ca. 568nm besser detektiert werden kann.

Das Ergebnis der Schnelltests für die Stämme *H. sal.* R1, R1Δ3107F, R1Δ3136F und S9H ist in (Abb. 67) dargestellt.

Aufgrund der starken Streuung ist in (Abb. 67 A) der BR-Peak nicht eindeutig zu sehen. Die auftretende Streuung kann durch eine mathematische Berechnung einer Anpassung von den Daten abgezogen werden. Dies erfolgte durch die Bestimmung einer exponentiellen Abnahme 1. Ordnung zwischen Messpunkten bei 350nm und 720nm, was in (Abb. 67 B) gezeigt wird.

Das Ergebnis der Berechnung und Subtraktion der Ausgleichsgeraden ist in (Abb. 67 C) zu sehen.

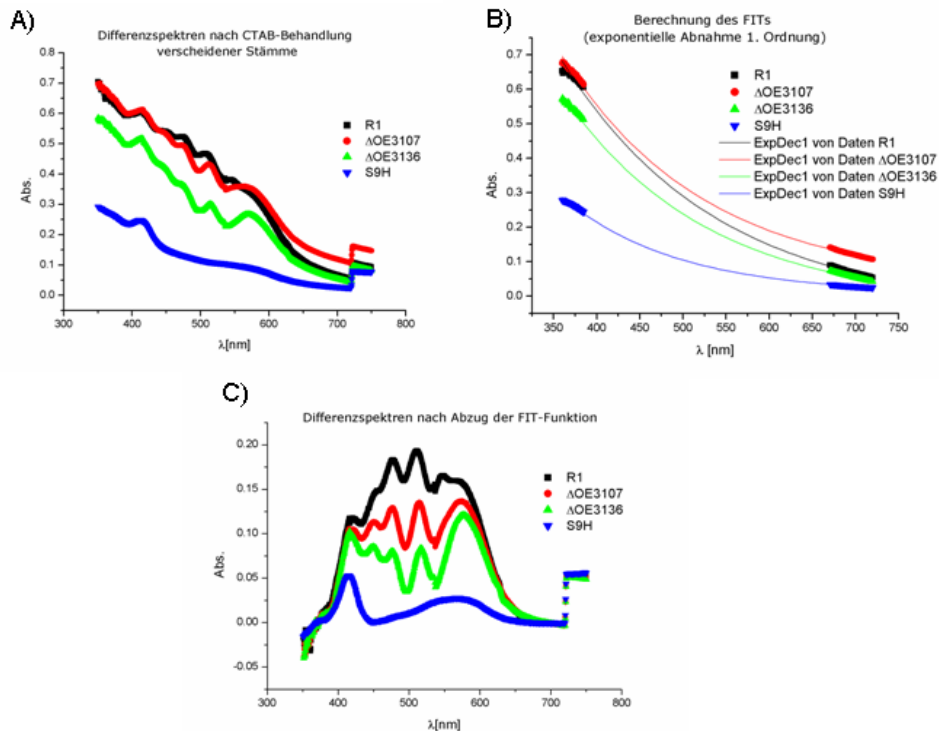
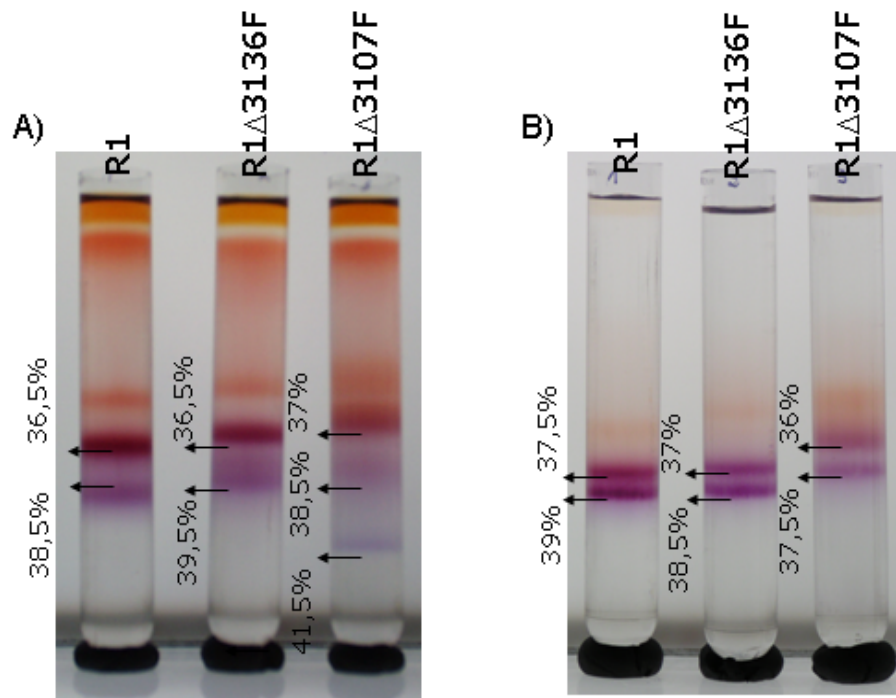


Abb. 67: Spektren der Stämme R1, R1 $\Delta$ 3107F und R1 $\Delta$ 3136F. A) Differenzspektren von Zellysaten der verschiedenen Stämme mit und ohne CTAB-Behandlung. B) Ergebnis der Berechnung von Ausgleichsgraden, zur Behebung der auftretenden Streuung. C) Spektren von Zellysaten nach Abzug der berechneten Ausgleichsgeraden.

Nach diesem BR-Schnelltest sind deutliche Unterschiede im Gehalt des Bakteriorhodopsins der aerob gewachsenen Kulturen auszumachen. So weist die farblose Mutante S9H den geringsten und der Wildtyp *H. sal.* R1 den größten Gehalt an BR auf. Für die beiden Deletionsstämme wurde im Vergleich zum Wildtyp-Stamm ein eindeutig geringerer Gehalt an BR nachgewiesen. Deshalb kann man aus diesen Ergebnissen schließen, dass die Deletionen der Gene OE3107F und OE3136F einen Einfluß auf die Synthese der Purpurmembran besitzen.

Die Durchführung eines BR-Schnelltest gibt zwar Aufschluss über den Gehalt von BR, jedoch über eine mögliche Änderung anderer Pigmente erhält man mit diesem Experiment keine Informationen. Zur vollständigen Charakterisierung des gesamten Pigmentgehalts der Zellysate bietet sich deshalb die Durchführung eines kontinuierlichen Saccharosegradienten an.

Die Ergebnisse dieser Analyse sind in (Abb. 68) dargestellt.



**Abb. 68:** Photographien von kontinuierlichen Saccharosegradienten (25 – 45%) auf denen das Zelllysate verschiedener Stämme direkt nach der Dialyse über Nacht (A) oder nach vorheriger Reinigung (B) auf die Gradienten aufgebracht wurden.

Auch anhand der Gradienten aus der (Abb. 68 A) kann man den deutlichen Unterschied im Bakteriorhodopsingehalt sehen. Die BR-Banden auf der Höhe von 39% Saccharose zeigen für die Deletionsmutante R1 $\Delta$ 3107F und R1 $\Delta$ 3136F einen geringeren Anteil von Bakteriorhodopsin. Auch an der Zusammensetzung der Ruberine können leichte Unterschiede detektiert werden, die im Rahmen der Arbeit nicht weiter untersucht wurden. Die Bande bei 41,5% für die Mutante R1 $\Delta$ 3107F ist darauf zurückzuführen, dass die Gradienten nicht gepuffert waren und deshalb ein Teil des Bakteriorhodopsins dort zu finden ist. Auch die offensichtlichen Doppelbanden des Bakteriorhodopsins sind auf Gradientenartefakten zurückzuführen. Der Unterschied zwischen (Abb. 68 A) und (B) besteht in einem vorgeschaltetem Reinigungsschritt, der in einer zusätzlichen Zentrifugation besteht, wodurch ein erheblicher Anteil der Ruberin haltigen Zellmembranfragmente abgetrennt wird.

#### 5.4.1.1 Diskussion

Als Fazit kann man feststellen, dass die Deletionsmutanten R1 $\Delta$ 3107F und R1 $\Delta$ 3136F im Vergleich zum Wildtyp-Stamm *H. sal.* R1 deutliche Unterschiede in

ihrer Pigmentzusammensetzung besitzen. So wurde in zwei unabhängigen Experimenten nachgewiesen, dass der Bakteriorhodopsingehalt der Deletionsmutanten erheblich geringer ist als im *H. sal.* R1. Dadurch wurde für zwei bisher unbekannte Gene erstmals die indirekte Beteiligung an der BR-Synthese dokumentiert.

In den folgenden Abschnitten soll die weitere Charakterisierung der Deletionsmutanten R1 $\Delta$ 3107 und R1 $\Delta$ 3136 geschildert werden.

Im Rahmen dieses Projektes wurde mittels RT-PCR damit begonnen eine Expressionsanalyse durchzuführen. Die Ergebnisse für die 12 ausgewählten Gene sind in Tab. 19 zu finden.

Rang FS-Statistik	Gen	OE-Nummer	Funktionsklasse	Ergebnis RT-PCR (R1 $\Delta$ 3107F/WT-R1)	Ergebnis RT-PCR (R1 $\Delta$ 3136F/WT-R1)
10	-	OE3107F	HY	-	0,41
213	-	OE3136F	CHY	35,75	-
2	bop	OE3106F	EM	1,07	0,6
11	bat	OE3101R	MIS	1,02	0,39
1273	brp	OE3102R	MIS	3,34	1,45
NA	blp	OE3100F	MIS	0,34	0,19
2411	blh	OE3980R	CHY	1,04	0,39
4	crtB1	OE3093R	LIP	0,23	0,33
560	arcB	OE5205R	AA	1,91	0,93
31	arcC	OE5206R	AA	1,16	0,38
1030	arcA	OE5208R	AA	2,35	0,6
12	carA	OE3556R	AA	0,35	0,3

Tab. 19: Ergebnisse der RT-PCR zur Bestimmung der Unterschiede zwischen den Deletionsmutanten R1 $\Delta$ 3107F und R1 $\Delta$ 3136F und dem Wild-Typ-Stamm R1. Beim Farbkode bedeutet rot = induziert in der R1 $\Delta$ 3107F oder R1 $\Delta$ 3136F Mutante im Vergleich zum WT-R1, gelb = nicht reguliert und grün = reprimiert in der R1 $\Delta$ 3107F oder R1 $\Delta$ 3136F Mutante im Vergleich zum WT-R1,. Die Abkürzungen bedeuten: CHY = konserviert hypothetisches Protein, EM = Energie-Stoffwechsel, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes und NA = „not available“, nicht verfügbar.

Durch die RT-PCR konnte für beide Mutanten ein weiteres Mal die Deletion der ORFs OE3107F und OE3136F nachgewiesen werden, da in den jeweiligen RT-PCR-Reaktionen kein Transkript zu detektieren war.



Weiter ist aus der Tab. 19 klar abzulesen, dass in der Mutante  $\Delta$ 3136F von den 12 Kandidatengen bis auf zwei alle im Vergleich zum Wildtyp *H. sal.* R1 reprimiert sind. Die Ausnahmen sind das *brp*-Gen, das eine Induktion und das *arcB*-Gen das keine Regulation zeigt. Für die Mutante des ORFs OE3107F zeigen die Ergebnisse der RT-PCR einen starken Einfluss auf die Expression von OE3136F mit der höchsten Induktion. Neben dem ORF OE3136F sind die Gene *brp*, *arcA,B,C* in der Mutante induziert, während das *bop*-, *bat*- und *blh*-Gen keine Regulation aufzeigen.

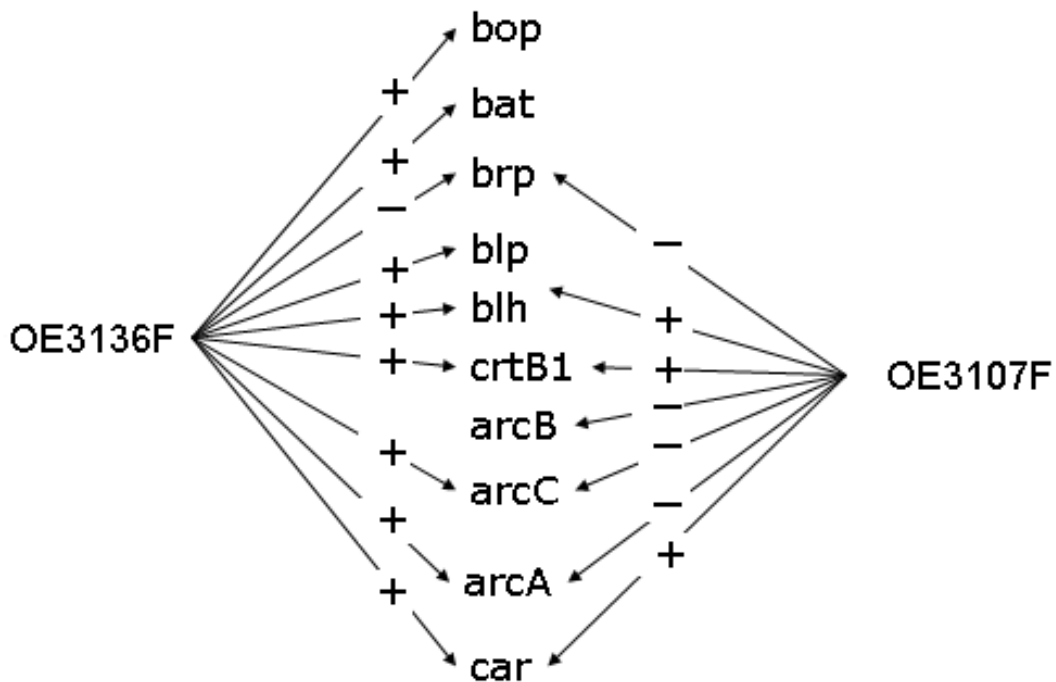


Abb. 69: Grafische Darstellung des Einflusses der Deletion von OE3107F und OE3136F auf die Expression verschiedener Gene, wie z.B. der Gene des *bop*-Gen Clusters. Die Abbildung basiert auf den RT-PCR Ergebnissen aus Tab. 19. „+“ bedeutet Induktion und „-“ bedeutet Repression

Damit bestätigt die nähere Charakterisierung der beiden Mutanten  $\Delta$ OE3107F und  $\Delta$ OE3136F, dass diese Gene einen Einfluss auf die Expression von Bakteriorhodopsin ausüben.

Wie der Abb. 69 und der Tab. 19 zu entnehmen ist, besteht kein direkter Einfluss von OE3107F auf die Expression der Gene *bop*, *bat* und *blh*.

Die Gene *brp*, *blp* und *crtB1* unterliegen jedoch einer direkten Wechselwirkung. Die Gene *brp*, das von OE3107F reprimiert wird, und das *crtB1* Gen, welches induziert wird, sind an der Retinalsynthese beteiligt sein. Daraus läßt sich schließen, dass mit dem Gen OE3107F ein bisher unbekanntes Mitglied der

Regulation der Carotinoid-Biosynthese gefunden wurde. Vielleicht könnte man in diesem Zusammenhang dem *blp* auch eine Beteiligung an der Retinalbiosynthese zuschreiben?

Im Gegensatz zum OE3107F zeigt das Gen OE3136F einen direkten Einfluss auf das *bop*, *bat*, *brp*, *blp*, *blh* und *crtB1* Gen. Damit ist neben der Beteiligung an der Retinalbiosynthese, auch der direkte Einfluss auf die Expression von Bakteriorhodopsin gegeben. Dadurch wird deutlich, dass das bisher bekannte Netzwerk der Regulation der Biosynthese von BR weitaus komplexer ist als bisher angenommen. Außerdem wird offensichtlich, dass ein für *H. sal.* R1 lebensnotwendiger Mechanismus, von einer weitaus größeren Anzahl von Kandidatengenen abhängig ist.

Durch die Entdeckung der Beteiligung dieser beiden Gene am Regulationsnetzwerk der Bakteriorhodopsinbiosynthese, wird für *H. sal.* R1 ebenfalls bewiesen, dass die in diesem Projekt etablierte Transkriptomanalyse mittels der Mikroarray-Technologie, einen hervorragenden Ansatz darstellt zur detaillierteren Untersuchung von bisher unbekanntem Genen und deren Funktion.

Für eine abschließende Expressionsanalyse der beiden Mutanten wäre die Durchführung eines Transkriptom-Experimentes sehr hilfreich.

Neben experimentellen Untersuchungen könnte auch eine intensive bioinformatische Analyse, wie z.B. eine Motivsuche, einen erheblichen Anteil zur Aufklärung der genauen Funktion der Genprodukte liefern. Eine erste Suche nach bekannten Motiven oder Domänen in den Sequenzen von OE3107F und OE3136F hat erbracht, dass in OE3136F mit einem Helix-Turn-Helix-Motiv eine DNA-bindende Domäne enthalten ist. Wobei eine Datenbanksuche nach bekannten Motiven und Domänen für OE3107F bisher erfolglos war.

Das Auffinden einer DNA-bindenden Domäne könnte eine Beteiligung des Gens OE3136F an der Expressionsregulation auf DNA-Ebene vermuten lassen. Zur Bestätigung dieser Vermutung, könnte eine *in-vitro* Expression des Gens durchgeführt werden, um mit dem gereinigten Produkt z.B. einen DNA-Bindungs-Assay auszuführen.

Im Rahmen dieses Projektes wurde für das Gen OE3136F eine *in-vitro* Expression mit dem RTS 100 System der Firma Roche durchgeführt. Das Ergebnis dieser erstmaligen *in-vitro* Expression von OE3136F ist an Hand eines Western-Blots in Abb. 70 zu sehen.

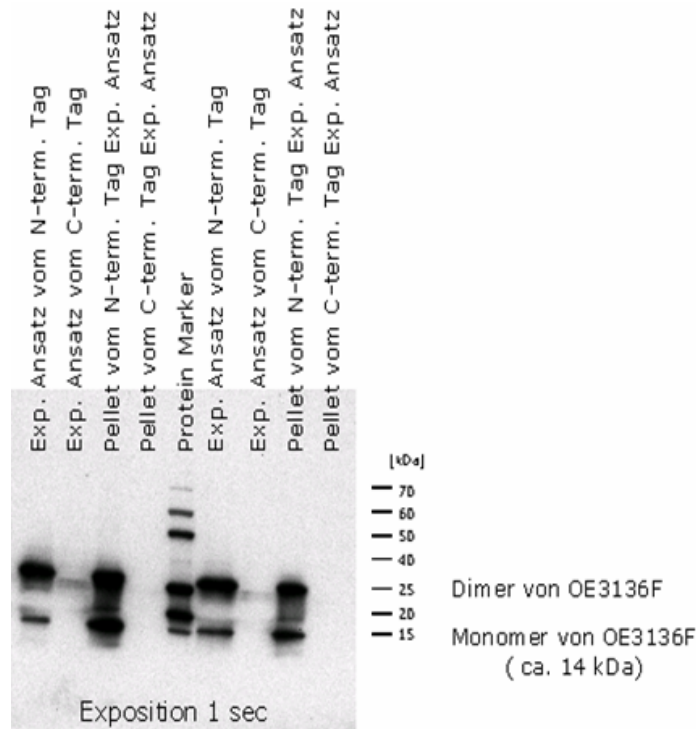


Abb. 70: Western-Blot der in-vitro Biotinylierungs-Expression mit dem RTS 100 System von Gen OE3136F. Die Proben rechts vom Protein Marker sind die gleichen wie auf der linken Seite. Das Monomer des Proteins hat eine Größe von ca. 14 kDa.

Der Abb. 70 ist zu entnehmen, dass eine in-vitro Expression durchgeführt werden kann. Das Konstrukt mit dem N-terminal enthaltenen Avi-Tag ist sowohl löslich als auch im Pellet nachzuweisen, während das Konstrukt mit der C-terminalen Markierung nur schwach als Dimer in löslicher Form zu detektieren ist.

Mit diesem Ergebnis stehen einer weiteren Funktionsanalyse des Genproduktes alle Möglichkeiten offen.

#### 5.4.2 Stoffwechselwege

Die Voraussetzung der Kombination von Transkriptom-Daten mit Stoffwechselwegen wurde im Rahmen der Doktorarbeit von Frau Michaela Falb und der Diplomarbeit von Frau Kerstin Müller geschaffen.

Die rekonstruierten Stoffwechselwege von *H. sal.* R1 sind in der HaloLex-Datenbank angebunden und bildeten die Grundlage die erhaltenen Transkriptom-Daten darauf abzubilden. Zu diesem Zweck wurden die gesamten Informationen des Zweigruppenvergleichs von aerob mit phototroph verwandt, ohne eine Begrenzung des statistischen Rangs vorzunehmen.

Für alle bis zu diesem Zeitpunkt rekonstruierten Stoffwechselwege besteht die Möglichkeit die Ergebnisse der Kombination der Transkriptom-Daten mit den Stoffwechselwegen in der bisher nichtöffentlichen Version der HaloLex-Datenbank abzufragen.

Dadurch existiert neben der Liste der differentiell exprimierten Gene eine weitere Möglichkeit die große Datenmenge der Transkriptom-Daten in Bezug auf eine bestimmte Fragestellung hin abzufragen und darstellen zu können.

Unter der Verknüpfung „pathway tools“ in der HaloLex-Datenbank besteht für *H. sal.* R1 nicht nur die Option die rekonstruierten Stoffwechseldaten in Bezug zu den Transkriptom-Daten darstellenzulassen, sondern auch in Bezug zu Genom- und Proteom-Daten.

**Auswahlmenü zur Selektion des Stoffwechselweges und der Darstellungsform**

**Legende für die grafische Darstellung der Stoffwechselwege**

**Kombination der Transkriptom-Daten mit den beteiligten Genen eines Stoffwechselweges**

EC: 1.1.1.37	Gene: OE4323F	Reg: unreg	ratio: 1.067	pvalue: 0.83356	rank: 980
EC: 1.1.1.41	Gene: OE3634F	Reg: unreg	ratio: 1.087	pvalue: 0.74645	rank: 666
EC: 1.1.1.42	Gene: OE3634F	Reg: unreg	ratio: 1.087	pvalue: 0.74645	rank: 666
EC: 1.2.7.3	Gene: OE1711R	Reg: up	ratio: 1.180	pvalue: 0.50222	rank: 209
EC: 1.2.7.3	Gene: OE1710R	Reg: up	ratio: 1.268	pvalue: 0.21379	rank: 60
EC: 1.3.5.1	Gene: OE2866R	Reg: up	ratio: 1.132	pvalue: 0.74056	rank: 604
EC: 1.3.5.1	Gene: OE2868R	Reg: down	ratio: 0.794	pvalue: 0.63608	rank: 388
EC: 1.3.5.1	Gene: OE2867R	Reg: unreg	ratio: 0.918	pvalue: 0.61676	rank: 323
EC: 1.3.5.1	Gene: OE2865R	Reg: unreg	ratio: 1.011	pvalue: 0.98855	rank: 2269
EC: 1.3.99.1	Gene: OE2866R	Reg: up	ratio: 1.132	pvalue: 0.74056	rank: 604
EC: 1.3.99.1	Gene: OE2868R	Reg: down	ratio: 0.794	pvalue: 0.63608	rank: 388
EC: 1.3.99.1	Gene: OE2867R	Reg: unreg	ratio: 0.918	pvalue: 0.61676	rank: 323
EC: 1.3.99.1	Gene: OE2865R	Reg: unreg	ratio: 1.011	pvalue: 0.98855	rank: 2269
EC: 1.8.1.4	Gene: OE4116F	Reg: ratio:		pvalue:	rank:
EC: 2.3.3.1	Gene: OE3934R	Reg: unreg	ratio: 1.024	pvalue: 0.96024	rank: 1928
EC: 2.3.3.3	Gene: OE3934R	Reg: unreg	ratio: 1.024	pvalue: 0.96024	rank: 1928

Abb. 71: Screenshot einer Abfrage von der Kombination der Transkriptom-Daten mit einem Stoffwechselweg in der HaloLex-Datenbank (Link zur Zeit noch nicht öffentlich).

Wie in Abb. 71 gezeigt, kann in einem Auswahlmenü ein bestimmter Stoffwechselweg selektiert werden und der Bezug in welchem die Daten dargestellt werden sollen ausgewählt werden.

Zur näheren Erläuterung der Darstellung ist eine Legende für die graphische Ausgabe sowie eine Auflistung mit Ergebnissen der Transkriptomanalyse aller beteiligten Gene des ausgewählten Stoffwechselweges enthalten.

An diese Auflistung der Gene schließt sich die graphische Darstellung des Stoffwechselweges in Form der Abbildungen der so genannten „KEGG Karten“ an. KEGG steht dabei für: **K**yoto **E**ncyclopedia of **G**enes and **G**enomes und repräsentiert eine der größten öffentlichen Datenbanken für Stoffwechselwege und deren Zusammenhänge.

Beispiele für Abfragen der Daten in Bezug auf alle Transkriptom-Daten sind in dem folgenden Abbildungen gezeigt.

Als Beispiel der Kombination von Transkriptom-Daten mit Stoffwechselwegen ist die Biosynthese der Carotinoide in Abbildung 72 als Ausschnitt der KEGG-Karte zur Biosynthese der Steroide gezeigt.

Man sieht, dass unter phototrophen Bedingungen nur die Gene *crtB1* und *crtB2* hoch reguliert werden.

Der dargestellte Stoffwechselweg der Biosynthese der Carotinoide ist für *H. sal.* R1 deshalb von Bedeutung, da die gezeigten Enzyme die Vorstufe des Retinals bilden, welches als prosthetische Gruppe in das Apoprotein Bakteriorhodopsin als Schiff'sche Base eingebaut wird, um die funktionelle lichtgetriebene Protonenpumpe zu bilden.

Somit steht die Induktion der Phytoensynthesen *crtB1,B2* im Einklang mit der Induktion des Apoproteins Bakteriorhodopsin und liefert einen Beweis für die Signifikanz der Transkriptom-Analyse von *H. sal.* R1.

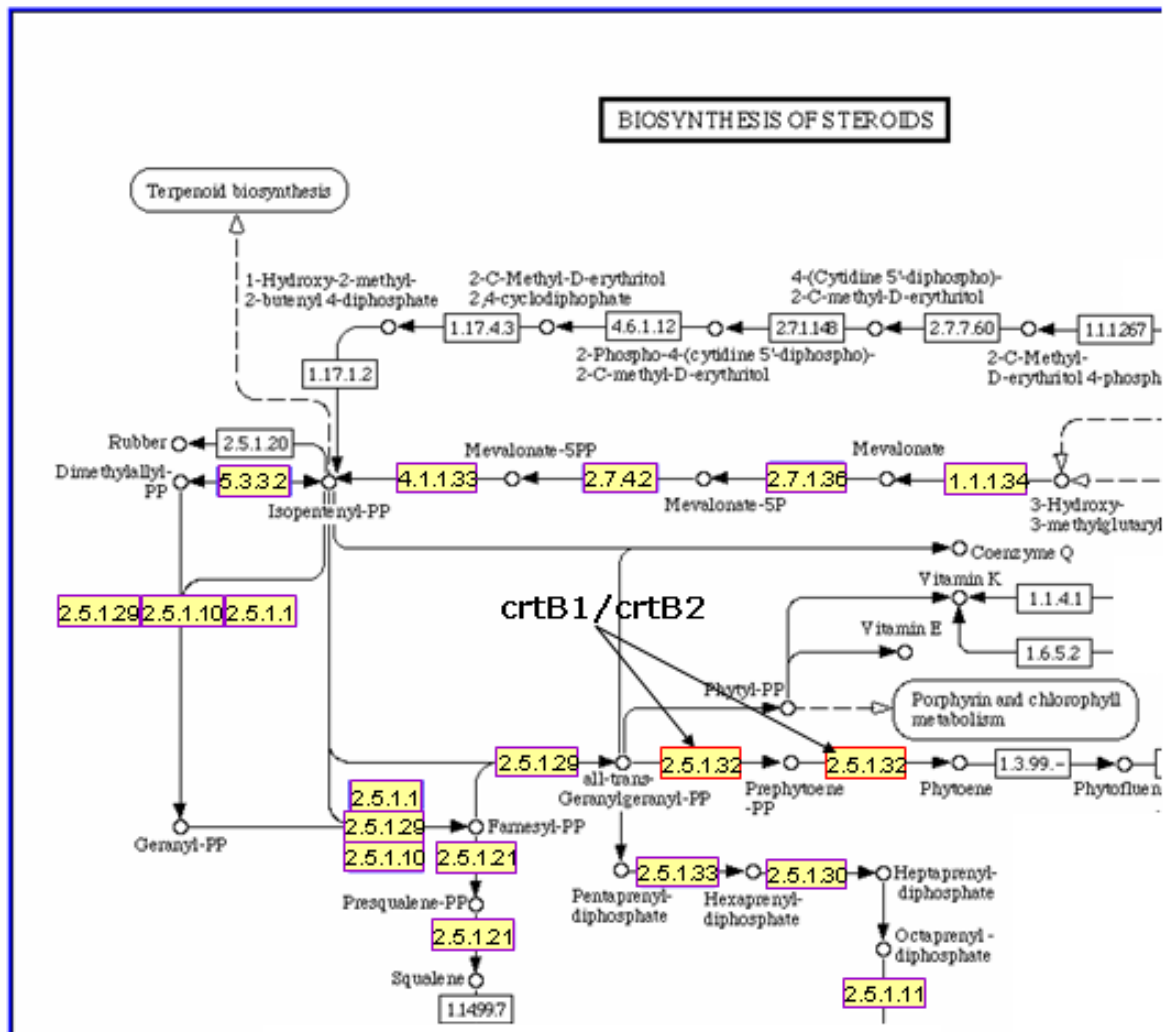


Abb. 72: Regulation von Enzymen, Isoprenoidsynthese beim Übergang zum phototropem Wachstum. Gelbgefüllte Kästchen mit EC-Nummern stehen für vorhandene Gene und Reaktionen in *H. sal.* R1, während die nicht gefüllten (weißen) Kästchen noch nicht näher analysierte Reaktionen und Gene repräsentieren. Farbcode: rot weist auf eine Induktion und grün auf eine Reduktion hin. Gene die keine Regulation aufweisen sind in violett dargestellt.

Die bioinformatische Analyse der Genomsequenz von *H. sal.* R1 hat ergeben, dass die notwendigen Enzyme für einen reduktiven Citrat-Zyklus vorhanden sind und deshalb seine Existenz als wahrscheinlich anzusehen ist.

Bildet man die Transkriptom-Daten des Zweigruppenvergleichs auf den Citrat-Zyklus ab, so sieht man an Abb. 73 dass zwei beteiligte Enzyme, die Aconitathydratase und die 2-Oxoglutaratsynthase, unter phototropen Bedingungen induziert sind, während die Mehrzahl der anderen Enzyme keine Regulation zeigt.

Die Succinatdehydrogenase, die aus den Ketten A-D (*sdhA-D*) besteht, zeigt in der Expressionsanalyse ein uneinheitliches Bild. So sind die Ketten A und D nicht

reguliert, während die Kette B leicht induziert und die Kette C leicht reprimiert ist. Alle vier beteiligten Gene zeigen einen tiefen statistischen Rang, was auf ein wenig signifikantes Ergebnis hindeutet. Ein ähnliches Bild ergibt sich für die Pyruvatsynthese (*porA,B*), die für eine Kette leichte Induktion und für die andere keine Regulation nach der statistischen Analyse zeigt.

Dieses Beispiel zeigt anschaulich, dass die Transkriptom-Analyse und deren Interpretation leicht an ihre Grenzen stößt. Um eine eindeutige Aussage über die Regulation der Succinatdehydrogenase und der Pyruvatsynthese treffen zu können, sind weiterführende z.B. RT-PCR Experimente nötig.

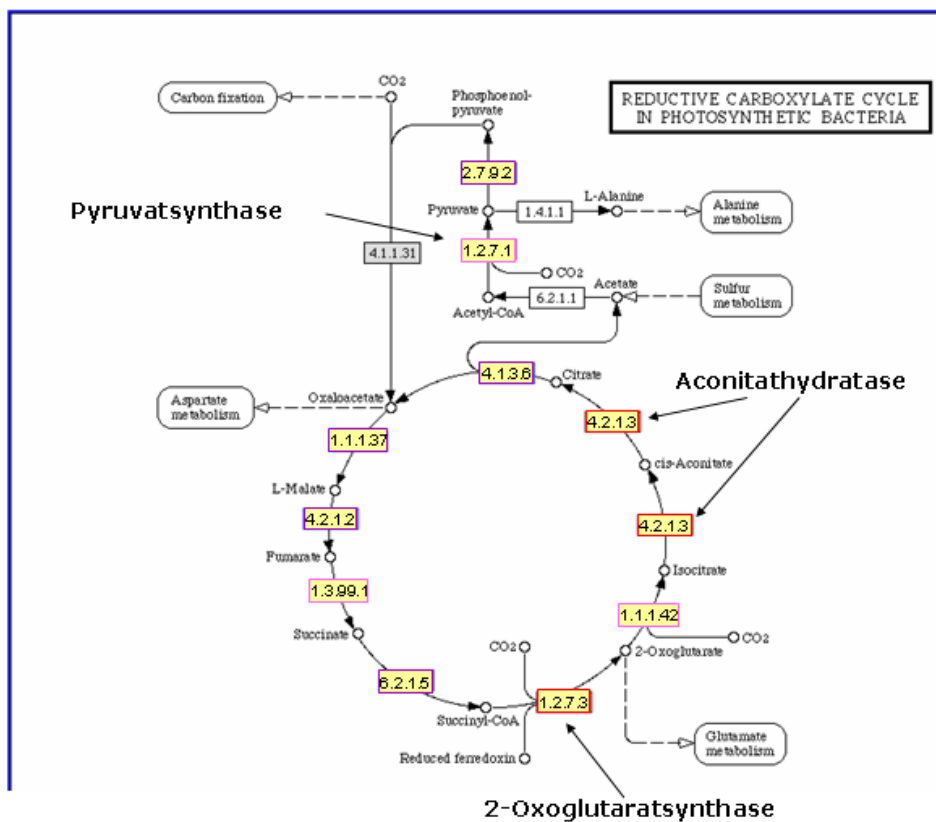


Abb. 73: Dargestellt ist der Weg des reduktiven Citrat-Zyklus in *H. sal.* R1 in Verbindung mit Informationen der Transkriptomanalyse. Gelbgefüllte Kästchen mit EC-Nummern stehen für vorhandene Gene und Reaktionen in *H. sal.* R1, während die grau gefüllten Kästchen Gene repräsentieren, die nicht in *H. sal.* R1 existieren und die nicht gefüllten (weißen) Kästchen für noch nicht näher analysierte Reaktionen und Gene. Farbkode: rot weist auf eine Induktion und grün auf eine Reduktion hin. Gene die keine Regulation aufweisen sind in violett dargestellt, während pink auf einen Proteinkomplex verweist dessen Transkriptom-Daten nicht konsistent sind.

### 5.4.2.1 Diskussion

Durch das Abbilden der Expressionsdaten des Zweigruppenvergleichs auf die Stoffwechselwege von *H. sal.* R1 ist eine weitere Verknüpfung von zwei Disziplinen, der Bioinformatik und der Transkriptomik, erfolgt.

Es zeigt sich deutlich, dass die modernen wissenschaftlichen Teilgebiete immer stärker verzahnt sind und im Grunde nicht mehr isoliert existieren können.

Die Kombination der validierten experimentellen Daten mit den rekonstruierten Stoffwechselwegen bildet einen wichtigen Baustein zur Modellierung des Stoffwechsels einer Zelle.

Da viele Bereiche des Stoffwechsels von *H. sal.* R1 zur Zeit noch nicht vollkommen verstanden sind, können diese Ergebnisse dazu dienen, z.B. für die Spate der Metabolomik, potentielle experimentelle Ansätze zu liefern. D.h. statt nach der Nadel im Heuhaufen zu suchen, können gezielt Teilbereiche von Stoffwechselwegen untersucht werden, um so Schritt für Schritt die Lücken zu füllen. Denn wiederum nur mit sicheren experimentellen Daten kann es der Systembiologie in Zukunft gelingen die erhaltenen einzelnen Bausteine zu einem kompletten Modell der Zelle zusammenzufügen. Mit den Ergebnissen dieser Arbeit ist dazu der Grundstein für *H. sal.* R1 gelegt worden.

### 5.4.3 Analyseansatz der S9H-Mutante

Die Mutante mit dem Namen S9H ist aus dem Stamm S9 durch eine nicht gezielte spontane Mutation entstanden. Für den Mutterstamm S9 ist die violette Färbung auch unter aeroben Wachstumsbedingungen charakteristisch, da dieser Stamm ein konstitutiver Überproduzent der lichtgetriebenen Protonenpumpe Bakteriorhodopsions (BR) ist. Im Gegensatz dazu findet im WT *H. sal.* R1 Stamm die Induktion des BRs in der stationären Wachstumsphase oder unter phototrophen Wachstumsbedingungen statt, also unter starker Lichteinstrahlung und / oder Sauerstoffausschluss.

Die so genannte Mutante S9H ist dadurch entstanden, dass S9-Zellen unter phototrophen Bedingungen bis zur stationären Wachstumsphase kultiviert wurden. Ein Aliquot dieser Kultur diente als Inokulum für eine neue phototrophe S9-Kultur und ein weiteres Aliquot wurde auf Kulturplatten ausgebracht und unter aeroben Bedingungen solange inkubiert bis deutlich sichtbare Klone gewachsen sind. Der Zyklus des phototrophen Wachstums bis zur stationären



Wachstumsphase und des erneuten Kultivierens, wurde 10-11 mal wiederholt. Die Kulturplatten der 10. und 11. Runde zeigten eine ca. 50/50-Verteilung von violetten und farblosen Klonen. Von diesen Kulturplatten wurden violette und farblose Klone vereinzelt und kultiviert, bis von den farblosen Klonen der stabile Stamm S9H hervorgegangen ist (siehe Methoden).

Im Rahmen der näheren Charakterisierung der farblosen Mutante S9H wurden Wachstumskurven unter aeroben und phototrophen Bedingungen im Vergleich zum Ausgangsstamm S9D aufgenommen, die in der folgenden Abb. 67 zu sehen sind.

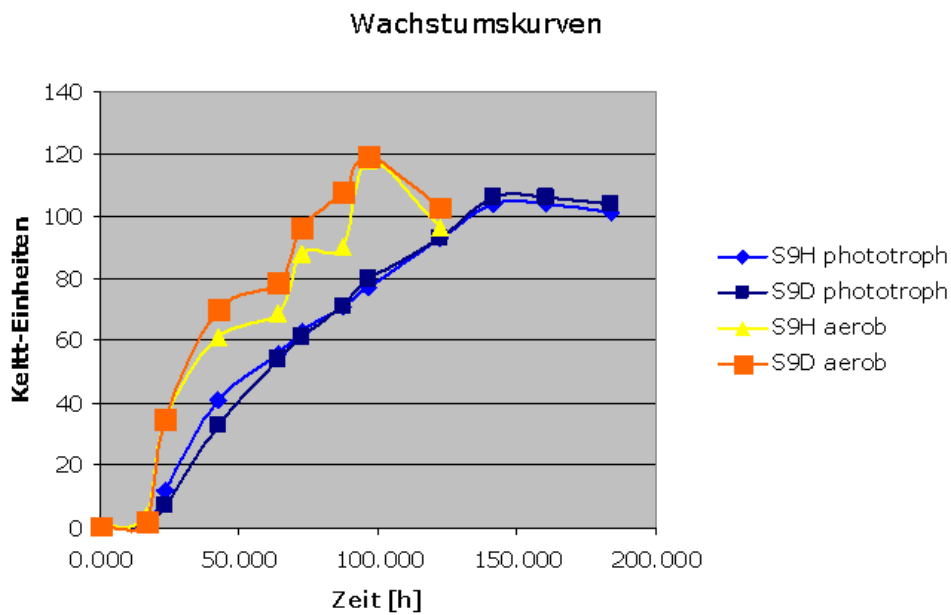


Abb. 74: Wachstumskurven der Stämme S9H und S9D, die unter aeroben oder phototrophen Wachstumsbedingungen kultiviert wurden.

Die Wachstumskurven aus der Abb. 74 lassen keine Unterschiede im Wachstumsverhalten der beiden Stämme erkennen.

Beim Vergleich von Photographien aerob und phototroph gewachsener Kulturen der Stämme S9H und S9D sieht man, dass die phototrophe Kultur von S9H sehr stark violett gefärbt ist. Dies zeigt, daß die Induzierbarkeit des Bakteriorhodopsions in dieser Mutante der des Wildtyps entspricht, d.h. Die konstitutive Expression in S9D wieder zu induzierbarem in S9H revertiert ist.



Abb. 75: Photographien der Stämme S9H / S9D, die aerob oder phototroph kultiviert wurden. Aerobes Wachstum: 1=S9D, 3=S9H; Phototrophes Wachstum: 2=S9D, 4=S9H.

Die starken Unterschiede in der Pigmentierung der Kulturen sind in der (Abb. 75) eindeutig zu sehen. Die größten Unterschiede weisen die Kulturen in den Kolben 1 und 3 auf. Die Kultur 3 (S9H) ist gelblich, während die Kultur 1 die typische violette Färbung einer S9-Kultur zeigt. Beim Vergleich der phototroph gewachsenen Kulturen in den Kolben 2 und 4 sind nur kleine Unterschiede in der Färbung auszumachen. Als Nachweis der Induzierbarkeit von BR in dem S9H-Stamm, wurde mit allen vier in (Abb. 75) dargestellten Kulturen ein BR-Test durchgeführt.

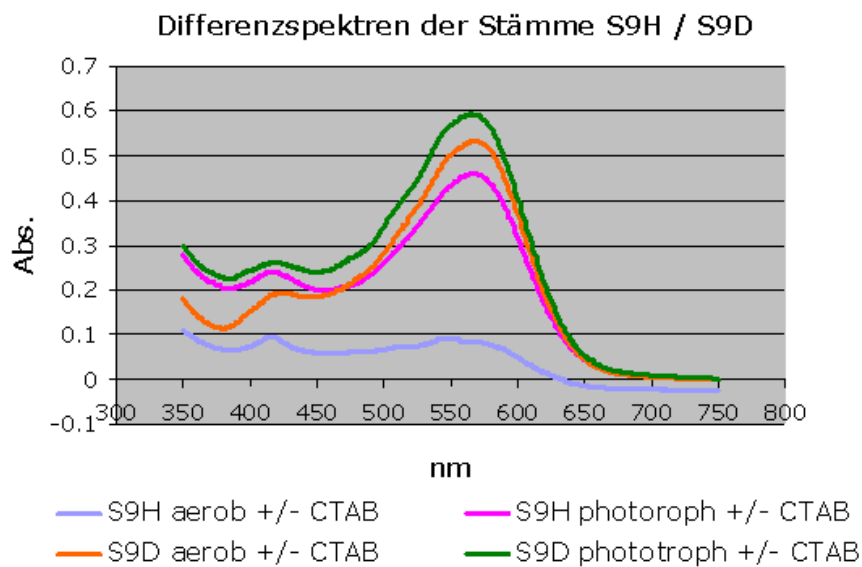


Abb. 76: Differenzspektren nach CTAB-Behandlung von Zelllysaten der Stämme S9H und S9D, die aerob oder phototroph kultiviert wurden.

Die Ergebnisse belegen, dass wie erwartet die violette Färbung der S9H Kultur durch eine stark erhöhte Bakteriorhodopsin-Expression hervorgerufen wurde. An

Hand des pinken Graphen aus (Abb. 76) ist die drastische Zunahme des BR-Gehalts der phototrophen Kultur von S9H eindeutig dokumentiert.

Aus diesen Ergebnissen kann geschlossen werden, dass die Mutante S9H sich wie der *H. sal.* R1-Wildtyp verhält und den Phänotyp des konstitutiven BR-Überproduzenten vom S9D-Stamm verloren hat.

Auf der Grundlage dieser Ergebnisse stellt die Mutante S9H eine ausgezeichnete Chance dar, mit Hilfe der Transkriptom-Analyse der Aufklärung der genauen Regulation des phototrophen Wachstums von Halobakterien einen Schritt näher zu kommen. Zu diesem Zweck wurde im Rahmen eines dye-swap DNA-Mikroarray-Experimentes ein Vergleich der Stämme S9H und S9D durchgeführt. Dabei wurden fluoreszenzmarkierte cDNAs von aerob kultivierten Zellen von S9H und S9D auf dem Gesamtgenom-Mikroarray von *H. sal.* R1 hybridisiert.

Die Auswertung dieser DNA-Mikroarray-Experimente wurde nicht anhand der beschriebenen statistischen Auswertung gemacht, sondern auf Grundlage des dye swap Experimentdesigns auf der Basis einer Excel-Berechnung. Details sind in der Promotion von Herrn Andy Wende zu finden.

Es fanden sechs Hybridisierungen statt, drei mit jeder Farbkombination. Die Ergebnisse dieser Auswertung werden in den folgenden Abschnitten vorgestellt.

## Ergebnisse und Diskussion

OE-Nummer	Funktionsklasse	Genname	Regulation	Proteinname
OE2173F	AA	cgs	1,94	cystathionine gamma-lyase (EC 4.4.1.1)
OE4408F	AA	serA1	1,68	phosphoglycerate dehydrogenase (EC 1.1.1.95)
OE4759F	CE	csg	2,15	cell surface glycoprotein precursor
OE3136F	CHY	-	2,28	conserved hypothetical protein
OE3822R	CHY	-	13,55	conserved hypothetical protein
OE4223R	CHY	-	2,1	conserved hypothetical protein
OE2473F	CHY	-	2,26	conserved hypothetical protein
OE5276F	CHY	-	1,81	conserved hypothetical protein
OE1789R	CHY	-	3,43	conserved hypothetical protein
OE1156F	CP	hsp20	2,1	heat shock protein 20 homolog
OE3925R	CP	cctB	2,06	thermosome beta chain
OE5125F	CP	gvpA2	1,88	gas-vesicle protein gvpA2
OE5125F	CP	gvpA2	2,13	gas-vesicle protein gvpA2
OE3106F	EM	bop	2,43	bacteriorhodopsin precursor
OE3545F	HY	-	2,35	hypothetical protein
OE3342R	HY	-	1,95	hypothetical protein
OE3314R	HY	-	4,21	hypothetical protein
OE8007F	HY	-	1,71	hypothetical protein
OE5163R	HY	-	4,25	[del] hypothetical protein (overlaps a more likely protein)
OE5032R	HY	-	2,78	hypothetical protein
OE3766R	HY	-	2,23	hypothetical protein
OE3942R	HY	-	2,17	hypothetical protein
OE1409F	HY	-	2,36	hypothetical protein
OE1971F	HY	-	1,77	hypothetical protein
OE2160R	HY	-	1,89	hypothetical protein
OE3107F	HY	-	3,51	hypothetical protein
OE2421R	HY	-	2,4	hypothetical protein
OE1040R	ISH	-	3,56	probable IS1341-type transposase (ISHnew88)
OE1019R	ISH	-	2,97	probable IS1341-type transposase (ISHnew88)
OE3468R	LIP	crt12	4,85	phytoene dehydrogenase (EC 1.14.99.-) (phytoene desaturase) 2 (alternative EC 1.3.-.-)
OE3093R	LIP	crtB1	3,62	geranylgeranyl-diphosphate geranylgeranyltransferase (EC 2.5.1.32) (phytoene synthase)
OE1465F	MIS	lon	2,62	endopeptidase La (EC 3.4.21.53)
OE5118R	MIS	gvpI2	1,88	gas-vesicle operon protein gvpI2
OE7033R	MIS	gvpD1	2,08	gas-vesicle operon protein gvpD1 (probable repressor protein)
OE3312R	MIS	oxr3	1,8	probable oxidoreductase (EC 1.1.1.-) (aryl-alcohol dehydrogenase homolog)
OE2853R	MIS	slyD	1,88	probable peptidylprolyl isomerase (EC 5.2.1.8)
OE4065R	MIS	hit1	1,81	histidine triad family protein
OE2708R	MIS	sod1	2,47	superoxide dismutase (EC 1.15.1.1) 1 (Mn containing, validated)
OE3542R	NOF	-	2,52	protein OE3542R
OE2252R	NOF	-	1,66	conserved protein
OE2515F	NOF	-	2,1	conserved protein
OE5106F	NOF	trkA7	1,95	trkA domain protein
OE2186R	NOF	-	3,47	conserved protein
OE1475F	NOF	-	1,95	conserved protein
OE1797R	REG	sirR	1,76	transcription regulator sirR

Tab. 20: Liste mit den Genen die beim Vergleich von S9H mit S9D reprimiert sind. Blaue Färbung: Gene mit stärkster Regulation; orangene Färbung: Übereinstimmung in drei Wiederholungen; hellgelbe Färbung: Übereinstimmung in zwei Wiederholungen.

In Tab. 20 sind die reprimierten Gene der S9H Mutante aufgelistet. Unter den reprimierten Kandidatengen sind das *bop*-, *crtB1*- und *crt12*-Gen zu finden, was die Ergebnisse der vorher durchgeführten Experimente bestätigt. Denn all diese Gene sind an der Pigmentsynthese in *H. sal.* R1 beteiligt. Da das *crt12*-Gen das am zweit stärksten reprimierte Gen darstellt deutet dies auf eine entscheidende Rolle in der Carotinoidbiosynthese hin.

Die stärkste Regulation zeigt das Gen OE3822R. Da dieses Gen zur Funktionsklasse CHY gehört, kann über seine Funktion und damit über die Rolle bei der Regulation des phototrophen Wachstums zur Zeit keine weitere Aussage getroffen werden. Auf Grund dessen stellt dieses Gen einen Kandidaten zur näheren Analyse in späteren Experimenten dar.

Analog zu den Transkriptom-Experimenten von *H. sal.* R1, zeigt sich auch in diesem ersten Experiment, dass der Anteil der drei Funktionsklassen CHY, HY und NOF mit 78,1% sehr hoch ist.

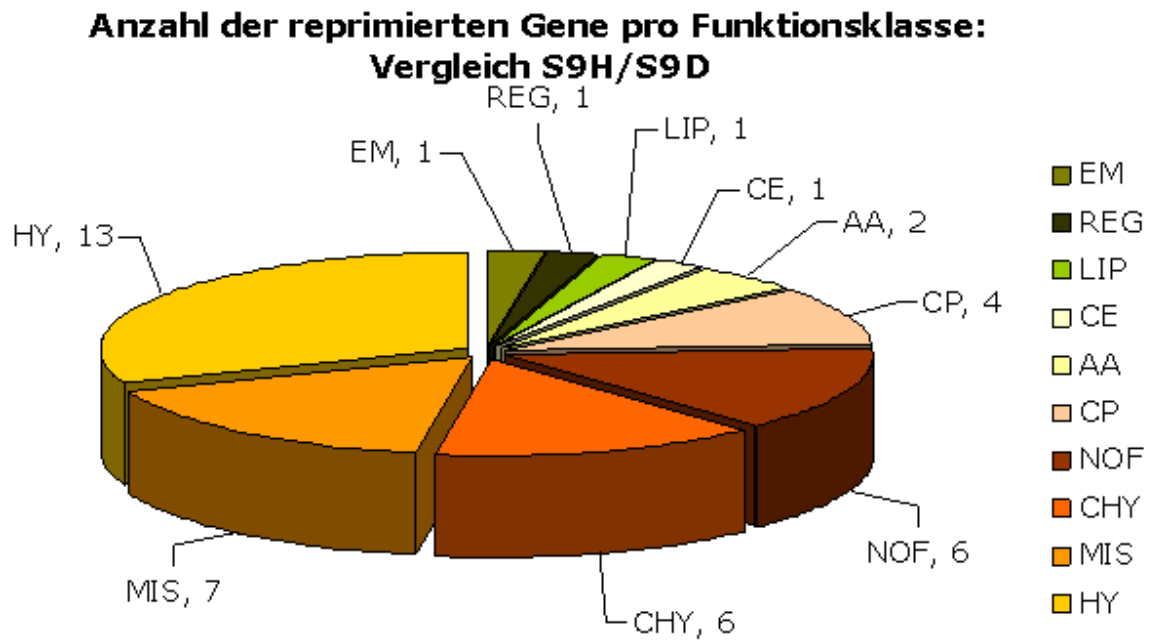


Abb. 77: Tortendigramm der reprimierten Gene (S9H) der Expressionsanalyse vom Vergleich der Stämme S9H und S9D. Die Anzahl der Gene pro Funktionsklasse ist neben den Abkürzungen angegeben. Die Abkürzungen bedeuten: AA = Aminosäurestoffwechsel, CE = Zellhülle, CHY = konserviert hypothetisches Protein, CP = zelluläre Prozesse, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, ISH = ISH-kodierte Transposasen, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes, NOF = noch keine Funktion, REG = Genregulation.

Bei Betrachtung der Kandidatengene der Funktionsklasse MIS, die einen Anteil von 21,9% der reprimierten Gene besitzt, findet man mit Gasvesikelgenen und z.B. den vier Kandidatengenen *hit1*, *slyD*, *sod1* und *oxr3* eine gute Übereinstimmung mit den Ergebnissen des Transkriptom-Experiments der zweiten Zeitreihe. Ein weiterer Kandidat zur näheren Analyse stellt der Transkriptionsregulator *sirR* da. Denn dieser wurde ebenfalls parallel zum Zeitreihenexperiment identifiziert.

Die zweite Hälfte der regulierten Gene des Expressionsvergleichs von S9H mit S9D ist in Tab. 21 zu finden. Auffällig an dieser Liste ist, dass sie mit 4 *fla*-Genen (*flaD*, *E*, *I* und *K*) und 5 *flg*-Genen (*flgA1*, *A2*, *B1*, *B2* und *B3*) sehr viele Gene enthält, die an der Motilität und der Flagellinsynthese beteiligt sind.

Im Hinblick auf eine mögliche Klärung der Regulation ist ebenfalls wichtig, dass mit *cheJ* und *cheJ3* zwei Sensorkinasen und mit *cheR* die Methyltransferase als induzierte Gene detektiert wurden.

OE-Nummer	Funktionsklasse	Genname	Regulation	Proteinname
OE2410R	CHY	cheJ3	2.9	taxis protein cheJ homolog
OE7065F	EM	cydA1	2.75	cytochrome d ubiquinol oxidase (EC 1.10.3.-) chain I
OE3991R	EM	atpI	1.86	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain
OE3992R	EM	atpH	1.97	H <sup>+</sup> -transporting two-sector ATPase (EC 3.6.3.14) chain
OE2706R	HY	-	1.85	hypothetical protein
OE3842R	HY	-	1.74	hypothetical protein
OE6034F	ISH	-	2.24	probable IS1341-type transposase (ISHnew98)
OE1934R	MIS	edp	1.9	proteinase IV (EC 3.4.-.-)
OE2378R	MIS	flaK	5.09	fla operon protein flaK (parA domain protein)
OE2373F	MIS	pta	5.32	probable phosphate acetyltransferase (EC 2.3.1.8)
OE2390R	MIS	flaD	1.96	flagella accessory protein
OE2386R	MIS	flaE	12.42	flagella accessory protein
OE2397F	MOT	flgB1	2.2	flagellin B1 precursor
OE2470F	MOT	flgA2	2.11	flagellin A2 precursor
OE2399F	MOT	flgB3	2.21	flagellin B3 precursor
OE2398F	MOT	flgB2	2.2	flagellin B2 precursor
OE2469F	MOT	flgA1	1.94	flagellin A1 precursor
OE7187F	NOF	-	2.28	conserved protein
OE5089F	NOF	-	2.02	conserved protein
OE5083R	NOF	hsp5	2.81	heat shock protein homolog
OE2402F	NOF	-	5.43	conserved protein
OE2404R	NOF	-	3.48	protein OE2404R
OE2401F	NOF	cpcE	9.96	phycocyanin alpha phycocyanobilin lyase homolog
OE2380R	SEC	flaI	4.72	flagella biogenesis protein flaI (probable flagellin
OE2406R	SIG	cheR	3.88	protein-glutamate O-methyltransferase (EC 2.1.1.80)
OE2414R	SIG	cheJ	3.85	taxis protein cheJ
OE7062R	TC	tbpB1a	1.67	TATA-binding transcription initiation factor

Tab. 21: Liste mit den Genen die beim Vergleich von S9H mit S9D induziert sind. Blaue Färbung: Gene mit stärkster Regulation; orangene Färbung: Übereinstimmung in drei Wiederholungen; hellgelbe Färbung: Übereinstimmung in zwei Wiederholungen.

Bei der Statistik der Verteilung der induzierten Genen auf die Funktionsklassen von *H. sal.* R1 ist in der Abb. 78 dargestellt. Aus der Abbildung ist abzulesen, dass der Anteil der Funktionsklassen CHY, HY und NOF bei 33,3% liegt.

Neben den bereits erwähnten Gene der Funktionsklassen MOT, die einen Anteil von 18,5% aufweist und der Funktionsklasse SIG mit einem Anteil von 6,3%, ist die Funktionsklasse EM mit 3 Kandidaten und 9,4% die nächst größere Gruppe. Zu diesen Genen zählen zwei Untereinheiten der ATP-Synthase (*atpH* und *I*) und die Kette I der Cytochrome-d-ubiquinoloxidase (*cydA1*).

### Anzahl der induzierten Gene pro Funktionsklasse: Vergleich S9H/S9D

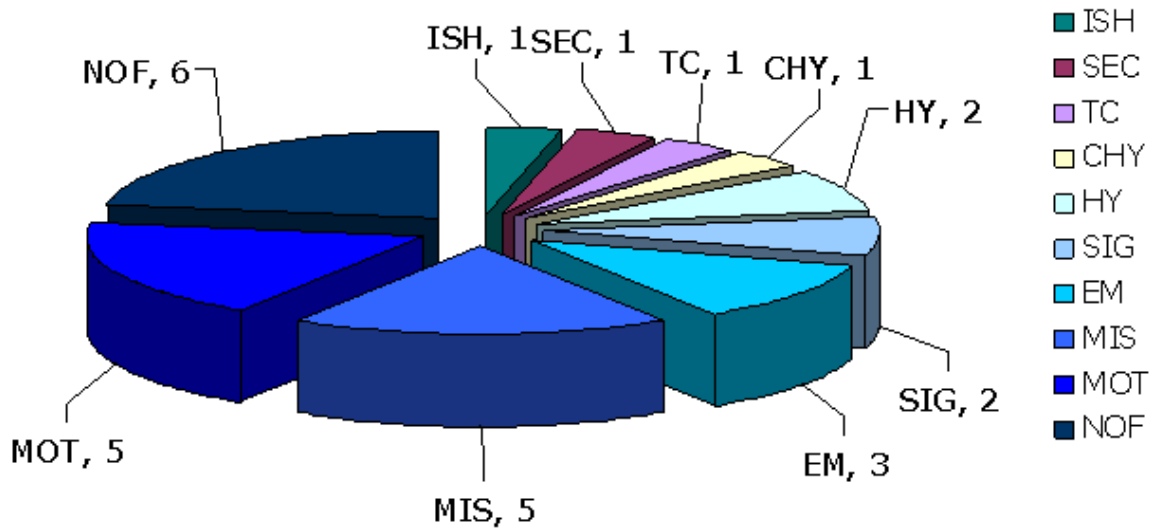


Abb. 78: Tortendigramm der induzierten Gene der Expressionsanalyse vom Vergleich der Stämme S9H und S9D. Die Anzahl der Gene pro Funktionsklasse ist neben den Abkürzungen angegeben. Die Abkürzungen bedeuten: CHY = konserviert hypothetisches Protein, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, ISH = ISH-kodierte Transposasen, MIS = Verschiedenes, MOT = Motilität, NOF = noch keine Funktion, SEC = Sekretion, TC = Transkription.

#### 5.4.3.1 Diskussion:

Zur Diskussion der Ergebnisse, muss beachtet werden, dass das phototrophe Wachstum keine Kombination der Signale Mangel an Sauerstoff und starker Lichtintensität ist. Deshalb warfen die Ergebnisse die Frage auf, ob der Mangel an Sauerstoff oder das Vorhandensein von Licht die Induktion von BR in der Mutante induziert. Zur Beantwortung dieser Frage wurde ein Experiment zur möglichen Untersuchung einer Entkoppelung des Licht- und Sauerstoffsignales durchgeführt.

Dazu wurden je drei Kulturen der Stämme S9H und S9D unter aeroben Bedingungen kultiviert. Wiesen die Kulturen eine Zelldichte von 30-40 Klett-Einheiten auf, wurden zwei Kolben mit einem Gummistopfen luftdicht verschlossen und alle Kolben für weitere 24h im Dunkeln inkubiert. Dadurch ist gewährleistet, dass sich in den luftdicht verschlossenen Kolben kein Sauerstoff mehr befindet. Nach Ablauf von weiteren 24h wurde einer der luftdicht verschlossenen Kolben in einen Lichtschüttler gestellt, während die anderen

Kolben im Dunkeln inkubiert wurden. Alle drei Kulturen wurden unter den jeweiligen Bedingungen für weitere 24h inkubiert. Zur Veranschaulichung des Versuchsaufbau, ist dieser in Abb. 79 dargestellt.

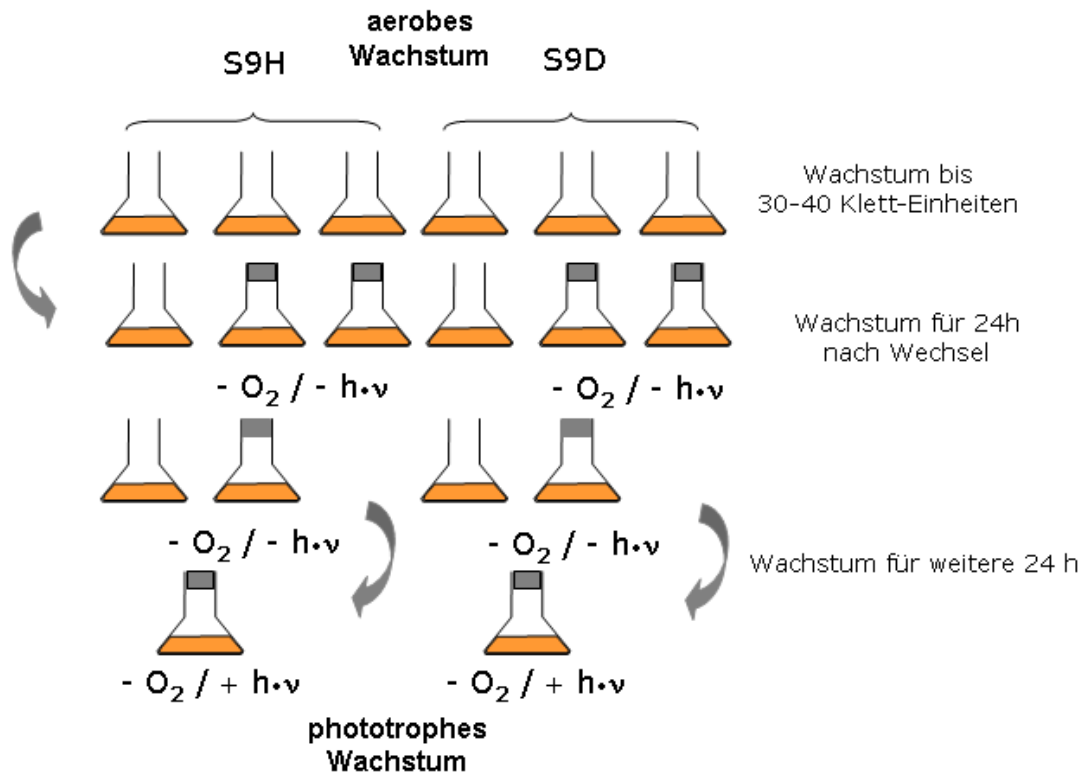


Abb 79: Versuchsaufbau zur Entkoppelung des Licht- und des Sauerstoffsignals.

Die Zellysate der so kultivierten Kulturen der Stämme S9H und S9D wurden anschließend zu einer Untersuchung der Pigmentzusammensetzung mittels eines Saccharosegradienten herangezogen.



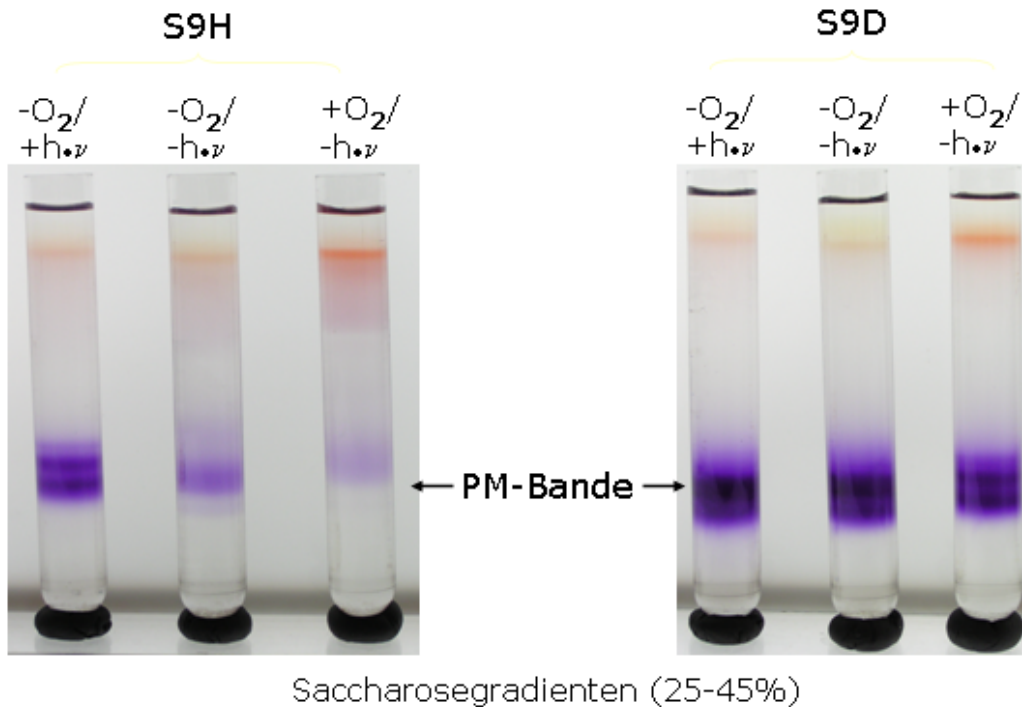


Abb. 80: Photographie von kontinuierlichen Saccharosegradienten (25-45%) auf denen das Zelllysat verschiedener Stämme nach dem Versuch der Entkoppelung des Licht- und des Sauerstoffsignals aufgebracht wurde. Die Wachstumsbedingungen sind über den Röhrrchen angegeben.

Wie in der (Abb. 80) angegeben sind links die Gradienten mit den phototroph gewachsenen Kulturen, in der Mitte die Gradienten der Kulturen die fermentativ und rechts die Gradienten der Kulturen die aerob gewachsen sind.

Als Ergebnis dieses Versuchs kann festgehalten werden, dass erhebliche Unterschiede in der Expression von Purpurchromophoren (PM) bei der Mutante S9H zu beobachten sind. So tritt aufgrund des Sauerstoffmangels in der fermentativ gewachsenen Kultur eine Induktion der BR-Synthese im Vergleich zur aerob gewachsenen Kultur ein. Diese Induktion ist im Verhältnis zur phototroph gewachsenen Kultur allerdings nicht so groß. Die Lichtinduktion scheint also einen stärkeren Einfluß auf die BR-Synthese zu haben, als die Induktion durch Sauerstoffmangel.

Bei dem S9D-Stamm ist diese Regulation offensichtlich verloren und eine fast konstante maximale BR-Konzentration vorhanden. Nur unter phototrophen Bedingungen wird eine leichte Zunahme der PM-Bande beobachtet.

Eine Untersuchung der Entkopplung des Licht- und Sauerstoffsignals wurde auch auf RNA-Ebene mittels RT-PCR durchgeführt. Dazu wurde neben den beiden Stämmen S9H und S9D auch der Wildtyp *H. sal.* R1 nach der in (Abb. 79) beschriebenen Weise kultiviert und Gesamt-RNA isoliert. Die Ernte der Zellen für

diesen Versuch erfolgte gleichzeitig für alle Stämme. Hieraus resultiert das Problem, dass die Zelldichten je nach Kulturbedingung differieren. Während sich die fermentativ und phototroph gewachsenen Kulturen noch in der logarithmischen Wachstumsphase befinden, ist die aerob gewachsene Kultur bereits deutlich in der stationären Wachstumsphase. Dies bedeutet, dass somit in dem RT-PCR-Experiment nicht nur die verschiedenen Wachstumsbedingungen, sondern auch die unterschiedlichen Wachstumsphasen einen Einfluss auf das Ergebnis haben. Diese Tatsache ist dem Abschnitt 5.4.1 schon beschrieben worden. Ein sinnvoller Vergleich ist unter diesen Voraussetzungen nicht durchzuführen, da die Wachstumsphase einen erheblichen Einfluss auf die Transkriptmenge ausübt.

Deshalb wurden die Zellen der drei Stämme unter allen Bedingungen nur bis zur Zelldichte von 30-40 Klett kultiviert und dann die RNA extrahiert.

Die Ergebnisse dieser RT-PCR-Experimente für die drei Stämme sind in den Tabellen 22 bis 24 zusammengestellt.

Rang FS-Statistik	Gen	OE-Nummer	Funktionsklasse	Ergebnis RT-PCR (fermentativ/aerob)	Ergebnis RT-PCR (phototroph/aerob)
11	bat	OE3101R	MIS	1,9	2,71
2	bop	OE3106F	EM	3,52	5,02
4	crtB1	OE3093R	LIP	0,36	1,13
235	crtB2	OE3376F	LIP	1,74	4,13
2290	crtI1	OE3381R	LIP	0,47	0,97
332	crtI2	OE3468R	LIP	3,52	14,03
2038	crtI3	OE1808F	LIP	0,13	0,76
500	crtY	OE3983R	LIP	0,54	2,52
2411	blh	OE3980R	CHY	1,16	4,71
NA	blp	OE3100F	MIS	5,98	26,91
1273	brp	OE3102R	MIS	3,93	4,56

Tab. 22: Ergebnisse der RT-PCR zur Bestimmung der Lichtinduzierbarkeit von Genen für den Wild-Typ Stamm **R1**. Beim Farbkode bedeutet rot = induziert unter fermentativen oder phototrophen Wachstumsbedingungen im Vergleich zum aeroben Wachstum, gelb = nicht reguliert und grün = reprimiert unter fermentativen oder phototrophen Wachstumsbedingungen im Vergleich zum aeroben Wachstum. Die Abkürzungen bedeuten: CHY = konserviert hypothetisches Protein, EM = Energie-Stoffwechsel, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes und NA = „not available“, nicht verfügbar.

## Ergebnisse und Diskussion

Rang FS-Statistik	Gen	OE-Nummer	Funktionsklasse	Ergebnis RT-PCR (fermentativ/aerob)	Ergebnis RT-PCR (phototroph/aerob)
11	bat	OE3101R	MIS	1,02	1,34
2	bop	OE3106F	EM	1,56	4,03
4	crtB1	OE3093R	LIP	0,55	1,68
235	crtB2	OE3376F	LIP	1,03	3,86
2290	crtI1	OE3381R	LIP	1,02	2,08
332	crtI2	OE3468R	LIP	2,12	7,06
2038	crtI3	OE1808F	LIP	4,68	32,98
500	crtY	OE3983R	LIP	4,48	12,41
2411	blh	OE3980R	CHY	3,36	8,84
NA	blp	OE3100F	MIS	1,93	10,22
1273	brp	OE3102R	MIS	7,35	10,13

Tab. 23: Ergebnisse der RT-PCR zur Bestimmung der Lichtinduzierbarkeit von Genen für den Wild-Typ Stamm **S9D**. Beim Farbkode bedeutet rot = induziert unter fermentativen oder phototrophen Wachstumbedingungen im Vergleich zum aeroben Wachstum, gelb = nicht reguliert und grün = reprimiert unter fermentativen oder phototrophen Wachstumbedingungen im Vergleich zum aeroben Wachstum. Die Abkürzungen bedeuten: CHY = konserviert hypothetisches Protein, EM = Energie-Stoffwechsel, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes und NA = „not available“, nicht verfügbar.

Rang FS-Statistik	Gen	OE-Nummer	Funktionsklasse	Ergebnis RT-PCR (fermentativ/aerob)	Ergebnis RT-PCR (phototroph/aerob)
11	bat	OE3101R	MIS	1,71	3,74
2	bop	OE3106F	EM	4,1	9,56
4	crtB1	OE3093R	LIP	0,3	1,07
235	crtB2	OE3376F	LIP	1,44	4,84
2290	crtI1	OE3381R	LIP	0,9	1,71
332	crtI2	OE3468R	LIP	1,79	16,91
2038	crtI3	OE1808F	LIP	0,63	11,53
500	crtY	OE3983R	LIP	0,77	4,12
2411	blh	OE3980R	CHY	3,28	7,71
NA	blp	OE3100F	MIS	2,53	20,07
1273	brp	OE3102R	MIS	13,8	26,05

Tab. 24: Ergebnisse der RT-PCR zur Bestimmung der Lichtinduzierbarkeit von Genen für die Mutante **S9H**. Beim Farbkode bedeutet rot = induziert unter fermentativen oder phototrophen Wachstumbedingungen im Vergleich zum aeroben Wachstum, gelb = nicht reguliert und grün = reprimiert unter fermentativen oder phototrophen Wachstumbedingungen im Vergleich zum aeroben Wachstum. Die Abkürzungen bedeuten: CHY = konserviert hypothetisches Protein, EM = Energie-Stoffwechsel, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes und NA = „not available“, nicht verfügbar.

Durch das erneute Kultivieren von aerob gewachsenen Zellen sollte das Problem der Vergleichbarkeit der RNA-Proben gelöst werden, jedoch zeigte sich, dass das Referenzgen *fdx* im Vergleich zu den verschiedenen Wachstumsbedingungen eine Regulation aufweist, weshalb die in den Tabellen 22 bis 24 zusammengestellten Ergebnisse ohne Bezug auf das Standardgen berechnet wurden.

Der Vergleich fermentativ mit aerob (f/a) ergibt für R1 und S9H identische Tendenzen, während S9D davon abweicht. So zeigen z.B. die Ergebnisse für die *bop*-Gen Expression eine Korrelation mit vorangegangenen phänotypischen Beobachtungen.

Der Stamm S9D exprimiert konstitutiv Bakteriorhodopsin und hiermit übereinstimmend zeigt sich nach der RT-PCR das *bop*-Gen im Vergleich f/a nicht reguliert und beim Vergleich phototroph und aerob (p/a) nur schwach induziert. Die Stämme *H. sal.* R1 und S9H weisen bei der Gegenüberstellung f/a eine erhöhte Bakteriorhodopsin-Expression auf, die unter phototrophen Wachstumsbedingungen noch gesteigert wird. Die Regulationsraten von *H. sal.* R1 und S9H entsprechen sich, was nochmals bestätigt, dass die S9H Mutante eine analoge BR-Expressionsregulation besitzt. Es könnte sich auch um eine Revertante handeln.

In der Literatur wird beschrieben, dass eine Mutation im *bat* Gen für den Phänotyp des S9-Stammes verantwortlich ist. Das Multidomänen Protein BAT besitzt laut einer in silico Domänensuche eine PAS/PAC- und eine GAF-Domäne. Dadurch läßt sich vermuten, dass BAT für die Zelle als Redoxsensor dient.

Parallele Sequenzierungen der vorliegenden Stämme S9H, S9D und WT R1 zeigen für das *bat* Gen im speziellen und dem *bop* Gen Cluster im allgemeinen, keine Unterschiede in der DNA Sequenz. Somit stellt das *bat* Gen einen wertvollen Kandidaten für zukünftige Versuche da.

In diesem Zusammenhang wäre denkbar Deletionsmutanten herzustellen oder eine in vitro Translation durchzuführen. Auf jeden Fall würden die Ergebnisse hilfreiche Bausteine zur Aufklärung der genauen Funktion des *bat* Gens sein.

## **6. Validierung der Daten / Ergebnisse**

DNA-Mikroarrays bieten eine hervorragende Möglichkeit funktionelle Zusammenhänge auf mRNA Ebene zwischen einer Vielzahl von Genen zu

untersuchen. Allerdings müssen aufgrund verschiedener Faktoren, wie z.B. der sequenzabhängigen Hybridisierungseigenschaften und deren Unterschiede die den Hybridisierungsreaktionen zugrunde liegen, die Array Ergebnisse als semiquantitativ angesehen werden. Als eine optimale Ergänzung zu den signifikanten Array Ergebnissen bietet sich die sensitive Verifikation der Expressionsdaten mittels quantitativer RT-PCR an.

Die DNA-Mikroarray basierenden Expressionsverhältnisse können oftmals um einen Faktor von 2-10 von den realen Werten abweichen. Diese Tatsache führte dazu, im Rahmen der Arbeit stichprobenartig die quantitative RT-PCR Methode anzuwenden, um die Ergebnisse der DNA-Mikroarray-Experimente zu überprüfen und zu quantifizieren.

### 6.1 Quantifizierung mittels RT-PCR

Die hier angewandte quantitative RT-PCR Methode basiert auf zwei Schritten. Im ersten Arbeitsschritt wird in einer reversen Transkription ausgehend von Gesamt-RNA von *H. sal.* R1 cDNA hergestellt. Diese dient im zweiten Schritt als Matrize in einer quantitativen genspezifischen PCR-Reaktion, wobei die Quantifizierung des Produktes über eine fluoreszenzbasierte Endpunktsbestimmung erfolgt. Das Prinzip der Quantifizierung beruht auf der Messung der Amplikonmenge nach jedem Temperaturzyklus der PCR mit Hilfe des Fluoreszenzfarbstoffs SYBR Green, der an die entstehende doppelsträngige DNA bindet. Die damit verbundene Fluoreszenz wird photometrisch detektiert und mit steigender Zyklenzahl graphisch und tabellarisch ausgegeben. Der aus diesen Kurven errechnete  $c_t$ -Wert, dies ist der Schnittpunkt der Kurve der signifikanten Fluoreszenzzunahme des Amplikons mit einem Schwellenwert, ermöglicht die Berechnung der Ausgangsmenge an mRNA-Templat. Dazu ist der Schwellenwert so zu wählen, dass er im linearen Bereich der logarithmischen Auftragung der Fluoreszenzzunahme des Amplikons liegt, siehe folgende Abbildung.

Zum Abschluss der Vorstellung der Ergebnisse wird in diesem Kapitel die Validierung der DNA-Mikroarray Auswertung beschrieben.

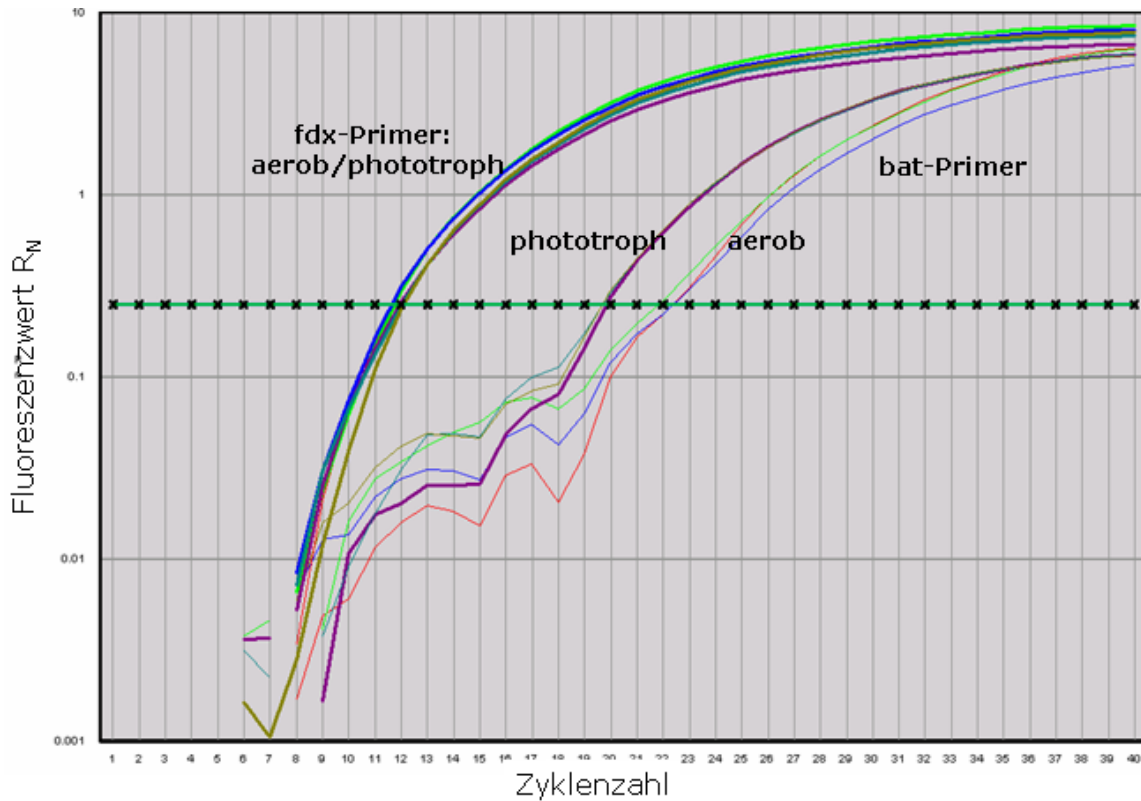


Abb. 81: Logarithmische Auftragung der Fluoreszenzwerte  $R_n$  gegen die Zyklenzahl mit  $C_t$  0,25 eines Ergebnisses eines RT-PCR-Experiments für die Gene *fdx* und *bat* zum Vergleich aerober und phototropher RNA von *H. sal.* R1.

Für jedes Gen werden drei Wiederholungen der PCR-Reaktion durchgeführt und der Mittelwert dieser Bestimmungen gebildet. Von diesen Mittelwerten werden die des nicht regulierten Referenzgens („house-keeping gene“) subtrahiert. Als Referenzgen für dieses Projekt wurde das Ferredoxin-Gen OE4217R (*fdx*, *fer1*) gewählt. Das Transkript des *fer1* Genes dient als Referenzgen, da an Hand von Mikroarray-Experimenten und durch RT-PCR-Analysen von *H. sal.* R1 gezeigt werden konnte, dass das Expressionsniveau unter aerober und phototropher Bedingungen für *fer1* in der midlog-Wachstumsphase gleich ist. Dies konnte unabhängig durch zwei verschiedene *fer1* spezifische Primerpaare bestätigt werden. Die Zellen wurden entsprechend des Protokolls der DNA-Mikroarrayexperimente kultiviert.

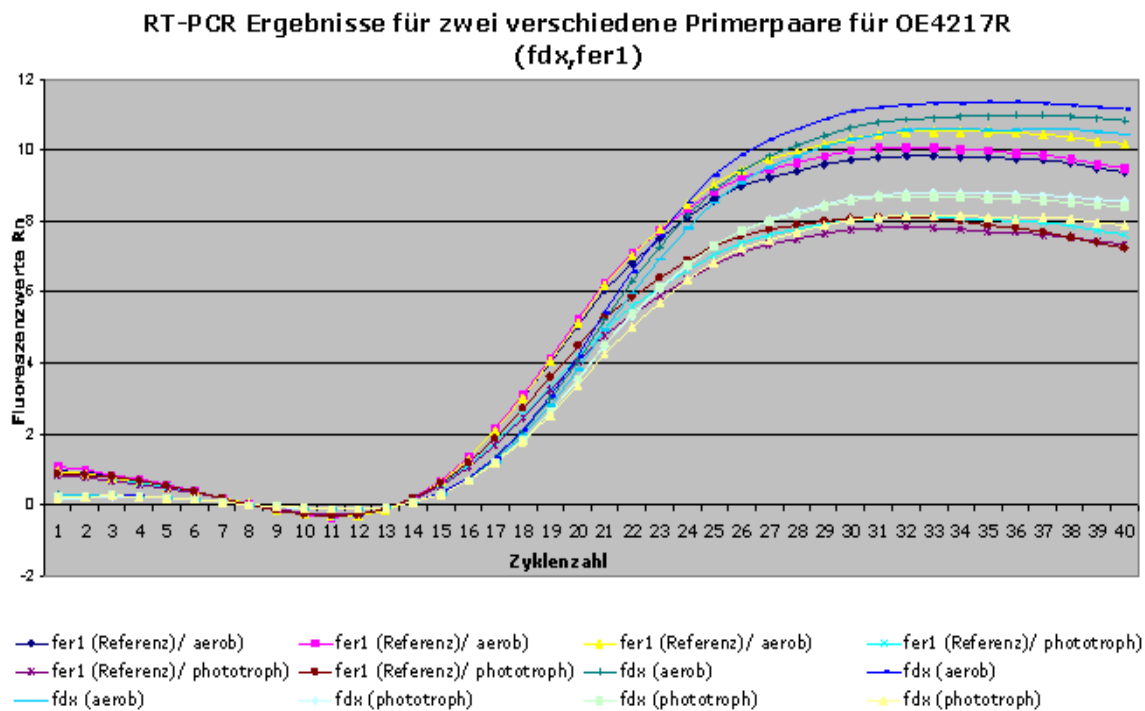


Abb. 82: Lineare Auftragung der Fluoreszenzwerte Rn eines RT-PCR-Experiments zum Vergleich aerob/phototroph für zwei unterschiedliche Primerpaare für das Referenzgen fdx (OE4217R).

Zum Vergleich zweier unterschiedlicher cDNAs z.B. von verschiedenen Wachstumsbedingungen, wird die Differenz der Mittelwerte der beiden Bedingungen gebildet. Dieser Differenzwert repräsentiert die unterschiedlichen Mengenverhältnisse der RNA-Templatmenge der zu vergleichenden Wachstumsbedingungen in Bezug auf das Referenzgen. Die Berechnung der Expressionsunterschiede begründet sich auf der Annahme, dass sich die Amplikonmenge bei jedem PCR-Zyklus verdoppelt.

Die Möglichkeit, einzelne Transkriptspezies mittels der RT-PCR Methode selektiv zu analysieren basiert auf dem Einsatz spezifischer Primerpaare. Für dieses Projekt wurde die Primerpaaroptimierung für jedes Gen mit Hilfe des Programms „Primer Express 2.0“ durchgeführt. Die Auswahl der Primer erfolgte zum einen aufgrund der Länge des Amplikons, die ca. 100 bp betragen sollte und der Lage des zu erwartenden PCR-Produktes, welche nicht zu nahe an dem 3´- bzw. 5´-Ende der Gensequenz sein sollte. Bevor die Primer für die RT-PCR eingesetzt wurden, sind sie mittels des „Oligonucleotide Properties Calculator“ ([www.basic.nwu.edu/biotools/oligocalc.html](http://www.basic.nwu.edu/biotools/oligocalc.html)) auf Sequenzinteraktionen und zudem anhand einer in-silico-PCR ([www.in-silico.com/s\\_pcr.html](http://www.in-silico.com/s_pcr.html)) auf ihre Spezifität hin überprüft worden. Die Primerauswahl erfolgte nur wenn beide

Kriterien erfüllt wurden. Analog zur Sondenauswahl für den DNA-Mikroarray sollen keine Kreuzreaktionen stattfinden.

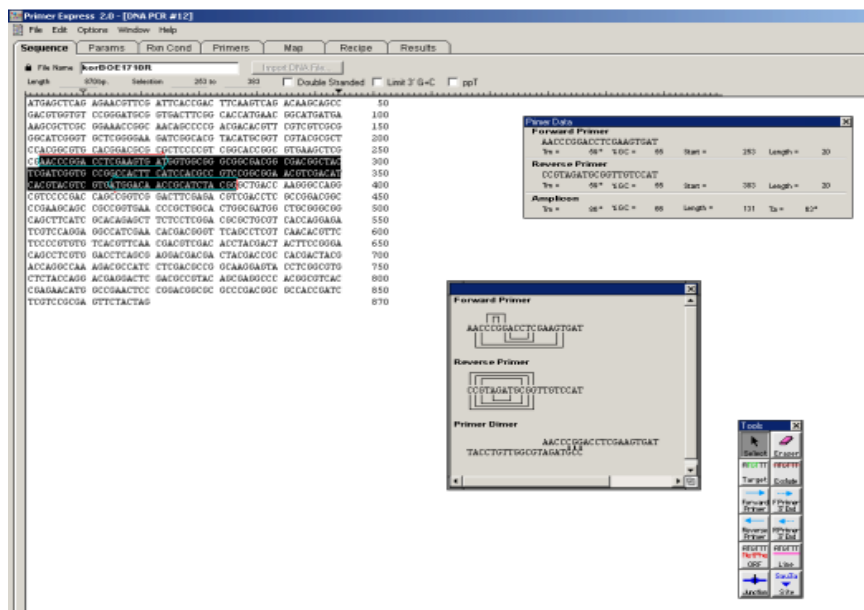


Abb. 83: Screenshot der Software Primer Express 2.0 zur Optimierung der Primerpaare zur RT-PCR Analyse.

Die Auswahl der Gene, die zur Verifizierung des ersten Transkriptom-Experiments eingesetzt wurden, erfolgte stichprobenartig basierend auf der Liste regulierter Gene nach der statistischen Auswertung. Die Gene sind dort nach ihrer statistischen Signifikanz analog ihres Ranges nach der Anwendung der FS-Statistik zur Lösung des ANOVA-Modells angeordnet. Bei der Zusammenstellung der zu kontrollierenden Gene wurde darauf geachtet, dass sie aus verschiedenen Signifikanzbereichen der Liste ausgewählt wurden. Neben der statistischen Signifikanz, war die Zugehörigkeit der Gene zu einem bestimmten Stoffwechselweg, wie z.B. der Carotinoidbiosynthese, der Argininfermentation oder der Beteiligung an der Bakteriorhodopsin-Biosynthese, ein weiteres Auswahlkriterium zum Einsatz in der Expressionsanalyse mittels RT-PCR.

Zur Validierung des ersten Mikroarray-Experiments wurden für die angewandte quantitative RT-PCR RNA-Proben eingesetzt, die auch für das erste Transkriptom-Experiment verwandt wurden. Von den zur Verfügung stehenden fünf RNA-Proben der Wiederholungen je Wachstumsbedingung, wurden drei eingesetzt. Die Ergebnisse der drei Einzelbestimmungen wurden, analog zum Mikroarray-Experiment, gemittelt. In der folgenden Tabelle sind die Ergebnisse des Mikroarray-Experiments und der RT-PCR-Analyse zusammengefasst und bewertet worden.



Als nächstes werden die Ergebnisse der RT-PCR zur Bestätigung der Ergebnisse der Transkriptom-Experimente vorgestellt.

Rang FS-Statistik	Gen	OE-Nummer	Funktionsklasse	Ergebnis Mikroarray	Ergebnis RT-PCR	Verifikation
4	crtB1	OE3093R	LIP	1,55	2,25	OK
10	OE3107F	OE3107F	HY	1,74	27,92	OK
12	carA	OE3556R	AA	1,64	2,23	OK
14	carB	OE3554F	AA	1,25	4,7	OK
31	arcC	OE5206R	AA	0,83	0,17	OK
60	korB	OE1710R	CIM	1,27	3,17	OK
87	arcB	OE5205R	AA	0,88	0,47	OK
140	dmsB	OE2225F	EM	1,31	1,21	OK
140	dmsB	OE2225F	EM	1,33	1,81	OK
199	cydA1	OE7065F	EM	0,78	0,67	OK
223	OE3136F	OE3136F	CHY	0,75	0,07	OK
332	crtI2	OE3468R	LIP	1,17	0,98	Halb OK
388	sdhC	OE2868R	CIM	0,79	0,88	OK
500	crtY	OE3983R	LIP	0,86	1,36	-
1030	arcA	OE5208R	AA	1,09	0,25	Halb OK
1273	brp	OE3102R	MIS	1,02	1,92	Halb OK
2290	crtI1	OE3381R	LIP	1,07	0,21	Halb OK
2411	blh	OE3980R	CHY	1,00	1,84	-
NA	blp	OE3100F	MIS	NA	1,19	-

Tab. 25: Ergebnisse der Validierung der Transkriptom-Analyse mittels RT-PCR für ausgewählte Gene. Beim Farbkode bedeutet rot = induziert im phototrophen Wachstum im Vergleich zum aeroben Wachstum, gelb = nicht reguliert und grün = reprimiert im phototrophen Wachstum im Vergleich zum aeroben Wachstum. Die Abkürzungen bedeuten: AA = Aminosäurestoffwechsel, CHY = konserviert hypothetisches Protein, CIM = zentraler Intermediatstoffwechsel, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes und NA = „not available“, nicht verfügbar.

Wie in der Tabelle 25 zu sehen ist, werden auch hier die in der Literatur beschriebenen Abweichungen der Mikroarray-Ergebnisse von denen einer RT-PCR beobachtet. Diese Unterschiede beziehen sich allerding auf die quantitative Aussage der Mikroarrays und nicht auf ihre qualitative. So werden die Ergebnisse

der Transkriptionsänderung bei 12 Genen von den 17 ausgewählten Genen durch die RT-PCR-Analyse bestätigt. Daher kann man zusammenfassend feststellen, dass die Ergebnisse der Transkriptionsanalyse mittels DNA-Mikroarrays mit denen der quantitativen RT-PCR im wesentlichen (etwa 71 %) übereinstimmen. Dies bedeutet, dass die Signifikanz der Ergebnisse der hier gezeigten Mikroarray-Analyse durch eine unabhängige Methode verifiziert wurden. Die Gene *blp* und *blh* wurden aus der Statistik herausgenommen, da für *blp* kein Mikroarray Ergebnis vorliegt und *blh* mit einem statistischen Rang von 2411 kein signifikantes Ergebnis liefert.

In einem zweiten großen Block von RT-PCR Experimenten wurde die Verifizierung der maximal möglichen Unterschiede zwischen aerob und phototroph vorgenommen.

Aus der Tabelle 25 ist ebenfalls zu entnehmen, dass die Expressionsverhältnisse basierend auf den Mikroarray- als auch auf den Ergebnissen der RT-PCR in ihren absoluten Werten nicht sehr hoch sind.

Aus diesem Grund wurde zur Bestimmung des maximal möglichen Expressionsunterschieds ein weiterer RT-PCR Versuch durchgeführt. Dazu wurden Kulturen aerob und phototroph inkubiert, mit dem Unterschied zum Transkriptom-Experiment, dass bei diesem Versuch die aerobe Kultur bei einer höheren Geschwindigkeit (250 rpm im Gegensatz zu 100 rpm) inkubiert wurde, um eine gesteigerte Sauerstoffversorgung zu gewährleisten. Eine weitere Änderung zum Standardprotokoll der Transkriptom-Analyse besteht darin, dass zur RNA-Isolierung einmal Zellen aus der logarithmischen (mid-log) und zum anderen aus der stationären (stat) Wachstumsphase eingesetzt worden sind. Durch dieses Experiment soll untersucht werden, in wieweit sich Änderungen in der Kultivierung des biologischen Ausgangsmaterials auf die Vergleichbarkeit der Transkriptmenge niederschlagen.

In der folgenden Tabelle sind, für 23 ausgewählte Gene, die Ergebnisse der RT-PCR Versuche mit denen der DNA-Mikroarrays zusammengefasst worden.

Rang FS-Statistik	Gen	OE-Nummer	Funktionsklasse	Ergebnis RT-PCR (mid-log Zellen)	Ergebnis RT-PCR* (stat. Zellen)	Ergebnis Mikroarray
2	bop	OE3106F	EM	6,5	3,45	2,09
4	crtB1	OE3093R	LIP	9,97	10,96	1,55
10	OE3107F	OE3107F	HY	19,74	6,56	1,6
11	bat	OE3101R	MIS	7,32	1,2	1,39
12	carA	OE3556R	AA	0,53	26,17	1,55
14	carB	OE3554F	LIP	1,85	6,22	1,25
31	arcC	OE5206R	AA	0,48	5,22	0,81
135	hly	OE4612F	CP	1,08	9,4	0,73
140	dmsB	OE2225F	EM	22,07	6,54	1,31
199	cydA1	OE7065F	EM	2,07	2,73	0,81
214	nuoCD	OE1956F	EM	0,98	3,27	1,27
223	OE3136F	OE3136F	CHY	4,53	5,72	0,83
235	crtB2	OE3376F	LIP	1,96	8,36	1,14
332	crtI2	OE3468R	LIP	1,64	1,88	1,17
500	crtY	OE3983R	LIP	2,14	0,91	0,86
560	arcB	OE5205R	AA	1,92	1,98	0,89
1030	arcA	OE5208R	AA	1,84	1,39	1,09
1684	fer4	OE4496R	EM	1,64	1,76	1,07
1850	arcR	OE5209R	REG	0,55	0,64	0,98
2038	crtI3	OE1808F	LIP	1,65	3,41	1,02
2290	crtI1	OE3381R	LIP	1,27	1,24	0,99
2398	-	OE3073R	MIS	1,94	1,62	1,01
NA	fer1	OE4217R	EM	1,01	4,0	NA

Tab. 26: Ergebnisse der quantitativen RT-PCR zur Bestimmung des maximale Unterschieds zwischen aerobem und phototrophem Wachstum. Beim Farbkode bedeutet rot = induziert im phototrophen Wachstum im Vergleich zum aeroben Wachstum, gelb = nicht reguliert und grün = reprimiert im phototrophen Wachstum im Vergleich zum aeroben Wachstum. (\*Ergebnis der RT-PCR nicht auf Referenzgen bezogen). Die Abkürzungen bedeuten: AA = Aminosäurestoffwechsel, CHY = konserviert hypothetisches Protein, CP = zelluläre Prozesse, EM = Energie-Stoffwechsel, HY = hypothetische Proteine, LIP = Lipid-Stoffwechsel, MIS = Verschiedenes, REG = Genregulation und NA = „not available“, nicht verfügbar.

Die hier angeführten Resultate der RT-PCR sind Mittelwerte aus mehreren Einzelbestimmungen. Zu beachten ist bei der Interpretation der RT-PCR-Ergebnisse, dass die Kurven der Zunahme der Amplikonmenge im Verlauf der PCR abhängig vom spezifischen Primerpaar und der Absolutkonzentration des jeweiligen Transkripts (Templats) sehr unterschiedliche Verläufe aufweisen und die Qualität der Ergebnisse beeinflussen können. So steigt die Konzentration des

Amplikons des Gens OE3107F erst nach 26 bis 29 PCR-Zyklen an und weist keinen exponentionellen Verlauf des Graphen auf. Die Interpretation solcher Ergebnisse muß daher sehr vorsichtig erfolgen.

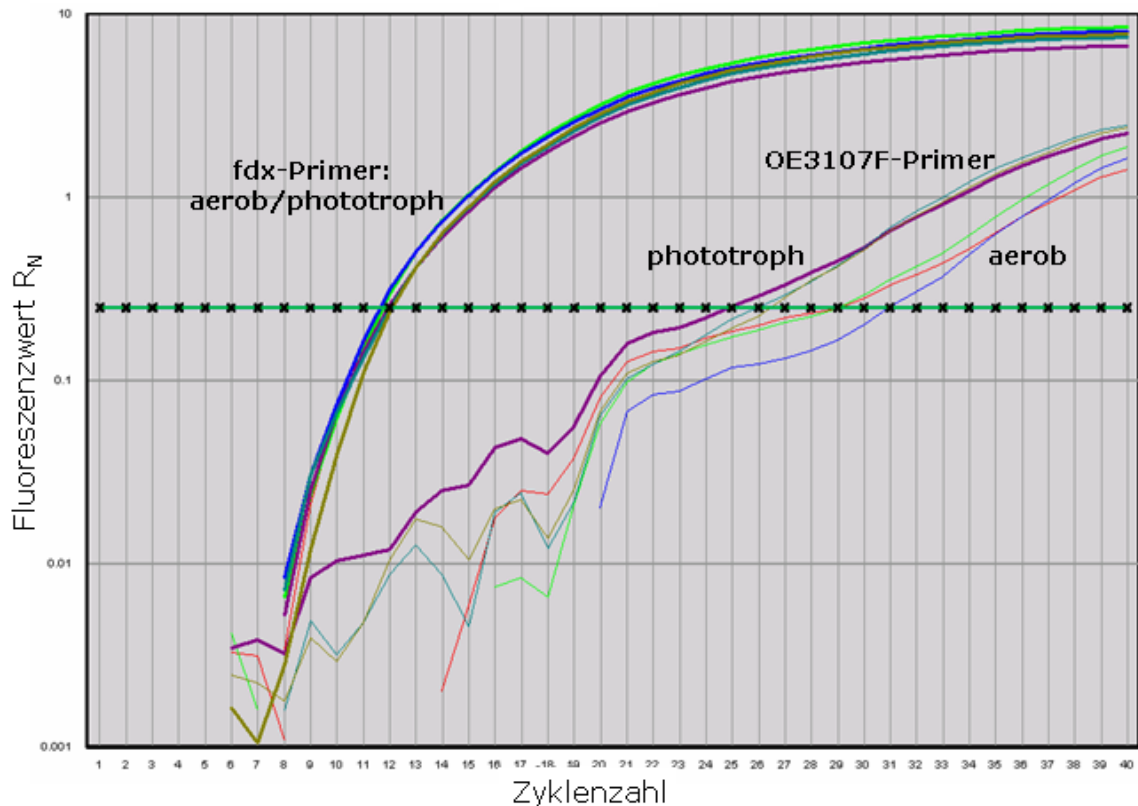


Abb. 84: Als Beispiel eine Darstellung einer logarithmischen Auftragung der Fluoreszenz  $R_n$  gegen die Zyklenzahl mit  $C_t$  0,25 einer RT-PCR für die Gene *fdx* und OE3107F.

Von den untersuchten 23 Genen zeigen 17 Gene für Zellen aus der logarithmischen Phase unter phototrophen Bedingungen eine induzierte Expression beim Vergleich mit aerob gewachsenen Zellen. Darunter befinden sind vor allem die untersuchten Gene der Carotinoidbiosynthese (*crtB1*, *crtB2*, *crtI1*, *crtI2*, *crtI3*, *crtY*), so wie das *bop* und das *bat* Gen. Eine ebenfalls erhöhte Expression beim phototrophen Wachstum weisen die Gene OE3107F und OE3136F auf, deren exakte Funktion in der Adaption an phototrophes Wachstum derzeit noch nicht geklärt ist. Eindeutig nicht reguliert zeigen sich die Gene *hly*, *nuoCD* und *fer1*. Diese Tatsache unterstreicht die Wahl des *fer1* Gens als Referenz zur Auswertung der RT-PCR (siehe Abb. 82). Ein reprimiertes Expressionsniveau zeigen die Gene *arcC* und *arcR*, die an der Argininfermentation beteiligt sind.

Der Vergleich von phototrophen und aeroben Zellen aus der stationären Wachstumsphase zeigt für 21 der 23 analysierten Gene eine Induktion der

Expression. In der stationären Wachstumsphase sind physiologische Anpassungen erforderlich, die sich auf der Ebene der Transkripte widerspiegeln sollte. Diese Unterschiede zwischen logarithmischer und stationärer Wachstumsphase sind hier zu sehen.

Entscheidend für jede weitere Betrachtung ist an dieser Stelle, dass in der stat-Wachstumsphase das Referenzgen *fer1* eine 4-fache Induktion aufweist, was in drei unabhängigen Experimenten mit zwei unterschiedlichen Primerpaaren bestätigt wurde. Hierdurch entfällt die Möglichkeit *fer1* als Standardgen zu verwenden. Die in Tabelle 26 enthaltenen Resultate für die RT-PCR mit RNA-Proben von stationären Zellen, wurden nicht auf das Referenzgen bezogen. Die Kurvenverläufe der logarithmischen Auftragung der Amplikonmengen gegen die Zyklenzahl, sind in der folgenden Abbildung zu sehen.

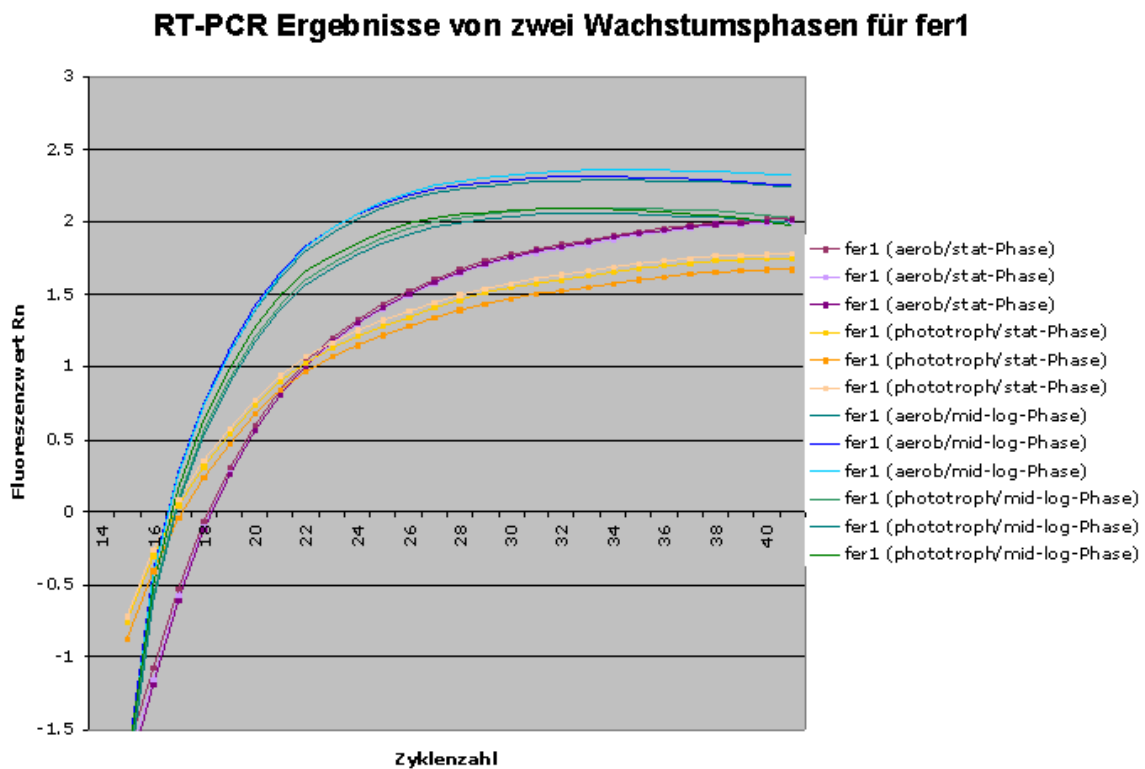


Abb. 85: Logarithmische Auftragung der Fluoreszenzwerte Rn gegen die Zyklenzahl für das Referenzgen *fer1*. Dargestellt sind die Kurvenverläufe von *fer1* für den nicht regulierten Fall (blaue und grüne Graphen) und den regulierten Fall (violette und orangene Graphen).

Die Abb. 85 verdeutlicht, dass in der logarithmischen Wachstumsphase die Kurven des *fer1* Gens unter aeroben und phototrophen Bedingungen übereinander liegen und damit sein Einsatz als Referenzgen gerechtfertigt ist, da es unter diesen Bedingungen als nicht reguliert angesehen werden kann. Im

Gegensatz dazu verlaufen die Graphen für das *fer1* Gen der stationären Phase deutlich unterschiedlich. Das *fer1* Gen ist bei phototrophen Wachstum stärker exprimiert und der Einsatz als Referenz ist nicht mehr möglich.

Jedoch zeigen die Resultate eindeutig, dass es bei einer Expressionsanalyse auf genau definierte Bedingungen bei der Kultivierung des biologischen Ausgangsmaterials ankommt, um eine Vergleichbarkeit der Transkriptmengen zu gewährleisten. Vor allem für die quantitative Beurteilung der Regulation bedarf es der exakten Standardisierung.

Trotz der unterschiedlichen biologischen und experimentellen Voraussetzungen zeigt der Vergleich der RT-PCR Ergebnisse der RNA-Proben der mid-log Phase mit den Mikroarray-Experimenten eine Übereinstimmung in 15 Genen, was einer Quote von 68% entspricht. Dies bestätigt das Ergebnis der direkten Verifizierung des ersten Transkriptom-Experiments. Bei dem Anteil von 15 Genen sind auch 6 Gene enthalten, die bei den Mikroarray-Ergebnissen keine Regulation zeigen, also deren Werte zwischen 0,9 und 1,1 liegen.

Des Weiteren konnte durch dieses Experiment eindeutig bestätigt werden, dass der maximale Unterschied zwischen aeroben und phototrophen Bedingungen im RT-PCR-Experiment und im Mikroarray-Experiment auf dem gleichen Niveau liegen, was bedeutet, dass die Induktion z.B. des *bop* Gens für beide Methoden im selben Bereich liegt und die Regulation der Proteinexpression nicht nur auf der Transkriptionsebene abläuft. Man kann also abschließend sagen, dass durch die höhere Sauerstoffdurchlüftung zwar Unterschiede auf Transkriptebene bestehen, es aber nicht zu einer wesentlich stärkeren Induktion der untersuchten Gene kommt. Zum Vergleich der Ergebnisse dieser Arbeit mit früheren Experimenten ist es wichtig, genau über die experimentellen Wachstumsbedingungen und die Wachstumsphase der Zellen Bescheid zu wissen, um ausschließen zu können, dass eventuelle Abweichungen auf Unterschiede im Zeitpunkt der Probennahme zurückzuführen sind.

## **6.2 Interpretation auf Proteinebene**

Neben der Transkriptomik stellt die Proteomik eine weitere Methode zur Gesamtgenomanalyse dar. Proteomik beschreibt das Proteininventar einer Zelle zu einem bestimmten Zeitpunkt und unter einem gegebenen Umwelteinfluss.

Eine Kombination der beiden Analyse Methoden kann einen großen Schritt zum Verständnis der Zusammenhänge der Regulation auf der RNA- und der Proteinebene bedeuten.

In den folgenden Abschnitten wird die Kombination der Ergebnisse der Transkriptom-Experimente und der Proteomanalyse von *H. sal.* R1 vorgestellt. Die Ergebnisse der Proteomik beziehen sich auf die Arbeiten von Frau Birgit Bisle und Herrn Andreas Tebbe, die eine Analyse des Membran-Proteoms und der cytosolischen Proteine von *H. sal.* R1 vornehmen. Ziel auch dieser Arbeiten ist die Identifikation der Unterschiede zwischen der aeroben und der phototrophen Wachstumsbedingung.

In der Tab. 27 sind die Ergebnisse der Analyse der cytosolischen Proteine zusammengestellt, die auf den Experimenten von Herrn Andreas Tebbe basieren (Dr. Arbeit A. Tebbe).

## Ergebnisse und Diskussion

OE-Nummer	Funktionsklasse	Genname	Protein-Identifizierung	Übereinstimmung	Proteinname
OE1613R	MIS	-	Trusted	ZR II	probable acylaminoacyl-peptidase (EC 3.4.19.1)
OE1626F	NOF	-	Trusted		conserved protein
OE1633F	CHY	-	-		conserved hypothetical protein
OE1932R	NOF	-	Trusted		conserved protein
OE2057F	COM	thiC	Trusted	ZV+ZR I	thiamin biosynthesis protein thiC
OE2671R	CHY	hemK	-		HemK protein homolog (protoporphyrinogen oxidase homolog)
OE2703F	MIS	pan1	Trusted	ZR I	probable copper-containing oxidoreductase
OE3036F	AA	glyA	Trusted	ZV+ZR I	glycine hydroxymethyltransferase (EC 2.1.2.1)
OE3045F	NOF	-	Trusted	ZV	conserved protein
OE3206R	CHY	-	-		conserved hypothetical protein
OE3207F	COM	cbiT	Trusted		probable precorrin-8W decarboxylase (EC 1.-.-.-)
OE3209F	COM	cbiL	Trusted		probable precorrin-2 C20-methyltransferase (EC 2.1.1.130)
OE3212F	COM	cbiF	Insecure	ZR I	precorrin-4 C11-methyltransferase (EC 2.1.1.133)
OE3213F	MIS	cbiG	-	ZR I	CbiG protein (unassigned function)
OE3214F	COM	cbiH1	Insecure		precorrin-3B C17-methyltransferase (EC 2.1.1.131) 1
OE3216F	COM	cbiH2	Insecure	ZR I	precorrin-3B C17-methyltransferase (EC 2.1.1.131) 2
OE3218F	MIS	-	Insecure		cobalamin operon protein
OE3221F	MIS	-	-		conserved cobalamin operon protein
OE3224F	MIS	-	Trusted		conserved cobalamin operon protein
OE3237F	COM	cbiC	Trusted		precorrin-8X methylmutase (EC 5.4.1.2)
OE3314R	HY	-	-		hypothetical protein
OE3315R	TP	cbiO1	-		ABC-type cobalt transport system ATP-binding protein
OE3319R	TP	cbiM	-		CbiM protein (unassigned function) (probable ABC-type cobalt transport system permease protein 2)
OE3710R	NOF	trkA6	Trusted	ZR II	trkA domain protein
OE4230F	MIS	yuxL	Trusted		probable acylaminoacyl-peptidase (EC 3.4.19.1)
OE4311F	TP	appA	Trusted		ABC-type transport system periplasmic substrate-binding protein
OE4345R	NUM	nrdA	-	ZR I	probable ribonucleoside-diphosphate reductase (EC 1.17.4.1) small chain
OE4346R	NUM	nrdB1	Insecure		ribonucleoside-diphosphate reductase (EC 1.17.4.1) large chain
OE4398F	MIS	hal	Trusted	ZV	O-acetylhomoserine (thiol)-lyase (EC 4.2.99.10)
OE4551F	TP	dppD	Trusted	ZR I	ABC-type transport system periplasmic substrate-binding protein
OE4555F	TP	dppC1	-		ABC-type transport system permease protein
OE4651F	COM	thi1	Trusted	ZV	thiamine biosynthesis protein (thiazole biosynthetic enzyme)
OE4654F	COM	thiD	Trusted	ZR I	phosphomethylpyrimidine kinase (EC 2.7.4.7)
OE4655R	NOF	-	Trusted		conserved protein
OE4673F	MIS	cxp	Trusted	ZR II	carboxypeptidase (EC 3.4.-.-)
OE4674F	CIM	adh2	Trusted		alcohol dehydrogenase (EC 1.1.1.1)
OE5032R	HY	-	Insecure		hypothetical protein
OE5245F	TP	-	-		ABC-Type transport system ATP-binding protein
OE5248F	NOF	-	Trusted	ZV	conserved protein
OE5326R	HY	-	-		hypothetical protein
OE1651F	HY	-	-		hypothetical protein
OE1874R	EM	petB	Trusted		probable menaquinol--cytochrome-c reductase (EC 1.10.2.-) (cytochrome bc complex) cytochrome b subunit
OE1876R	EM	petA	Trusted		probable menaquinol--cytochrome-c reductase (EC 1.10.2.-) (cytochrome bc complex) Rieske iron-sulfur protein
OE2991F	NOF	-	Trusted	ZR I	transcription factor TFB homolog
OE3101R	MIS	bat	Questionable	ZV+ZR I	bacterioopsin activator
OE3468R	LIP	ort2	Trusted	ZR I+ZR II	phytoene dehydrogenase (EC 1.14.99.-) (phytoene desaturase) 2 (alternative EC 1.3.-.-)
OE3554F	AA	carB	Trusted	ZV+ZR I	carbamoyl-phosphate synthase (EC 6.3.-.-) large chain ((ammonia) (EC 6.3.4.16) or (glutamine-hydrolyzing) (EC 6.3.4.17))
OE3556R	AA	carA	Insecure	ZV	carbamoyl-phosphate synthase (EC 6.3.-.-) small chain ((ammonia) (EC 6.3.4.16) or (glutamine-hydrolyzing) (EC 6.3.4.17))
OE3815R	CHY	-	-	ZR II	conserved hypothetical protein
OE5083R	NOF	hsp5	Believed	ZR I+ZR II	heat shock protein homolog
OE6046F	CHY	-	-		conserved hypothetical protein

Tab. 27: Liste der regulierten Gene der Proteomanalyse der cytosolischen Proteine vom Vergleich aerobes und phototrophes Wachstum kombiniert mit den Ergebnissen der Transkriptom-Experimente von *H. sal.* R1. Gene der oberen Hälfte der Tabelle sind unter phototrophen Bedingungen reprimiert, die der unteren Hälfte induziert. (Quelle Promotion Andreas Tebbe) ZV = Zweigruppenvergleich; ZRI = erstes Zeitreihen Experiment; ZRII = zweites Zeitreihenexperiment. Der Farbkode ist in Tab. 26 aufgeschlüsselt.

Mit welchem Transkriptom-Experiment die Proteom-Daten übereinstimmen ist durch die Abkürzungen in der fünften Spalte von rechts angegeben, dabei beziehen sich die unterschiedlichen Färbungen der Spalten darauf mit welchem Experiment eine Übereinstimmung besteht. Die Aufschlüsselung des Farbkodes ist in Tab. 28 gegeben.

5		Gen ist in Proteom-Liste und Transkriptom-Liste Zweigruppenvergleich
8		Gen ist in Proteom-Liste und Transkriptom-Liste von Zeitreihe I
4		Gen ist in Proteom-Liste und Transkriptom-Liste von Zeitreihe II
4		Gen ist in Proteom-Liste und Transkriptom-Liste vom Zweigruppenvergleich und Zeitreihe I
2		Gen ist in Proteom-Liste und Transkriptom-Liste von Zeitreihe I und II
23		

Tab. 28: Aufschlüsselung des Farbkodes von Tabelle 27. Die Zahlen links geben die Anzahl der übereinstimmenden Gene pro Transkriptom-Experiment an.

Aus Tab. 28 ist zu entnehmen, dass 23 Gene aus den Transkriptom-Experimenten mit den Ergebnissen der Proteomanalyse übereinstimmen, was



einem Anteil von 56% entspricht, wenn man eine Gesamtzahl von 41 regulierten Proteine nach der Proteomanalyse zugrunde legt.

Man findet drei Blöcke, die Übereinstimmungen der beiden Methoden aufweisen. So werden in einem Block die Gene *thi1*, *thiC* und *thiD*, die an der Thiaminbiosynthese beteiligt sind, als übereinstimmend gefunden. Einen weiteren Block von übereinstimmenden Genen bilden die Gene *cbiF*, *cbiG* und *cbiH2* die an der Biosynthese des Percorrins beteiligt sind. Und als dritter Block kann die Übereinstimmung für die Gene *bat* und *crtI2* sowie der beiden Untereinheiten der Carbamoylsynthase *carA* und *carB* angesehen werden.

In die Liste der regulierten Gene der Proteomanalyse sind Gene aus der Transkriptomanalyse eingefügt worden, wenn sie in direkter Nachbarschaft mit einem Gen auf der Proteinliste stehen. Solche Einschübe von Genen in diese Listen könnte helfen die Interpretation der Ergebnisse zu erleichtern.

In der Liste der Membran-Proteomanalyse (Tab. 29) ist eine geringere Rate der Übereinstimmungen mit den Transkriptomdaten zu verzeichnen. Dies liegt zum einen daran, dass diese Liste mehr regulierte Gene im Vergleich zu der Liste der cytosolischen Proteine enthält. Die Aufschlüsselung des Farbkodes und der Anzahl der Gene pro Transkriptom-Experiment ist in Tab. 30 enthalten.

## Ergebnisse und Diskussion

OE-Nummer	Funktionsklasse	Genname	Protein-Identifizierung	Übereinstimmung	Proteinname
OE1202F	TP	trp1	Trusted	ZV + ZRI	probable ABC-type transport system ATP-binding/permease protein
OE1442R	CHY	-	Trusted	ZRI	conserved hypothetical protein
OE1675R	TP	pstB2	Trusted	-	probable ABC-type phosphate transport system ATP-binding protein
OE1676R	TP	pstA2	Trusted	ZV	probable ABC-type phosphate transport system permease protein
OE1679R	TP	phoX2	Trusted	-	probable ABC-type phosphate transport system periplasmic phosphate-binding protein
OE1806R	CE	-	Trusted	-	probable periplasmic protein
OE1932R	NOF	-	Trusted	-	conserved protein
OE1979R	EM	cox1A	Trusted	-	cytochrome-c oxidase (EC 1.9.3.1) chain I
OE1988R	EM	cox2A	-	ZRII	cytochrome-c oxidase (EC 1.9.3.1) chain II
OE2076F	NOF	-	Trusted	-	conserved protein
OE2964F	CHY	-	Trusted	-	signal-transducing histidine kinase homolog
OE3367F	CHY	-	-	-	dnaJ N-terminal domain protein
OE3597R	HY	-	-	-	hypothetical protein
OE3611R	SIG	basT	Trusted	-	transducer protein basT
OE3612R	SIG	basB	Trusted	-	chemotactic signal transduction system periplasmic substrate-binding protein basB
OE3641F	TP	thb	Insecure	-	ABC-type transport system periplasmic substrate-binding protein (thiamin-binding protein)
OE3642F	NOF	-	Trusted	-	conserved protein
OE3907R	TP	phnE	-	-	ABC-type transport system permease protein
OE3908R	TP	phnC	Insecure	-	ABC-type transport system ATP-binding protein
OE4070R	EM	cox1B	Trusted	-	cytochrome-c oxidase (EC 1.9.3.1) chain I
OE4197F	NOF	-	Trusted	-	conserved protein
OE4301R	TP	ykfD	Trusted	-	ABC-type transport system ATP-binding protein
OE4304R	TP	dppB1	Trusted	-	ABC-type transport system permease protein
OE4305R	TP	dppA	Trusted	-	ABC-type dipeptide transport system periplasmic dipeptide-binding protein
OE4311F	TP	appA	Trusted	-	ABC-type transport system periplasmic substrate-binding protein
OE4343R	NOF	-	Trusted	-	protein OE4343R
OE4345R	NUM	nrdA	-	-	probable ribonucleoside-diphosphate reductase (EC 1.17.4.1) small chain
OE4357F	TP	-	Trusted	-	ABC-type transport system periplasmic substrate-binding protein
OE4551F	TP	dppD	Trusted	ZRI	ABC-type transport system periplasmic substrate-binding protein
OE4555F	TP	dppC1	Trusted	-	ABC-type transport system permease protein
OE4576F	TP	-	Trusted	-	probable ABC-type transport system substrate-binding protein
OE4748F	NOF	-	Trusted	ZV	conserved protein
OE5032R	HY	-	Insecure	-	hypothetical protein
OE1211F	MIS	hsp4	Trusted	-	heat shock protein homolog
OE1294R	TL	rpl15R	Trusted	-	ribosomal protein L15.eR
OE1365F	CHY	-	Questionable	-	conserved hypothetical protein
OE1545F	LIP	rdh1	Trusted	-	probable oxidoreductase (EC 1.1.1.-) (cis-retinol dehydrogenase homolog)
OE1597F	NOF	-	Trusted	-	protein OE1597F
OE1724R	CHY	-	-	ZRII	conserved hypothetical protein
OE1847R	NOF	-	Trusted	-	conserved protein
OE2092F	HY	-	Insecure	-	hypothetical protein
OE2116R	MIS	-	Trusted	-	glutamate-rich protein
OE2120F	NOF	-	Trusted	-	conserved protein
OE2121F	NOF	-	Trusted	-	conserved protein
OE2159R	TL	rps3aR	Trusted	ZRI + ZRII	ribosomal protein S3a.eR
OE2160R	HY	-	-	-	hypothetical protein
OE2222F	HY	-	-	-	hypothetical protein
OE2223F	EM	dmsA	Questionable	-	probable dimethylsulfoxide reductase (EC 1.8.-.-) chain A (reductase subunit)
OE2225F	EM	dmsB	-	-	dimethylsulfoxide reductase (EC 1.8.-.-) chain B (electron transfer protein)
OE2316R	TP	rbsA	Trusted	-	ABC-type transport system ATP-binding protein
OE2317R	TP	-	Trusted	-	probable ABC-type transport system periplasmic substrate-binding protein (membrane lipoprot
OE2390R	MIS	flaD	Believed	-	flagella accessory protein
OE3106F	EM	bop	Trusted	ZV + ZRI + ZRII	bacteriorhodopsin precursor
OE3342R	HY	-	Trusted	-	hypothetical protein
OE3392F	TL	rpl2	Trusted	ZV + ZRII	ribosomal protein L2
OE3415F	TL	rps5	Trusted	ZRII	ribosomal protein S5
OE3533F	COM	hemY	Trusted	-	probable protoporphyrinogen oxidase (EC 1.3.3.4)
OE3762R	CIM	glpK	Insecure	-	glycerol kinase (EC 2.7.1.30)
OE3763F	CIM	gpdA2	Trusted	-	glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) chain A
OE3765F	CIM	gpdC	Questionable	-	glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) chain C
OE3947R	CHY	-	-	-	conserved hypothetical protein
OE4712F	NOF	-	Trusted	-	conserved protein
OE4735R	TL	rps7	Trusted	ZRI + ZRII	ribosomal protein S7
OE5187R	MIS	-	Trusted	ZV	probable hydrolase
OE5188F	CHY	-	Insecure	-	protein kinase weak homolog

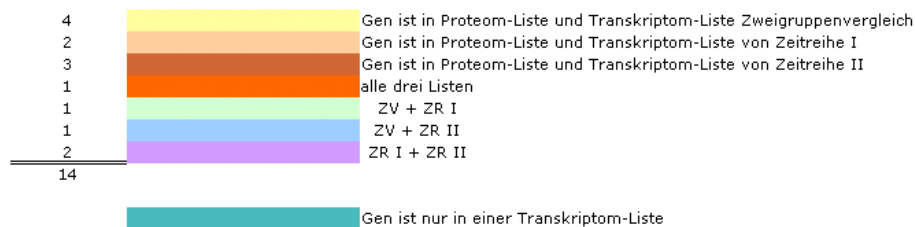
Tab. 29: Liste der regulierten Gene der Proteomanalyse der Membranproteine vom Vergleich aerobes und phototrophes Wachstum kombiniert mit den Ergebnissen der Transkriptom-Experimente von *H. sal.* R1. Gene der oberen Hälfte der Tabelle sind unter phototrophen Bedingungen reprimiert, die der unteren Hälfte induziert. (Quelle Promotion Birgit Bisle) ZV = Zweigruppenvergleich; ZRI = erstes Zeitreihen Experiment; ZRII = zweites Zeitreihenexperiment. Der Farbkode ist in Tab. 28 aufgeschlüsselt.

Auch in dieser Liste sind Gene eingefügt worden, die nicht in der Liste der regulierten Proteine enthalten sind.

Das *bop*-Gen wird in allen drei Transkriptom-Experimenten und der Proteomanalyse als reguliert gefunden.

Die offensichtlichsten Übereinstimmungen werden bei Proteinen von ABC-Transport-Systemen gefunden. Hier zeigt sich der positive Effekt des Einfügens von Genen aus der Transkriptomanalyse, denn so vervollständigen sich die Informationen über die Mitglieder dieser Transportsysteme. Als Beispiele wären da die Gene *pstB2* und *dppC1* zu nennen, die in der Transkriptomanalyse detektiert wurden und in direkter Nachbarschaft mit den Genen *pstA2* und *dppD* stehen, die in der Membranproteomanalyse gefunden wurden.

Die Rate der Übereinstimmungen, die der Tab. 30 zu entnehmen ist, liegt bei 14 Genen, was einem Anteil von 25,6% bei einer Gesamtzahl von 54 regulierten Membranproteinen entspricht.



Tab. 30: Aufschlüsselung des Farbkodes von Tabelle 25. Die Zahlen links geben die Anzahl der übereinstimmenden Gene pro Transkriptom-Experiment an.

Neben der Angabe über die Übereinstimmung enthält die Tab. 30 auch die Aufschlüsselung des Farbkodes für die Tab. 29.

Als Fazit dieser Kombination von Transkriptom- und Proteomdaten kann geschlossen werden, dass die Raten der Übereinstimmungen von 25,6% bzw. 56% im Hinblick auf die methodischen Unterschiede sehr groß sind.

Dieses Ergebnis zeigt ein weiteres Mal, dass die etablierte Methode der Transkriptomik verlässliche Ergebnisse liefert, die durch andere Methoden reproduziert werden können. Im Hinblick der Aufklärung der Regulation des phototrophen Wachstums stellt die Kombination von mehreren Gesamtgenomanalyse Methoden einen unschätzbaren Wert dar.

## 7. Mikrobiologische Methoden

### 7.1 Zellzucht

#### 7.1.1 Propagierung und Aufbewahrung von *Escherichia coli*

*Escherichia coli* (*E. coli*) Zellen wurden in LB Medium inokuliert. Für „über Nacht“ (üN)-Kulturen wurden 35 ml Medium angeimpft und bei 37°C bei 250 rpm geschüttelt. Bakterienkolonien auf Agarplatten können einige Tage verwendet werden. Für eine längere Aufbewahrung wurden Aliquots von üN-Kulturen mit derselben Menge an Glycerin versetzt und bei -70°C gelagert.

Medium für *Escherichia coli*:

LB(Luria-Bertani) Medium

Bacto™ Trypton	10 g
Bacto™ Hefeextrakt	5 g
NaCl	10 g
ad 1 l H <sub>2</sub> O bidest.	
mit NaOH auf pH 7,0-7,2 einstellen, autoklavieren	

#### 7.1.2 Propagierung und Aufbewahrung von *Halobacterium salinarum*

Für die Anzucht von *Halobacterium salinarum* (*H. sal.*) Stamm R1 (DSM 671) wurden 35 ml Halostandardmedium entweder von Agarplatten oder mit 1 ml Flüssigkultur angeimpft und bei 40°C und 100 rpm geschüttelt. Das Überimpfen erfolgte im Verhältnis 1:20 in 1 l Hauptkultur. Diese 1 l Kultur wurde als Stammkultur angesehen und diente für alle weitere Kulturen als Inokulum.

Für das aerobe Wachstum wurden Seitenarmkolben (KlettKolben) benutzt, die mit einem Zellstoffstopfen verschlossen wurden und bei 40°C im Dunkeln geschüttelt wurden. Bei Untersuchung des phototrophen Wachstums von *H. sal.* R1 wurden KlettKolben mit Schliff benutzt, die mittels eines Silikonseptums luftdicht verschlossen wurden. Die Inkubation erfolgte hier in einem Schüttler bei 40°C vor einer Hochdruck-Xenonlampe.

Eine stationäre 5-7 Tagekultur kann bei Raumtemperatur luftdicht verschlossen gelagert werden. Außerdem können 50 µl einer 10<sup>6</sup>-10<sup>7</sup> fachen Verdünnung

ausplattiert werden und die nach 8- bis 14-tägiger Inkubation bei 40°C erhaltenen Einzelkolonien bei Raumtemperatur aufbewahrt werden.

Medium für *H. sal.* R1:

Halostandardmedium (Oesterhelt & Krippahl, 1983)

4,3 M NaCl	250 g
80 mM MgSO <sub>4</sub> • 7H <sub>2</sub> O	20 g
27 mM KCl	2 g
10 mM Na <sub>3</sub> Citrat• 2H <sub>2</sub> O	3 g
Bacto™ Pepton	10 g
ad 1 l H <sub>2</sub> O bidest.	
mit NaOH auf pH 7,0-7,2 einstellen , autoklavieren	
für Haloplatten + 15 g Bacto™ Agar	

## Wechsel vom aeroben auf phototrophes Wachstum

### 7.2 Aufnahme einer Wachstumskurve

Zur Kontrolle der Wachstumsgeschwindigkeit wurden 35 ml-Kulturen von *H. sal.* R1 in 100 ml Klettkolben kultiviert und in bestimmten Zeitabständen die Lichtstreuung in einem Klett-Photometer bestimmt. Dabei entsprach eine Klett-Einheit etwa  $10^7$  Zellen in 1 ml Medium.

### 7.3 Antibiotika

Die verwendeten Antibiotika werden wie folgt eingesetzt:

*E. coli*:

Ampicillin 100 µg/ml (100 mg/ml Stammlösung)

*H. sal.* R1:

Mevinolin 10,1 µg/ml (25 mM Stammlösung)

## 8. Molekularbiologische Methoden

### 8.1 Isolierung chromosomaler DNA aus *H. sal.* R1

Für die Präparation sauberer chromosomaler DNA wurden 2 ml einer in der logarithmischen Wachstumsphase befindlichen Halokultur für 2 min bei 14000 rpm abzentrifugiert und das Zellpellet in 50 µl Basalsalz resuspendiert. Nach Zugabe von 400 µl Lysepuffer wurden die Zellen durch mehrmaliges Pipettieren bis zum vollständigen Aufschluß gemischt. Das Zellysate wurde mit 800 µl Ethanol vorsichtig überschichtet. Die chromosomale DNA befindet sich in der Interphase und kann von dort aus auf eine Pasteurpipette gewickelt werden. Anschließend wurden die beiden Phasen gemischt und die Pasteurpipette mit der DNA in dem Reaktionsgefäß für 5 min inkubiert. Zum Waschen der DNA wurde die DNA zweimal für 5 min in Ethanol und einmal für 1 min in Ether getaucht. Nach kurzem Trocknen der DNA an der Luft (ca. 1 min) wurde die Pipettenspitze in 400 µl TE-Puffer getaucht, um die DNA zu lösen (üN bei 4°C inkubieren). Zum vollständigen Lösen wurde die DNA-Lösung für mindestens 20 min bei 60°C inkubiert.

Lysepuffer:

0,1 M NaCl	60 mg
1mM EDTA	3,7 mg
0,05% (w/v) SDS	5,0 mg
10 mM Tris	12 mg
ad 100 ml H <sub>2</sub> O bidest.	
mit Salzsäure pH 8,0 einstellen, autoklavieren	

TE-Puffer:

10 mM Tris	12 mg
1 mM EDTA	3,7 mg
ad 100 ml H <sub>2</sub> O bidest.	
mit Salzsäure pH 8,0 einstellen, autoklavieren	

### 8.2 Isolierung genomischer DNA aus *H. sal.* R1

Die Isolierung von genomischer DNA aus *H. sal.* R1 erfolgte mit Qiagen G500/G-Säulen nach Angaben des Herstellers für die Isolierung von DNA aus Bakterien.

### **8.3 Isolierung von Plasmid-DNA aus *E. coli***

Für die Präparation von Plasmid-DNA aus *E. coli* werden 4,5 ml üN-Kultur in LB-Medium mit dem entsprechenden Antibiotikum bei 37°C und 250 rpm inkubiert. Die anschließende Isolierung der Plasmid-DNA erfolgte mit dem Plasmid-Mini-Kit der Firma Qiagen nach Herstellerangaben. Dabei wurden die Zellen durch alkalische Lyse aufgeschlossen und die Plasmid-DNA unter Hochsalz-Bedingungen an die Silicagelmembran adsorbiert.

### **8.4 Isolierung von Gesamt-RNA aus *H. sal.* R1**

Die Qualität der verwendeten RNA bestimmt in erster Line die Ergebnisse der nachfolgenden Experimente. Deshalb sollte idealer Weise die RNA nicht mit DNA, Proteinen, Zuckern, Lipiden oder anderen Verbindungen kontaminiert sein.

Da RNA, im Vergleich zu DNA, instabiler ist und sie enzymatisch, chemisch oder physikalisch degeneriert werden kann, wurde bei der Isolierung und dem Umgang mit der Gesamt-RNA aus *H. sal.* R1 bei jedem Schritt darauf geachtet die Möglichkeiten einer Degradation so gering wie möglich zu halten.

Für Mikroarray-Analysen:

Die Isolierung der Gesamt-RNA aus *H. sal.* R1 erfolgte im Wesentlichen nach der Methode von Chomczynski und Sacchi, unter Verwendung der peqGOLD RNAPure Lösung von peqLab (siehe [11.2.1](#)).

Für andere Anwendungen:

Für andere Anwendungen als die Mikroarray-Analysen, wurde die Gesamt-RNA aus *H. sal.* R1 mit den RNeasy Mini / Midi Kits von Qiagen unter Verwendung der mitgelieferten Reagenzien und nach Angaben des Herstellers isoliert. Standardmäßig wurde ein DNaseI-Verdau mit dem DNasefree Kit (Ambion) durchgeführt, um mögliche DNA-Kontaminationen zu beseitigen. Nach der Isolierung wurde die Konzentration und Unversehrtheit der RNA mittels des Agilent 2100 Bioanalyser und des RNA 6000 Nano LabChip® Kits bestimmt und bis zur weiteren Verwendung bei -80°C gelagert.

## 8.5 Konzentrationsbestimmung von doppelsträngiger DNA (dsDNA)

Ein Aliquot der DNA-Lösung wurde mit H<sub>2</sub>O verdünnt und die Extinktion bei 260 und 280 nm gemessen. In proteinfreien Proben sollte das Verhältnis  $A_{260}/A_{280}$  ca. 2 betragen. Die Konzentration der Ausgangslösung berechnet sich nach folgender Formel:

$$c(\text{dsDNA}) = 50 \times A_{260} \times \text{Verdünnungsfaktor} \quad [\mu\text{g/ml}]$$

## 8.6 Herstellung von RbCl-kompetenten Bakterien

Für die Herstellung chemisch kompetenter *E. coli* Zellen wurden üN-Kulturen der entsprechenden *E. coli* Stämme in LB-Medium mit dem jeweils erforderlichen Antibiotikum bei 37°C, 250 rpm herangezogen. Dann im Verhältnis 1:100 in 500 ml frisches Medium überimpft und bis zu einer OD<sub>600</sub> von 0,5 weiter inkubiert. Anschließend wurden die Zellen 15 min auf Eis gekühlt und dann bei 4°C und 3500 rpm zentrifugiert. Das Zellpellet wurde in 16 ml RF-1-Lösung resuspendiert und für 1 h auf Eis inkubiert. Es folgte ein weiterer Zentrifugationsschritt für 15 min bei 4°C und 3500 rpm. Zum Resuspendieren des Pellets wurden 4 ml RF-2-Lösung benutzt und die Zellen für 15 min auf Eis gekühlt. Abschließend wurde die Bakteriensuspension aliquotiert (200 µl/Reaktionsgefäß), in flüssigem Stickstoff schockgefroren und bei -80°C gelagert.

RF-1-Lösung:

10 mM RbCl	1,2 g
6 mM MnCl <sub>2</sub> •4H <sub>2</sub> O	0,99 g
30 mM Kaliumacetat	3 ml (1 M Lösung)
1 mM CaCl <sub>2</sub> •2H <sub>2</sub> O	0,15 g
Glyzerin	15 g
ad 100 ml H <sub>2</sub> O bidest.	
mit 0,2 M Essigsäure den pH 5,8 einstellen, steril filtrieren	

RF-2-Lösung:

0,1 M MOPS	2 ml (0,5 M Lösung)
------------	---------------------



10 mM RbCl	1,2 g
7,5 mM CaCl <sub>2</sub> ·2H <sub>2</sub> O	1,1 g
Glyzerin	15 g
ad 100 ml H <sub>2</sub> O bidest.	
mit NaOH pH 6,8 einstellen, steril filtrieren	

## 8.7 Ligation von DNA-Fragmenten

Für einen Ligationsansatz wurden 10-30 µl Insert- und linearisierte Vektor-DNA mit passenden kohäsiven Enden in einem molaren Verhältnis von 3:1 bis 5:1 mit etwa 1U T4-Ligase pro µg DNA und einer entsprechenden Menge 5 x Ligasepuffer üN bei 9-12°C inkubiert.

## 8.8 Transformation kompetenter Bakterien

Die RbCl-kompetenten *E. coli* Zellen wurden 30 min auf Eis aufgetaut. Anschließend erfolgte die Zugabe von 10 µl Ligationsansatz bzw. die zu transformierende Plasmid-DNA (20-200ng). Die Transformationsansätze wurden 20 min auf Eis inkubiert, nachfolgend für 2 min bei 42°C einem Hitzeschock unterzogen und für 2 min auf Eis gestellt.

Pro Ansatz wurden 790 µl SOC-Medium zugegeben und 1 h bei 37°C, 250 rpm geschüttelt. Abschließend wurden die Bakterien in verschiedenen Verdünnungen auf antibiotikahaltigen Agarplatten ausplattiert und 16-20 h bei 37°C inkubiert.

SOC-Medium:

2% (w/v)	Bacto™ Trypton	20 g
0,5% (w/v)	Bacto™ Hefeextrakt	5 g
10,3 mM	NaCl	0,5 g
2,68 mM	KCl	0,2 g
ad 970 ml H <sub>2</sub> O bidest.		
mit NaOH pH 7,0 einstellen, autoklavieren		
plus	5 ml 2 M MgCl <sub>2</sub> -Lösung (steril)	
	5 ml 2 M MgSO <sub>4</sub> -Lösung (steril)	
	20ml 1 M Glucose-Lösung (steril)	
steril filtrieren		

## 8.9 Präparation kompetenter *H. sal.* R1 Zellen und Transformation

Für die Herstellung kompetenter *H. sal.* R1 Zellen wurde eine 35 ml Kultur benutzt, die mindestens dreimal alle zwei Tage überimpft wurde. Nach 1 ½ Tagen Wachstum wurden 2 ml Kultur eingesetzt und 5 min bei 5000 rpm abzentrifugiert und der Überstand abgenommen. Die Zellen wurden in 200 µl sterilem SPH-Puffer durch leichtes Schwenken des Eppendorfgefäßes resuspendiert. Durch Zugabe von 10 µl 0,5 M EDTA in SPH-Puffer wurden Sphäroplasten erzeugt, deren Ausbildung im Mikroskop verfolgt wurde.

Von der Plasmid-DNA wurden 5-10 µl mit derselben Menge 4 M NaCl gemischt und zu der Sphäroplastenlösung gegeben. Der Transformationsansatz wurde vorsichtig gemischt und für 10 min bei RT inkubiert. Anschließend wurden 200 µl 60% PEG<sub>600</sub> in SPH-Puffer in den Deckel des Reaktionsgefäßes pipettiert und nach dem Schließen des Deckels wurde durch Schwenken der Reaktionsansatz gemischt.

Das Eppendorfgefäß wurde mit SVL-Puffer aufgefüllt und der Ansatz für 5 min bei RT und 4000 rpm zentrifugiert.

Das Zellpellet wurde in 3 ml Halostandardmedium aufgenommen und üN bei 37°C inkubiert (sehr sanft schütteln). Von der transformierten Halokultur wurden verschiedene Aliquots (50-200 µl) auf mevinolinhaltige (10 µg/ml) Agarplatten ausplattiert und solange bei 40°C inkubiert bis Einzelkolonien zu sehen waren.

SPH(Sphäroplastenpräparations)-Puffer:

2 M NaCl	11,7 g
25 mM KCl	186 mg
50 mM Tris	606 mg
15% (w/v) Sucrose	15,0 g
ad 100 ml H <sub>2</sub> O bidest.	
mit Salzsäure pH 8,75 einstellen, steril filtrieren	

SVL(Sphäroplastenverdünnungs)-Puffer:

4,3 M NaCl	25,1g
80 mM MgSO <sub>4</sub> •7H <sub>2</sub> O	1,97g
27 mM KCl	201 mg

---

1,4 mM CaCl <sub>2</sub> •2H <sub>2</sub> O	20,6 mg
10 mM Na <sub>3</sub> Citrat•2H <sub>2</sub> O	294 mg
50 mM Tris	606 mg
15% (w/v) Sucrose	15,0 g
ad 100 ml H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,4 einstellen, steril filtrieren	

PEG-SPH-Puffer:

60% (v/v) Polyethylenglycol 600	33,8 g
ad 50 ml SPH-Puffer	

EDTA-SPH-Puffer:

250 mM EDTA	3,65 g
ad 50 ml H <sub>2</sub> O	

## 8.10 Spaltung von DNA mit Restriktionsendonukleasen

Je nach Konzentration wurden 1-5 µl Vektor-DNA mit 1 µl der Restriktionsenzymlösung und 1/10 des Endvolumens an Enzympuffer (10 x Puffer von New England Biolabs) auf ein Endvolumen von 10-20 µl mit sterilem H<sub>2</sub>O verdünnt und bei der vom Hersteller für das Enzym angegebenen Temperatur und Zeit inkubiert (zwischen 1-3 h bei 30-37 °C). Zum Verdau von genomischer DNA werden 20 µl DNA Lösung mit 2 µl der Restriktionsenzymlösung und 10 µl Enzympuffer auf ein Endvolumen von 100 µl verdünnt und über Nacht bei der angegebenen Temperatur inkubiert.

## 8.11 Polymerasekettenreaktion

Die Polymerase-Kettenreaktion (PCR, polymerase chain reaction) (Saiki *et al.* 1988) dient der selektiven Amplifikation von DNA-Fragmenten mittels thermostabiler DNA-Polymerase und spezifischer Oligonukleotide (Primer). Eine Standard-PCR besteht aus drei Reaktionsschritten, die zyklisch wiederholt werden. Nach einer Denaturierungsphase der DNA-Vorlage (Templat) folgt eine Hybridisierungsphase zur Anlagerung (Annealing) der einzelsträngigen Primer an

das DNA-Templat, im letzten Schritt eines Zyklus erfolgt die Verlängerung der Oligonukleotide durch eine thermostabile DNA-Polymerase (DNA-Synthese). Das Temperatur- und Zeitprofil und die Anzahl der Zyklen wurden entsprechend des DNA-Templats, der Größe des zu amplifizierenden Produktes und der Primer ausgewählt bzw. variiert. Die Annealing-Temperatur ( $T_M$ ) wurde nährungsweise mit der „2+4“-Regel ermittelt:

$$T_M = [(G+C) \times 4^\circ\text{C} + (A+T) \times 2^\circ\text{C}] - 5$$

Die Annealing-Temperatur lag für die Primer 5-10°C unter der berechneten Schmelztemperatur. Alternativ wurde die Schmelztemperatur mit dem Computerprogramm „Oligo Calculator“ berechnet (<http://www.basic.nwu.edu/biotools/oligocalc.html>).

Alle PCR-Reaktionen wurden mit dem vom Hersteller gelieferten Reaktionspuffer, 50 pmol Primern, je 1,25 mM dATP, dCTP, dGTP sowie dTTP, 100 ng DNA-Template und 2,5 U DNA-Polymerase durchgeführt. Bei hohem GC-Gehalt der zu amplifizierenden DNA wurden 1-3 µl DMSO hinzugefügt, was die Bildung von stabilen Sekundärstrukturen in bestimmten Bereichen unterdrücken und somit die Synthese begünstigen sollte.

Standard PCR-Ansatz:

DNA-Template (100 ng)	x µl
sense-Primer [10 pmol/µl]	5 µl
reverse-Primer [10 pmol/µl]	5 µl
10x PCR-Puffer	10 µl
dNTP-Mix (je 1,25mM pro dNTP)	16 µl
DMSO	3 µl
DNA-Polymerase	1 µl
ad 100 µl H <sub>2</sub> O bidest.	

Temperaturprofil einer Standard PCR:

	96°C	2 min
30 Zyklen	96°C	30 sec
	55°C	30 sec
	72°C	1 min/1kb
	72°C	7 min

96°C	2 min
4°C	∞

## 8.12 DNA-Sequenzierung

Die Sequenzierung von Plasmiden und PCR-Fragmenten erfolgte nach der Kettenabbruchmethode durch PCR mit fluoreszenzmarkierten Didesoxynukleotiden unter Gebrauch des ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kits von Perkin-Elmer Applied Biosystems. Der Reaktionsmix wurde gebrauchsfertig geliefert und enthielt DNA-Polymerasepuffer, Desoxynukleotide, basenspezifisch Fluoreszenzfarbstoff markierte Didesoxynukleotide und eine thermostabile DNA-Polymerase. Für eine Sequenzierung wurden 0,5-1 µg Vektor-DNA aus einer Standard-Minipräparation oder 0,2-0,5 µg eines gereinigten PCR-Produkts eingesetzt. Eine typische Sequenzierung hatte folgende Zusammensetzung:

DNA	1,5 µl
Primer [10 pmol/µl]	1,5 µl
BigDye™ Reaktionsmix	3,0 µl
BigDye Puffer	5,0 µl
5M Betain	3,0 µl

PCR-Bedingungen:

	94°C	30 s
25-30 Zyklen	60°C	4 min
	4°C	∞

### 8.12.1 Reinigung von Sequenzierungsproben

Die DNA im Sequenzierungsansatz wurde über Gelfiltrationssäulen MicroSpin™ G50 Säulchen von Amersham Biosciences nach Angaben des Herstellers aufgereinigt und anschließend in einer Vakuumzentrifuge (Speed Vac Concentrator) zur Trockne eingengt. Die Sequenzierung erfolgte per Gelelektrophorese in einem DNA-Sequenzierer (Modell 373A von Applied Biosystem).

### 8.13 Elektrophoretische Auftrennung von DNA-Fragmenten

Zur Auftrennung von DNA-Fragmenten wurde eine Flachbettgelelektrophorese in TAE-Puffer bei 60-80 V durchgeführt.

Die Agarosegele enthielten 0,8-2 % Agarose in TAE-Puffer und wurden nach abgeschlossener Elektrophorese 10-20 min in einer Ethidiumbromid-Lösung [0,1 mg/l] gefärbt und 10-20 min in H<sub>2</sub>O gewaschen. Zur Detektion und Aufnahme der DNA-Fragmente wurde das Gel auf einem UV-Schirm ( $\lambda = 366 \text{ nm}$ ) belichtet und mit einer Kamera photographiert. Ethidiumbromid interkaliert in die DNA und fluoresziert sehr stark unter UV-Licht, wodurch die DNA-Fragmente sichtbar gemacht werden konnten. Anhand der DNA-Längenstandards konnten die Fragmentgrößen von den DNA-Fragmenten bestimmt werden.

(50x) TAE-Puffer:

2 M Tris	242 g
Eisessig	57,1 ml
50 mM EDTA	37,2 g
ad 1l H <sub>2</sub> O bidest.	
mit Essigsäure pH 8,5 einstellen	

### 8.14 Elektrophoretische Auftrennung von RNA-Fragmenten

Die Gelelektrophorese wurde in horizontalen Gelapparaturen durchgeführt. Zur Inaktivierung von RNasen wurden der Gelschlitten, der Gelkamm und die Gelkammer mit RNaseZap<sup>®</sup> von Ambion behandelt. Zur Herstellung des Gels wurde 1,0 g Agarose in 72 ml H<sub>2</sub>O (bidest.) vollständig gelöst. Anschließend wurde auf ca. 60°C abgekühlt und 10 ml 10x MOPS-Puffer und 18 ml 37%-iges Formaldehyd zugegeben, vorsichtig durchmischt und die Lösung in eine Flachbettkammer gegossen. Das Aushärten des Gels erfolgte bei RT für ca 30 min. Die RNA-Proben wurden mit 2/3 RNA-Auftragspuffer versetzt, 15 min auf 65°C erhitzt und auf Eis abgekühlt. Anschließend wurden die Proben in die Geltaschen pipettiert und die Elektrophorese wurde in einer Flachbett-Gelkammer mit 1x MOPS-Puffer als Laufpuffer durchgeführt. Es wurde eine Spannung von 80 V angelegt und die Elektrophorese beendet, wenn das Bromphenolblau des Auftragspuffers das untere Ende des Gels erreicht hat.

10x MOPS-Puffer:

500 mM MOPS	104,7 g
10 mM EDTA	3,7 g
ad 1l H <sub>2</sub> O bidest.	

RNA-Auftragspuffer:

1x MOPS-Puffer	1 ml (10x MOPS-Puffer)
50% (v/v) Formamid	5 ml
6,8% (v/v) Formaldehyd	1,84 ml (37%ige Lösung)
11,7% (v/v) Glyzerin	1,34 ml (87%ige Lösung)
1% (v/v) Xylencyanol	100 µl (gesättigte Lösung)
1% (v/v) Bromphenolblau	100 µl (gesättigte Lösung)
ad 10 ml H <sub>2</sub> O bidest.	

## 8.15 Extraktion von DNA aus Agarosegelen

Nach der Auftrennung der DNA-Fragmente wurde das zu isolierende Fragment unter UV-Beleuchtung aus dem Gel herausgeschnitten und in ein Eppendorfgefäß mit Mobicol-Zentrifugeneinheit überführt und 1 h (zweimal 30 min) bei 6000 rpm zentrifugiert.

Die Isolierung der DNA aus dem Eluat erfolgte anschließend mit dem PCR-Purification-Kit von Qiagen mit den gelieferten Reagenzien nach den Angaben des Herstellers.

## 8.16 Southern-Hybridisierung mit Digoxigenin(DIG) -markierten Sonden

### 8.16.1 Herstellung DIG-markierter Sonden

Eine DIG-markierte DNA-Sonde für eine Southern-Hybridisierung wurde mittels einer PCR hergestellt, wobei zum Standard-Reaktionsansatz zusätzlich 1 µl DIG-markiertes dUTP zugegeben wurde. Die Sonde wurde über ein Agarose-Gel gereinigt und aus dem Gel extrahiert (siehe 8.15).

### 8.16.2 Fixierung der DNA auf Membranen (Vakuum-Blot)

Für einen Southern-Blot wurde die zu untersuchende genomische DNA über Nacht mit Restriktionsenzymen verdaut und 20 µl eines 100 µl-Ansatzes in einem 1%-igen Agarose-Gel aufgetrennt, wobei 10 µl eines DIG-markierten Größenstandards (Roche) aufgetragen wurden. Anschließend wurde das Gel auf dem Schüttler für 5 min in 0,25 M HCl, 30 min in 0,5 M NaOH und zuletzt bei 4°C 30 min in Tris-Waschlösung gewaschen.

Eine Vakuum-Blot-Apparatur wurde so abgedichtet, daß nur noch eine Fläche in der Größe des Gels offen war. Darauf wurden 3 mittelstarke, in 10x SSC-Puffer getränkte, Whatman-Papiere und die Nylon-Membran (Hybond-N, Amersham Biosciences) gelegt, die etwas größer sein sollte als das Gel. Auf diesen Sandwich wurde das Gel so aufgelegt, daß keine Luftblasen eingeschlossen wurden. Zur Abdeckung des Gels wurde abschließend ein weiteres Whatman-Papier benutzt. Danach wurde für 2-3 h ein Vakuum angelegt. Nachdem die Dicke des Gels auf etwa 1 mm abgenommen hatte, wurde die Nylon-Membran abgenommen und die DNA in einem UV-Crosslinker (zweimal auto-crosslink, je 1200x 100 µJ) kovalent auf der Membran fixiert.

Tris-Waschlösung:

1,5 M NaCl	87,6 g
0,5 M Tris	60,5 g
ad 1l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,5 einstellen	

20 x SSC-Puffer:

3 M NaCl	175,2 g
0,3 M Na <sub>3</sub> Citrat•2H <sub>2</sub> O	88,2 g
ad 1l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,0 einstellen, autoklavieren	

### 8.16.3 Hybridisierung DIG-markierter DNA-Sonden

Nach der Fixierung der DNA wurde die Membran mit etwa 10-15 ml Prähybridisierungspuffer (DIG Easy Hyb Lösung von Roche Diagnostics) in einer



Rollerflasche im Hybridisierungssofen bei 42°C ca. 60 min inkubiert. Inzwischen wurden 5 µl der DIG-Sonde in 3 ml DIG Easy Hyb Lösung 10 min in kochendem H<sub>2</sub>O erhitzt und in Eis abgeschreckt. Die denaturierte DIG-Sonden Lösung wurde zu der Membran und dem Prähybridisierungspuffer in die Rollerflasche gegeben und üN bei 42°C unter Drehen inkubiert. Die Membran wurde danach zunächst zweimal in Hybridisierungs-Waschpuffer I für 15 min bei RT und anschließend zweimal 15 min bei 68°C in Hybridisierungs-Waschpuffer II im Hybridisierungssofen gewaschen.

Hybridisierungs-Waschpuffer I:

2x SSC-Puffer	100 ml (20x SSC-Puffer)
0,1% (v/v) SDS	10 ml (10%-gen SDS-Lösung)
ad 1l H <sub>2</sub> O bidest.	

Hybridisierungs-Waschpuffer II:

0,1x SSC-Puffer	5 ml (20x SSC-Puffer)
0,1% (v/v) SDS	10 ml (10%-gen SDS-Lösung)
ad 1l H <sub>2</sub> O bidest.	

10% SDS:

10% (w/v) SDS	40 g
ad 400 ml H <sub>2</sub> O bidest.	

#### 8.16.4 Detektion der hybridisierten DNA

Die Membran wurde nach der Hybridisierung für 5 min in Waschpuffer und 30 min in Puffer II geschüttelt. Zur immunologische Detektion wurde das Anti-Digoxigenin Alkalische Phosphatase Fab-Fragment, als Antikörper gegen die DIG-Sonde, in einer Verdünnung von 1:10000 (in 15 ml Puffer II) verwendet und die Membran darin 30 min bei RT inkubiert. Nach zweimaligem Waschen im Waschpuffer (je 30 min) wurde 1 ml der gebrauchsfertigen CDP-Star Detektionslösung (Amersham Biosciences) zwischen zwei Folien auf der Membran verteilt und diese für 5 min im Dunkeln bei RT inkubiert. Danach wurde die Membran kurz auf ein Papiertuch gelegt und zwischen zwei neuen Folien möglichst falten- und luftblasenfrei eingeschweißt. In einer Expositionskassette

wurde ein Röntgenfilm auf die Membran gelegt und dieser je nach Intensität der Lumineszenz exponiert und entwickelt.

Puffer I:

0,1 M Maleinsäure	11,7 g
0,15 M NaCl	8,8 g
ad 1l H <sub>2</sub> O bidest.	
mit NaOH pH 7,5 einstellen, autoklavieren	

Blocking Lösung:

10% (w/v) Blocking Reagent	25 g
ad 250 ml Puffer I, autoklavieren (Lagerung bei 4°C)	

Puffer II:

1% Blocking Lösung	10 ml (10% Blocking Lösung)
Ad 100 ml Puffer I	

Waschpuffer:

0,3% (v/v) Tween 20	3 ml
ad 1l Puffer I	

## 8.17 Northern-Hybridisierung mit radioaktiv markierten Sonden

Für die Northern-Blot Analysen wurden das NorthernMax™ und DECAprime™ II DNA-Markierungs Kit von Ambion verwandt.

### 8.17.1 Herstellung der radioaktiv markierten Sonden

Die radioaktive Markierung der Sonden wurde mit dem DECAprime™ II DNA-Markierungs Kit von Ambion wie folgt durchgeführt.

Es wurden 25 ng der zu markierenden DNA mit 2,5 µl 10x „decamer solution“ gemischt und 5 min bei 96°C inkubiert. Danach wurden die Proben in fl. N<sub>2</sub> schockgefroren und auf Eis langsam aufgetaut. Zu der denaturierten DNA/Decamer-Mischung wurden 5 µl vom 5x dATP-Puffer, 5 µl [α-<sup>32</sup>P]dATP, 1,5 µl

H<sub>2</sub>O und 1 µl der Exonuklease-freien Klenow-DNA-Polymerase (Endvolumen 25 µl) dazugegeben und gemischt. Anschließend wurde die Mischung 10 min bei 37°C inkubiert und die Reaktion durch Zugabe von 1,0 µl 0,5 M EDTA-Lösung gestoppt. Zur Abtrennung von nicht eingebauten markierten Nukleotiden und zur Reinigung der markierten Sonden wurden die Gelfiltrationssäulen MicroSpin™ G50 von Amersham Biosciences benutzt (siehe Reinigung von Sequenzierungsproben).

### **8.17.2 Fixierung der RNA auf Membranen (Vakuum-Blot)**

Die Auftrennung der RNA Proben wurde durch denaturierende formaldehyd-basierende Gelelektrophorese mit dem NorthernMax™ Kit von Ambion nach Herstellerangaben durchgeführt (analog zu 8.14).

Der Transfer der RNA auf die Hybond-N+ Membran wurde analog dem Protokoll zur Southern-Blot Analyse mittels einer Vakuum-Blot-Apparatur durchgeführt. Nach dem Transfer wurde die RNA in einem UV-Crosslinker (zweimal auto-crosslink, je 1200x 100 µJ) kovalent auf der Membran fixiert.

### **8.17.3 Hybridisierung radioaktiv markierter RNA-Sonden**

Nach der Fixierung der RNA wurden die Membranen in Rollerflaschen mit 15 ml ULTRAhyb™-Lösung mindestens 30 min bei 68°C prähybridisiert.

Inzwischen wurde zu der radioaktiv markierten Sonde 50 µl einer 10 mM EDTA-Lösung dazugegeben und 10 min bei 90°C inkubiert. Zu der denaturierten Sonde wurden 500 µl ULTRAhyb™-Lösung gegeben und diese Mischung zu der Membran und dem Prähybridisierungspuffer in der Rollerflasche pipettiert. Die Membran wurde bei 68°C für 2 h hybridisiert und anschließend zweimal 15 min bei RT mit Waschlösung I und zweimal 20 min bei 68°C mit Waschlösung II gewaschen. (Die Waschlösungen I und II sind äquivalent zu den Hybridisierungs-Waschpuffern I und II beim Southern-Blot.) Abschließend wurde die Membran auf ein Whatman-Papier gelegt, in Frischhaltefolie gewickelt und die Signale mittels Autoradiografie visualisiert.

## 8.18 Quantitative RT-PCR

Die Quantifizierung der RNA Transkriptmenge basiert auf einer Zweischnitt RT-PCR basierend auf dem SYBR Green PCR Master Mix Kit von Applied Biosystems. Das Prinzip der Quantifizierung beruht auf der Messung der Amplikonmenge nach jedem Temperaturzyklus. Die Detektion der Fluoreszenzsignale von SYBR Green erfolgt mit dem GeneAmp 5700 Sequence Detection System von Applied Biosystems.

Das Primerdesign für die RT-PCR wurde nach Herstellerangaben mit der Software Primer Express 2.0 (Applied Biosystems) durchgeführt. Die vorgeschlagenen Primer der Software wurden vor einem Einsatz in der RT-PCR mittels des Oligonucleotide Properties Calculator auf Sequenzinteraktionen hinüberprüft und nur eingesetzt wenn auch die Prüfung der Sequenzspezifität mit Hilfe einer in-silico PCR erfolgreich war ([www.basic.nwu.edu/biotools/oligocalc.html](http://www.basic.nwu.edu/biotools/oligocalc.html) / [www.in-silico.com/s\\_pcr.html](http://www.in-silico.com/s_pcr.html)).

Die Synthese der cDNA erfolgte mit der reversen Transkriptase SuperscriptIII von Invitrogen nach Angaben des Herstellers. Eingesetzt wurden 5µg Gesamt-RNA von *H. sal. R1* und 550ng random hexamer Primer von Promega.

Die anschließende Quantifizierung der Transkriptmenge erfolgte als dreifaches Replikat in 25µl Reaktionsansätzen mit einer Primerendkonzentration von 0,2µM. Das Reaktionsgemisch und das Temperaturprofil der Quantifizierungs PCR entsprachen den Herstellerangaben.

Als interner Standard zur Bestimmung der Transkriptmenge wurde das Ferredoxin-Gen OE4217R (*fdx*, *fer1*) verwendet. Die Berechnungen basieren auf der  $2^{\Delta\Delta Ct}$ -Methode, die eine Effizienz von 100% für jeden PCR-Zyklus voraussetzt.

Bei dieser Methode wurde so vorgegangen, dass am Anfang aus den Mehrfachbestimmungen die Mittelwerte errechnet wurden. Von den Mittelwerten der zu untersuchenden Genen wurde der Mittelwert des Referenzgens (*fer1*) subtrahiert und man erhält die  $\Delta Ct$ -Werte der einzelnen Gene. Auf der Grundlage der ermittelten  $\Delta Ct$ -Werte können z.B. zwei Wachstumsbedingungen miteinander verglichen werden, in dem man die  $\Delta Ct$ -Werte der verschiedenen Wachstumsbedingungen eines Gens von einander abzieht. Als Ergebnis bekommt man die  $\Delta\Delta Ct$ -Werte. Der Unterschied der RNA-Transkripte wurde berechnet durch  $2^{\Delta\Delta Ct}$ .

## 9. Zellfreie Expression von Proteinen

### 9.1 *In vitro* Expression (RTS-100 System)

#### 9.1.1 Einleitung

Zur *in vitro* Expression von Proteinen wurde das „Rapid Translation System“ (RTS) von Roche Diagnostics benutzt. Mit Hilfe des RTS AviTag *E. coli* Biotinylation Kits (Linear Template), wurde mittels PCR ein lineares Template hergestellt und eine *in vitro* mono-Biotinylierung unter Verwendung des RTS 100 Expressionssystems durchgeführt.

Das Rapid Translation System (RTS) basiert auf der Aktivität einer Zellextraktion von *E.coli*-Zellen. Mit diesem System ist die präparative Synthese von bis zu 5 mg eines gewünschten Proteins innerhalb 24 Stunden möglich. Die Proteinproduktion in einer solchen Reaktionslösung hat den Vorteil, dass ein passender Expressionsvektor direkt in die Flüssigkeit gegeben werden kann. Gegenüber der Expression in Zellen, die zunächst mit dem Vektor transformiert und danach kultiviert werden mussten, bedeutet dies einen erheblichen Zeitvorteil. Aufgrund der hohen Effizienz und Ausbeute der T7-RNA-Polymerase, wird ausschließlich das mit dem Vektor eingebrachte Gen transkribiert und anschließend translatiert.

Biotin (Vitamin H) ist ein kleines Molekül und besitzt eine große Affinität und sehr hohe Spezifität zu den Proteinen Avidin und Streptavidin. Die Wechselwirkungen zwischen Biotin und Avidin / Streptavidin sind, mit einer Bindungskonstante von  $10^{-15} \text{ M}^{-1}$ , extrem stark und zählen deshalb zu den stärksten nicht kovalenten biologischen Wechselwirkungen. So ermöglicht der Einbau von Biotin an spezifischen Stellen in Proteinen den Zugang zu der leistungsfähigen Avidin / Streptavidin Technologie in der Immobilisierung, der Reinigung und Detektion von Proteinen.

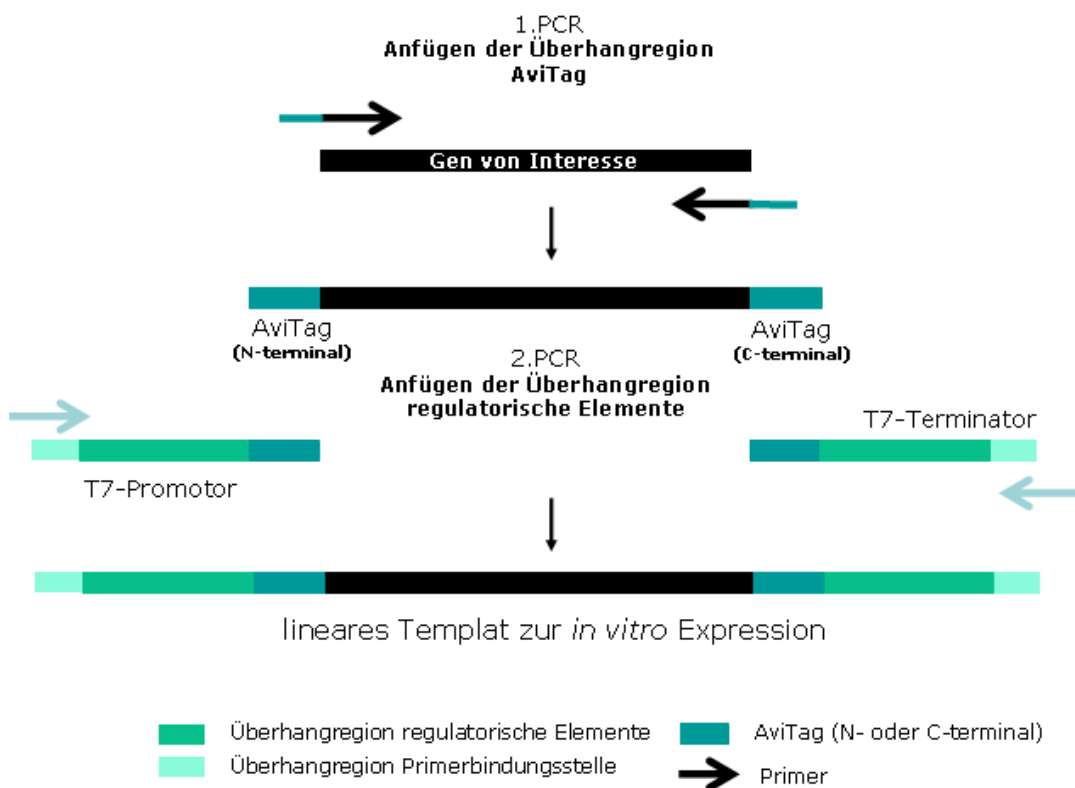
Mit Hilfe der AviTag Technologie (Avidity Inc., Denver Colorado, USA) in Kombination mit dem RTS-System, ist es möglich, eine *in vitro* Expression von mono-biotinylierten Proteinen durchzuführen. Die Nukleotidsequenzen der RTS kompatiblen Expressionsvektoren (pIVEX2.7d und pIVEX2.8d) enthalten Sequenzen für die Einführung eines C-terminalen oder N-terminalen AviTags. Der AviTag, eine 15 Aminosäure lange Peptidkette, dient als Sequenz für die spezifische Biotinylierung durch die sogenannte Biotin Protein Ligase (BirA) aus

*E. coli*, die in einer ATP verbrauchenden spezifischen Reaktion an eine  $\epsilon$ -Aminogruppe eines Lysins ein Biotin anhängt.

## 9.2 Herstellung des linearen Expression-Templates

Das lineare DNA Fragment, welches geeignet ist zur *in vitro* Expression von Proteinen, wurde durch eine Zweischritt PCR hergestellt.

Die erste PCR Amplifikation wurde mit einem Paar von genspezifischen Primern durchgeführt, die an beiden Enden der Gensequenz einen Überhang von 20-21 Basen einführen. Diese zusätzlichen Sequenzen dienen während der zweiten PCR als Bindungsstellen für die Primer, die die regulatorischen Elemente, wie T7



Promotor, ribosomale Bindungsstelle und T7 Terminator, einführen.

Abb. 86: Schema der Zweischritt PCR zur Herstellung des linearen Templates zur *in vitro* Expression.

Erste genspezifische PCR:

5x Phusion GC-Puffer	10 $\mu$ l
dNTP-Mix (je 1,25mM pro dNTP)	8 $\mu$ l
DMSO	1,5 $\mu$ l
sense-Primer (10 pmol/ $\mu$ l)	5 $\mu$ l
reverse-Primer (10 pmol/ $\mu$ l)	5 $\mu$ l

## Methoden

DNA-Template (100 ng)	1 $\mu$ l
Phusion DNA-Polymerase	0,5 $\mu$ l
H <sub>2</sub> O bidest.	19 $\mu$ l
	$\Sigma$ 50 $\mu$ l

PCR-Programm:

	98°C	30 sec
	98°C	10 sec
30 Zyklen	60°C	10 sec
	72°C	15 sec
	72°C	6 min
	4°C	$\infty$

Nach erfolgreicher Amplifikation des DNA Fragments, wurde dieses nach Agarosegelelektrophorese aufgereinigt und für die Überhang-PCR eingesetzt.

Überhang-PCR:

5x Phusion GC-Puffer	10 $\mu$ l
dNTP-Mix (je 1,25mM pro dNTP)	8 $\mu$ l
DMSO	1,5 $\mu$ l
T7 Promotor Primer (6 $\mu$ M)	4 $\mu$ l
T7 Terminator Primer (6 $\mu$ M)	4 $\mu$ l
Erstes PCR-Produkt als	5 $\mu$ l
DNA-Template (150-300 ng)	
C-terminale oder N-terminale	1 $\mu$ l
AviTag DNA	
Phusion DNA-Polymerase	0,5 $\mu$ l
H <sub>2</sub> O bidest.	16 $\mu$ l
	$\Sigma$ 50 $\mu$ l

PCR-Programm:

	98°C	30 sec
	98°C	10 sec
30 Zyklen	60°C	10 sec
	72°C	17 sec

---

98°C	30 sec
72°C	6 min
4°C	∞

Das Produkt der zweiten PCR wurde ohne weitere Reinigung für die *in vitro* Protein Expression eingesetzt.

### **9.3 *In vitro* Protein Expression und Biotinylierungsreaktion**

Die Proteinexpression wurde mit dem RTS 100 *E. coli* HY Kit nach Angaben des Herstellers wie folgt durchgeführt.

Bevor mit der *in vitro* Proteinexpression begonnen werden konnte, wurde als erster Arbeitsschritt das lyophilisierte BirA-Enzym durch Zugabe des Rekonstitutions-Puffers rekonstituiert. Anschließend wurde der Biotinylierungs Mix hergestellt aus einer 1:1 Mischung des rekonstituierten BirA-Enzyms und der Biotin Energie Mischung.

Für einen 50 µl Expressionsansatz wurden folgende Lösungen in ein Reaktionsgefäß pipettiert:

- 12 µl *E. coli* Lysat
- 10 µl Reaktionsmix
- 12 µl Aminosäuren
- 1 µl Methionin
- 5 µl Biotinylierungsmix
- 10 µl lineares Expressionstemplate

Der Expressionansatz wurde für 2 h bei 30°C in einer PCR-Maschine inkubiert. Nach der Expression wurden die Reaktionsgefäße 10 min bei 14000 rpm und 4°C zentrifugiert, der Überstand in ein neues Reaktionsgefäß überführt und das Pellet in einer 1%-igen SDS-Lösung resuspendiert. Der Überstand und das resuspendierte Pellet wurden mittels SDS-PAGE analysiert.

#### **9.3.1 SDS-PAGE und Streptavidin Western Blot**

Die Analyse der exprimierten Proteine mittels SDS-PAGE wurde mit kommerziell erhältlichen vorgefertigten SDS-Gelen und Pufferlösungen (NuPAGE / Invitrogen) durchgeführt.



Die aufgetrennten Proteine wurden auf eine PVDF-Membran transferiert. Dafür wurde eine Transfermembran auf die Größe des Gels zurechtgeschnitten und kurz mit Methanol benetzt und danach für 5 min im Transferpuffer (20 % v/v Methanol, 192 mM Glycin, 0.01 % w/v Natriumdodecylsulfat, 25 mM Tris/HCl, pH 8.6) äquilibriert. Das Gel wurde blasenfrei auf der Anode der Blotapparatur auf drei zurechtgeschnittene, in Puffer eingeweichte Whatman-Papiere gelegt und mit der Transfermembran bedeckt. Darauf wurden wiederum drei feuchte Whatman-Papiere gelegt und anschließend die Kathode aufgesetzt. Der Transfer erfolgte mit einer Stromdichte von  $1.2 \text{ mA cm}^2$  für 90 min. Beim Auseinandernehmen der Apparatur wurden zunächst die Ränder des Gels auf der Membran markiert, bevor diese abgezogen wurde. Danach erfolgte eine Inkubation für 1 h in TBST-Puffer mit 5% BSA. Es folgte eine Inkubation für 50 min mit dem Streptavidin-POD Konjugat (1:5000 in TBST-Puffer mit 5% BSA) (Roche Diagnostics). Abschließend wurde die Membran dreimal für 15 min mit TBST-Puffer gewaschen und mit einem 1:1 Gemisch der LumiLight Detektionslösungen (Roche Diagnostics) für 5 min im Dunkeln bei RT inkubiert. Danach wurde die Membran kurz auf ein Papiertuch gelegt und zwischen zwei Folien möglichst falten- und luftblasenfrei eingeschweißt. In einer Expositionskassette wurde ein Röntgenfilm auf die Membran gelegt und dieser je nach Intensität der Lumineszenz exponiert und entwickelt.

10x TBST-Puffer:

100 mM Tris	12,1g
1,5 M NaCl	87,7g
4,5 mM Tween-20	5 ml (5,53g)
ad 1 l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,4 einstellen	

## 10. Biochemische Methoden

### 10.1 Schnelltest für die Konzentrationsmessung von BR

Es wurden zweimal 4 ml von der zu untersuchenden Kultur 15 min bei 8000 rpm zentrifugiert und der Überstand sorgfältig abgenommen. Die eine Fraktion des Zellpellets wurde mit 1 ml, die andere mit 900  $\mu\text{l}$  H<sub>2</sub>O plus DNase I lysiert. Zu dem zweiten Zelllysate wurden 100  $\mu\text{l}$  einer 0,1 M CTAB-Lösung (N-Cetyl-N,N,N-

trimethylammoniumbromid) dazugegeben und gemischt. Beide Lysate wurden ün bei 4°C gelagert und die Spektren an einem UV/VIS-Spektrometer aufgenommen. Die Konzentration von Bakteriorhodopsin wurde anhand der Absorption in den Differenzspektren der Proben mit und ohne CTAB-Behandlung bei 568 nm berechnet.

## 10.2 Isolierung der BR-Membranfraktion

Die Isolierung der BR-haltigen Membranfraktionen wurde nach dem Protokoll von Oesterhelt und Stoeckenius (1974) durchgeführt.

Es wurden 500 ml der zu untersuchenden Kultur in GSA-Bechern bei 10°C und 9500 rpm für 20 min zentrifugiert. Die Zellen wurden in 3 ml Basalsalz plus DNase I resuspendiert und mit 12 ml H<sub>2</sub>O lysiert. Das Zelllysat wurde 1 h bei 40000 rpm und 4°C in einem Ti60-Rotor zentrifugiert und das Pellet in 2 ml H<sub>2</sub>O resuspendiert, gepottert und in einem TL100.1 Rotor 45 min bei 60000 rpm und 10°C zentrifugiert. Der Überstand wurde vorsichtig abgesaugt und das Pellet in 200 µl H<sub>2</sub>O resuspendiert. Für die Vergleichbarkeit der produzierten Proteinmengen wurden die Proben entsprechend ihren Klett-Werten verdünnt. Anschließend wurde die Gesamtmembranfraktion (GMF) auf einen Saccharosegradienten (25-45 %, (w/w)) aufgetragen und 14 h bei 25000 rpm und 10°C im TST41-Rotor zentrifugiert. Die erhaltenen Gradienten wurden fotografiert und die BR-haltigen Banden isoliert und ohne weitere Reinigung im UV/VIS-Spektrometer vermessen.

## 10.3 Sauerstoffmessung in *H. sal.* R1 Kulturen

Die Messungen des Sauerstoffgehaltes in *H. sal.* R1 Kulturen wurde mit dem Microprocessor Oximeter OXI 2000 von WTW durchgeführt. Der WTW-Sauerstoffsensoren arbeitet nach dem Prinzip der membranbedeckten polarographischen Sensoren. Die Oberfläche einer negativ geladenen Goldelektrode reduziert den Sauerstoff zum Hydroxid. Bei dieser chemischen Reaktion fließt ein elektrischer Strom, der der Menge des chemisch umgesetzten Stoffes proportional ist. Der Strom ist das Meßsignal des Sensors und das Meßgerät berechnet aus dem Sensorsignal die Konzentration des Sauerstoffs. Die Meßwerte wurden online von einem Apple Macintosh Computer aufgezeichnet

und anschließend grafisch ausgewertet. Zum Kalibrieren des Meßgerätes, nach dem Einlaufen (Polarisieren) des Sensors, wurde ein spezielles Luftkalibriergefäß benutzt, daß für ideale Kalibrierbedingungen mit definierter Luftfeuchte sorgt.

## 11. Transkriptom-Analyse

### 11.1 Herstellung des DNA-Mikroarrays

#### 11.1.1 Primeroptimierung zur Herstellung der Sonden

Die Primerauswahl und Optimierung zur Amplifikation der Gensonden wurde mit Hilfe des Programms PrimeArray durchgeführt. Es ermöglicht im Gesamt-Genom-Maßstab die automatische Extraktion von kodierenden Sequenzen aus verschiedenen Eingabedateiformaten.

Über die grafische Benutzeroberfläche ist es möglich wichtige Parameter zum Primerdesign, wie Schmelztemperatur (TM), Produktlängen, GC-Gehalt usw., frei zu wählen. Während des Optimierungsprozesses werden die Primerpaar-Kandidaten über ein iteratives Verfahren solange entlang der Gensequenz verschoben, bis die gewählten Parameter wie Schmelztemperatur, GC-Gehalt usw. erfüllt sind und Primer-Primer Wechselwirkungen ausgeschlossen werden können. Die Ergebnisse dieser Primeroptimierung können als Excel- oder ASCII-Datei exportiert werden.

Der erste Schritt beim Ablauf des Primerdesigns war das Öffnen und Einlesen der Dateien mit den Sequenzen der ORFs des *H. sal.* R1 Genoms. Die Sequenzen der Gene, für die die Primer berechnet werden sollten, waren in Dateien im FASTA-Format gespeichert und wurden, je nach Status der Annotation des *H. sal.* R1 Genoms, von der Bioinformatikgruppe für das Primerdesign bereitgestellt. Anschließend wurden über das grafische Eingabefenster die Optionen zur Berechnung der Primerpaare eingegeben. Dabei wurden folgende Parameter speziell für dieses Projekt gewählt:

Parameter	Einstellungen
Schmelztemperatur	
minimal	71°C
maximal	72°C
GC-Gehalt	

minimal	60%
maximal	70%
Einzug von der ersten Base	50 Basen
Länge der Sequenz (maximal)	550 Basen
Länge des PCR-Produkts (minimal)	100 Basen

Alle anderen Parameter wurden in ihren Standardeinstellungen nicht verändert.

Nach erfolgter Eingabe der Parameter wurde von dem Programm die Berechnung und Optimierung der Primerpaare durchgeführt und das Ergebnis mit der Dateiendung (\*.pri) versehen und gespeichert. Abschließend wurde das Ergebnis in eine Excel-Datei überführt und dort weiter bearbeitet.

So wurden zuerst die von PrimeArray automatisch vergebenen Namen für die Primer verändert, in dem der Namenszusatz *F* bzw. *R* durch *for* bzw. *rev* ersetzt wurde. Als nächstes wurden die Gene, für die mit den gewählten Einstellungen keine Primer berechnet werden konnten, aus der Datei gelöscht und das Format der Excel-Datei nach den Vorgaben der Firma *metabion* konvertiert. Die Konvertierung beinhaltete zum Einen das Einteilen der Gene / Primer in 96-iger Einheiten, da dies dem Synthese- und Versandformat der Primer entsprach und zum Anderen wurde festgelegt, dass sich in einer Spalte einer 96-iger Platte immer der *forward*-Primer und in der daneben liegenden Spalte der *reverse*-Primer befindet, so dass mit einer Primer Platte 48 Gene (ORFs) mittels PCR amplifiziert werden konnten. Durch diese Festlegung der Position der Primer im 96-iger Format wurde auch die Position des PCR-Produktes auf der PCR-Mikrotiterplatte bestimmt, d.h. man braucht für eine 96-iger PCR-Platte zwei 96-iger Platten mit Primern.

Zur Synthese wurden diese so modifizierten Dateien mit den Sequenzen für die Primer an die Firma *metabion* übermittelt.

### 11.1.2 Herstellung der Sonden

Für die Herstellung des Gesamtgenom-Mikroarrays von *H. sal.* R1 wurden alle „open reading frames“ (ORFs) mit genspezifischen Primern amplifiziert. Diese Primer wurden, wie unter 11.1.1 beschrieben, mit Hilfe des Computerprogramms PrimeArray berechnet und im Hochdurchsatz-Verfahren im 96-iger Format von der Firma *metabion* synthetisiert. Alle Primer wurden so geliefert, daß sie im gleichen Volumen eine Konzentration von 100 pmol/µl besaßen. Diese Vorgaben ermöglichten den Einsatz von Mehrkanalpipetten zum Pipettieren. Ausgehend von

diesen gelieferten 96-iger Platten wurde eine Arbeitslösung in der Konzentration von 10 pmol/μl in neuen 96-iger Platten angefertigt. Die Platten mit den Oligonukleotiden wurden bis zur weiteren Verwendung bei -20°C gelagert.

Zur PCR-Amplifikation der ORFs wurde die HotStarTaq™ DNA Polymerase von Qiagen benutzt, die zum Einen besonders für die Amplifikation von GC-reichen Sequenzen geeignet ist und zum Anderen ist die Polymerase so modifiziert, dass sie bei Raumtemperatur keine Polymeraseaktivität besitzt. Erst durch eine Inkubation bei 95°C von 15 min, wird die Polymerase aktiviert.

Für die PCR-Reaktionen im 96-iger Format wurde ein „Master-Mix“ für 100 Reaktionen in einem 15 ml Reaktionsgefäß mit folgender Zusammensetzung hergestellt:

	Reaktionsmix für 1 Ansatz	Master-Mix für 100 Ansätze
10x PCR-Puffer (+15mM MgCl <sub>2</sub> )	10 μl	1000 μl
5x Q-Solution	20 μl	2000 μl
dNTP-Mix (je 1,25 mM pro dNTP)	16 μl	1600 μl
genom. DNA ( <i>H. sal</i> R1 [100ng/μl])	1 μl	100 μl
H <sub>2</sub> O bidest.	42,75 μl	4275 μl
HotStarTaq™	0,25 μl	25 μl
	Σ 90 μl	Σ 9000 μl

Von diesem Master-Mix wurden 96 mal 90 μl in die PCR-Mikrotiterplatten vorgelegt und anschließend mit einer Mehrkanalpipette je 5 μl der genspezifischen Primer zu dem Reaktionsmix pipettiert.

Die PCR wurde in einem GeneAmp PCR-System 9700 Thermocycler mit folgendem Programm durchgeführt:

PCR-Programm zur Amplifikation der Sonden „probes“ :

	95°C	15 min
	96°C	1 min
30 Zyklen	55°C	30 sec
	72°C	50 sec
	4°C	∞

Es wurde für die Amplifikation sämtlicher ORFs des *H. sal.* R1 Genoms das gleiche PCR-Programm verwendet, da die Primer zur Vereinfachung der PCR mit Hilfe des Programms PrimeArray so optimiert wurden, daß sie annähernd alle die gleiche Schmelztemperatur besitzen.

Im Anschluss an die PCR wurden von jedem PCR-Ansatz 5 µl für eine Agarosegelelektrophorese eingesetzt, um die Effizienz der Reaktion zu überprüfen. Zur Kontrolle der Eindeutigkeit der generierten PCR-Produkte wurden stichprobenartig PCR-Produkte sequenziert und die Sequenzen mittels BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool)-Suche gegen das *H. sal.* R1 Genom überprüft.

Bis zur weiteren Verwendung wurden die PCR-Platten bei -20°C gelagert.

### **11.1.3 Aufreinigung der Sonden-PCRs**

Für die Aufreinigung der PCR-Produkte wurden MultiScreen 96 PCR-Filterplatten in Kombination mit dem MultiScreen „vakuum manifold“ von Millipore verwendet. Bei dieser Aufreinigungsmethode, ohne Zentrifugation oder Fällung, wurde eine Vakuumfiltration mittels einer Größenausschluß-Membran angewandt, um in einem Schritt eine hohe Reinheit und Ausbeute an PCR-Produkt zu gewährleisten.

Um die Ausbeute an gereinigtem PCR-Produkt zu erhöhen, wurden, bevor die PCR-Ansätze auf die Filterplatte pipettiert wurden, 100 µl H<sub>2</sub>O bidest. in jede Plattenvertiefung pipettiert und für 5-10 min ein Vakuum von 10 Zoll Hg-Säule eingestellt. Dadurch wurde die Membran mit H<sub>2</sub>O gesättigt, wodurch der Verlust um das Rückhaltevolumen der Membran von ca. 4 µl / Plattenvertiefung erniedrigt wurde. Jetzt wurden die PCR-Ansätze auf die MultiScreen-Platte gegeben und erneut ein Vakuum von 10 Zoll Hg-Säule angelegt. Zur Verbesserung der Reinheit wurde zweimal in jede Plattenvertiefung 40 µl H<sub>2</sub>O bidest. gegeben und per Vakuum abgesaugt. Abschließend wurden zur Rückgewinnung der PCR-Produkte 33 µl des „Spotting-Puffers“ auf die Membran pipettiert und die Filterplatte zum resuspendieren der DNA für 20 min auf einem Plattenschüttler kräftig geschüttelt. Die gereinigten PCR-Produkte wurden in neue Mikrotiterplatten überführt, verschlossen und bis zur weiteren Verwendung bei -20°C gelagert.

Spotting-Puffer:

3x SSC-Puffer	15 ml 20x SSC-Puffer
1,5 M Betain	23,04 g
ad 100 ml H <sub>2</sub> O bidest.	

20x SSC-Puffer:

3 M NaCl	175,3 g
0,3 M Na <sub>3</sub> Citrat • 2H <sub>2</sub> O	88,2 g
ad 1 l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,0 einstellen, autoklavieren	

### 11.1.4 Herstellung der DNA-Mikroarrays

Als Trägermaterial für die DNA wurden Glasobjektträger („Slides“) benutzt, die mit  $\gamma$ -Aminopropylsilan beschichtet wurden (CMT-GAPS II Slides, Corning), wodurch die Oberfläche eine erhöhte Hydrophobizität aufweist und eine größere Adhärenz der aufgetragenen DNA erzielt werden kann.

Für das Drucken der Mikroarrays in großem Maßstab wurde der Chipwriter™Pro von Virtek, ein vollautomatischer Roboter („Arrayer“), eingesetzt, dessen Hauptkomponente ein dreiachsiger Roboterarm mit einem Druckkopf („Print Head“) inklusive Drucknadeln („Pins“) bildet. Die Sondendichte, die mit einem solchen Roboter erzielt werden kann, ist neben der Anzahl der Nadeln im Druckkopf, auch von dem aufgetragenen Sondenvolumen (in Abhängigkeit von der Druckmethode und der Nadelbeschaffenheit) und dem Abstand zwischen den Sonden (limitiert durch die Bewegung des Roboterarms) abhängig.

Die *H. sal.* R1-DNA-Mikroarrays wurden mit folgender Konfiguration des Kontaktpotters Chipwriter™Pro nach einem Protokoll von Dr. Markus Panhuysen (Max Planck Institut für Psychiatrie) hergestellt.

#### Druckkopf

Nadelausführung	Quill Pin / Federkielnadel (SMP3 Telechem)
Anzahl der Nadeln (X/Y)	48 (X/Y = 12/4)
Größe der Slides (X/Y)	75mm x 25mm
Anzahl der verwendeten Slides (X/Y)	5 x 10 ( $\Sigma$ 50 Slides)

Rastergröße (X/Y)	X/Y 57,5mm x 17,5mm (10,05 cm <sup>2</sup> )
Abstände zwischen den Sonden	200 µm
Anzahl der Sonden pro Raster (X/Y)	20/16
Anzahl der Wiederholungen pro Gen	5
Auftrittsgeschwindigkeit der Nadeln	9 mm/sec
Druckrichtung des Rasters	links→rechts; dann oben→unten
<b>Bezugsquelle</b>	
Plattenausführung	348 Mikrotiterplatten
Anzahl der Platten	8
Anzahl der Plattenhalter	2
Benutzung Plattenabdeckung	Ja
<b>Waschschritte</b>	
Anzahl der Waschzyklen	4
Dauer des Waschzyklus	3 sec
Dauer Vakuumtrocknung	1 sec
<b>Klimaumgebung</b>	
Raumtemperatur	~ 20°C (20-25°C)
Luftfeuchtigkeit	~ 45% (40-50%)

Alle anderen Einstellungen waren Standardeinstellungen des Systems und wurden nicht verändert.

Die Konfiguration des Druckkopfes des ChipWriterPro mit 48 Nadeln machte es notwendig, die im 96-iger Format hergestellten Gensonden, in Mikrotiterplatten im 384-iger Format zu überführen. Dies wurde am MPI für Psychiatrie automatisch mit dem Robotersystem Biomek 2000 Orca (Beckmann Coulter) durchgeführt.

Vor jeder Arrayproduktion wurde routinemäßig das gesamte System des Spotters an Hand einer Checkliste, die ebenfalls am Max-Planck-Institut für Psychiatrie erarbeitet wurde, auf seine Funktionsfähigkeit hin überprüft.



### **11.1.5 Weiterverarbeitung der DNA-Mikroarrays**

Direkt nach dem Spotten der DNA-Mikroarrays, wurden die Slides mit der bedruckten Seite nach oben für 10 sec auf einen 80°C heißen Heizblock gelegt und abschließend zum weiteren Fixieren der DNA in einem UV-Crosslinker mit 3000x 100  $\mu$ J auf der Oberfläche fest verankert, da durch die UV-Quervernetzung die DNA kovalent an die Oberfläche des Slides gebunden wurde.

Zum Entfernen der nicht gebundenen DNA und zur Inaktivierung der unbedruckten Flächen auf dem DNA-Chip wurden diese mit Succinanhydrid behandelt. Dazu wurden 1,25 g Succinanhydrid in 250 ml 1,2-Dichlorethan gelöst und zu dieser Lösung 3,1 ml 1-Methylimidazol gegeben und gemischt. Die Mikroarray-Slides wurden in dieser Lösung 1 h bei Raumtemperatur unter Rühren inkubiert und zum Waschen in 250 ml frisches 1,2-Dichlorethan getaucht. Anschließend wurden die Arrays für 2 min in 95°C heißem H<sub>2</sub>O zur Denaturierung der DNA inkubiert, für 1 min in 95% Ethanol getaucht und durch Zentrifugation für 5 min bei 1500 rpm in 50 ml Reaktionsgefäßen getrocknet. Bis zur weiteren Verwendung wurden die DNA-Mikroarrays bei Raumtemperatur staubfrei in Objektträgerkästen aufbewahrt.

Beim Umgang mit den DNA-Mikroarrays wurde stets darauf geachtet, daß niemals mit den Fingern die Oberfläche der Slides berührt wurde, damit diese staub- und fettfrei bleibt.

### **11.2 Zellzucht für die Expressionsexperimente**

Ausgangspunkt für alle Experimente zur Transkriptomanalyse von *H. sal.* R1 ist eine Stammkultur, die bis zur mid-log Phase wachsen gelassen wurde. Zur Expressionsanalyse werden von dieser Stammkultur neue Kulturen angeimpft, und diese unter den jeweils zu untersuchenden Bedingungen wachsen gelassen. Erreichen die Kulturen unter den gewählten Bedingungen eine Zelldichte von 30-40 Klett-Einheiten, wird eine neue Kultur angeimpft und zur besseren Adaption der Zellen an die äußeren Bedingungen, wird vor der Probennahme diese Prozedur dreimal wiederholt.

Für jedes Mikroarrayexperiment wurde der Zyklus des dreimaligen Überimpfens unabhängig vier bis fünf mal wiederholt.

Neben den Experimenten zur Untersuchung der Expressionsunterschiede von adaptierten Zellen, wurden sogenannte „switch“-Experimente oder Wechselexperimente durchgeführt. Ziel dieser Experimente ist es, zeitaufgelöst die differentiell exprimierten Gene bei einem Wechsel von aerobem auf phototrophes Wachstum zu bestimmen.

Bei dieser Art der Experimente, wurden die Kulturen unter aeroben Bedingungen bis zur mid-log Wachstumsphase (30-40 Klett-Einheiten) wachsen gelassen und dann einem Wechsel auf phototrophes Wachstum ausgesetzt. Dazu wurden die Kolben mit Septen aus Silikon luftdicht verschlossen und in einen Schüttler unter einer Hochdruck-Xenonlampe bis zum Ende des Versuchs inkubiert. Die Probenahme erfolgte je Experiment einmal vor dem Setzen des Stimulus und dann anschließend nach bestimmten Zeitintervallen.

Die Zeitreihenexperimente wurden mit zwei verschiedenen Zeitintervallen durchgeführt. Beim ersten Zeitreihenexperiment erfolgte die Probenahme nach 3 min, 15 min, 45 min und 120 min. Die Probenahme beim zweiten Zeitreihenexperiment wurde nach 45 min, 90 min, 180 min, 360 min und 1440 min durchgeführt.

### **11.3 Herstellung der zu hybridisierenden DNA**

#### **11.3.1 Isolierung von Gesamt-RNA aus *Halobacterium salinarum***

Die Qualität der Ergebnisse eines Mikroarrayexperimentes ist in erster Linie von der Qualität der verwendeten RNA abhängig. Die RNA sollte idealerweise nicht mit DNA, Proteinen, Zuckern, Lipiden oder anderen Verbindungen kontaminiert sein. Ebenfalls ist zu beachten, dass RNA, im Vergleich zu DNA, relativ instabil ist und sie enzymatisch, chemisch oder physikalisch degradiert werden kann. Deshalb wurde bei der Isolierung und dem Umgang mit der Gesamt-RNA aus *H. sal. R1* bei jedem Schritt darauf geachtet, die Möglichkeiten einer Degradation so gering wie möglich zu halten.

Die Isolierung der Gesamt-RNA aus *H. sal. R1* erfolgte im Wesentlichen nach der Methode von Chomczynski und Sacchi, unter Verwendung der peqGOLD RNAPure Lösung von peqLab.

Für jedes Transkriptom-Experiment und alle Wiederholungen wurde so durchgeführt, dass zur Präparation der Gesamt-RNA aus *H. sal. R1* jeweils 10ml

der zu untersuchenden Kultur (Zelldichte 30-40 Klett-Einheiten) bei 8000 rpm (SS-34-Rotor) für 6min bei 4°C zentrifugiert wurden.

Das Zellpellet wurde mit 6 ml peqGOLD RNAPure Lösung versetzt und so lange gemischt, bis eine homogene Lösung entstanden ist. An dieser Stelle konnte die Präparation unterbrochen und die Proben bis zur weiteren Verwendung bei -20°C gelagert werden.

Nach dem Auftauen der Proben und Inkubation bei RT für 5 min, wurde das Zelllysat mit 1,2 ml Chloroform (0,2 Volumen der eingesetzten peqGOLD RNAPure Lösung) versetzt, für 15" gevortext und weitere 5 min bei Raumtemperatur inkubiert. Zur Phasentrennung wurde die Lösung für 5 min bei 10000 rpm (SS-34-Rotor) zentrifugiert und man erhielt eine untere gelbe phenolhaltige Phase, eine obere farblose wässrige Phase und eine dazwischenliegende Interphase. Die wässrige Phase, die ausschließlich RNA enthält, wurde vorsichtig abgenommen und auf drei frische Reaktionsgefäße verteilt. Die Präzipitation der RNA erfolgte durch Zugabe von 600 µl Isopropanol (0,5 Volumen der wässrigen Phase, je Reaktionsgefäß). Anschließend wurden die Proben 5-15 min bei RT inkubiert und für 10 min bei 12000xg und 4°C abzentrifugiert. Der Isopropanolüberstand wurde vorsichtig abgenommen und das RNA-Präzipitat mit 1 ml 70%-igem Ethanol versetzt und anschließend für 10 min bei 12000xg und 4°C zentrifugiert. Zum vollständigen Waschen der RNA wurde dieser Schritt wiederholt.

Nach der letzten Zentrifugation wurde der Überstand sehr sorgfältig entfernt und das RNA-Pellet für ca. 15 min an der Luft getrocknet. Zum Lösen der RNA wurde das Pellet mit 80 µl H<sub>2</sub>O (DEPC behandelt) versetzt und durch mehrmaliges pipettieren resuspendiert.

Um mögliche DNA-Kontaminationen zu beseitigen, wurden nach jeder RNA-Isolierung ein DNaseI-Abbau mit dem DNasefree Kit (Ambion) nach Herstellerangaben mit den gelieferten Reagenzien durchgeführt. Anschließend wurde die Konzentration und Unversehrtheit der RNA mittels des Agilent 2100 Bioanalyser und des RNA 6000 Nano LabChip® Kits nach Angaben des Herstellers bestimmt. Bis zur weiteren Verwendung wurden die RNA-Proben bei -80°C gelagert.

### 11.3.2 Synthese der fluoreszenzmarkierten cDNA

Bei allen Arbeitsschritten, bei denen mit Fluoreszenzfarbstoffen gearbeitet wurde, ist stets darauf geachtet worden, die Exposition der Reaktionsansätze mit Licht so gering wie möglich zu halten, um die Gefahr des Bleichens der Fluoreszenzfarbstoffe zu vermeiden, wodurch ein Informationsverlust der Mikroarray-Experimente entstehen würde. Deshalb wurden die cDNA-Synthesen und auch das Waschen der hybridisierten DNA-Mikroarrays in einem Dunkelraumlabor unter Rotlicht durchgeführt.

Für die Synthese der markierten cDNA wurde das „CyScribe First-Strand cDNA Labeling Kit“ von Amersham Biosciences verwendet. Die cDNA-Synthese wurde mit der „random“-Primer Methode (zufallsbedingten Primer Methode) von der Reversen-Transkriptase CyScript katalysiert, die in den cDNA-Einzelstrang die verwendeten fluoreszenzmarkierten Nukleotide, Cy3-dUTP, Cy5-dUTP einbaut. Als Ausgangsmaterial für die Synthese der fluoreszenzmarkierten cDNA wurde Gesamt-RNA von *H. sal.* R1 eingesetzt.

Im ersten Schritt, der Primer Anlagerung („Annealing“), wurden 5 µg Gesamt-RNA, 1 µl random Nonamer-Primer-Mix in einem Gesamtvolumen von 11 µl gemischt und 5 min bei 70°C inkubiert. Anschließend wurden die Reaktionsansätze 10 min auf RT abgekühlt. Inzwischen wurden für die unterschiedlichen cDNA-Ansätze folgende Reagenzien gemischt und anschließend zu den Reaktionsansätzen dazugegeben:

5x CyScript Puffer	4 µl
0,1 M DTT	2 µl
dUTP-Nukleotid Mix	1 µl
Cy3-dUTP / Cy5-dUTP	1 µl
CyScript Reverse Transcriptase	1 µl
	<hr/> <hr/>
	Σ 9 µl

Die Reaktionsansätze wurden für die cDNA-Synthese 1,5 h bei 42°C inkubiert und die Reaktion durch Zugabe von 2,5 µl 0,5 M EDTA-Lösung gestoppt. Nachfolgend wurde die RNA durch alkalische Hydrolyse durch Zugabe von 5 µl einer 1 M NaOH-Lösung und Inkubation für 1 h bei 65°C abgebaut. Die Proben wurden mit 12,5 µl 1 M Tris/HCl (pH 7,5) neutralisiert. Für die anschließende Reinigung wurden für die kompetitive Zweifarben-Hybridisierung die

entsprechenden Cy3- und Cy5-Ansätze vereint und mit 420 µl TE-Puffer und 1 µl Bierhefe tRNA ([4mg/ml]) versetzt. Durch eine Ultrafiltration mit MicroCon YM-30 Säulen (Milipore) wurden die Ansätze auf ein Endvolumen von ca. 17 µl gebracht. Zur Kontrolle der Markierungseffizienz der Fluoreszenzfarbstoffe wurden die Eluate (Durchläufe) der Ultrafiltration mit TE-Puffer auf 1 ml Endvolumen gebracht und mit einem Ultrospec 3000 UV/VIS-Spektrometer ein Spektrum von 450-700 nm aufgenommen. Anhand der Peaks (Cy3 550 nm / Cy5 650 nm) der nicht eingebauten Cy-Farbstoffe konnte dadurch indirekt auf die Einbaurate der Fluoreszenzfarbstoffe in den cDNA-Einzelstrang geschlossen werden.

TE-Puffer:

10 mM Tris	1,2 g
1 mM EDTA	0,37 g
ad 1 l H <sub>2</sub> O	
mit Salzsäure pH 8,0 einstellen, autoklavieren	

## 11.4 Hybridisierung

### 11.4.1 Prähybridisierung der DNA-Mikroarrays

Zur Prähybridisierung der Mikroarray-Slides wurden in einem Glasbehälter 250 ml Prähybridisierungs-Puffer 1 h bei 50°C in einem Hybridisierungssofen inkubiert. Nach der Temperierung (Vorinkubation) des Prähybridisierungs-Puffers wurden die Mikroarray-Slides in diesem Puffer 25 min bei 50°C inkubiert. Zum Waschen der prähybridisierten Slides wurden sie bei RT unter Rühren abwechselnd 30 sec in 250 ml H<sub>2</sub>O bidest. sowie 250 ml Isopropanol getaucht und unverzüglich durch Zentrifugation für 5 min bei 1500 rpm in 50 ml Reaktionsgefäßen getrocknet. Die so prähybridisierten DNA-Mikroarrays wurden sofort für die Hybridisierung verwendet.

Prähybridisierungs-Puffer:

3,5x SSC-Puffer	43,75 ml 20x SSC-Puffer
0,1% (v/v) SDS	2,5 ml 10%-igen SDS-Lösung
1% (w/v) BSA-Fraktion V	2,5 g
ad 250 ml H <sub>2</sub> O bidest.	

20x SSC-Puffer:

3 M NaCl	175,3 g
0,3 M Na <sub>3</sub> Citrat • 2H <sub>2</sub> O	88,2 g
ad 1 l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,0 einstellen, autoklavieren	

#### 11.4.2 Hybridisierung der fluoreszenzmarkierten cDNAs

Zur Hybridisierung wurden Hybridisierungsansätze mit einem Gesamtvolumen von 30 µl aus den fluoreszenzmarkierten und gereinigten cDNAs, 3,75 µl 20x SSC-Puffer und 1,25 µl 50x Denhardts-Lösung hergestellt. Die Hybridisierungsansätze wurden 2 min auf 98°C erhitzt und auf 45°C abgekühlt. Anschließend wurden zum Hybridisierungsansatz 2 µl einer 1%-igen SDS-Lösung pipetteirt, leicht gemischt und die Hybridisierungslösung in die Mitte des bedruckten Bereichs des prähybridisierten DNA-Mikroarray-Slides pipettiert, ohne dabei die Oberfläche des Slides zu berühren. Danach wurde auf dem bedruckten Bereich vorsichtig und möglichst luftblasenfrei ein Deckgläschen platziert, um zum Einen die Hybridisierungslösung durch Kapillarkräfte gleichmäßig auf dem Array zu verteilen und zum Anderen soll durch das Deckgläschen die Verdunstung der Lösung verhindert werden. Ein weiterer Vorteil der Hybridisierung mit einem Deckgläschen ist, dass für die Hybridisierung nur sehr wenig Volumen an Hybridisierungslösung eingestezt werden muss, wodurch die Konzentration der markierten cDNA erheblich vergrößert werden kann. Abschließend wurde jedes DNA-Mikroarray-Slide in eine Hybridisierungskammer von Corning gelegt, diese festverschlossen und sofort in einen auf 64°C temperierten Hybridisierungsöfen gebracht. Um zu verhindern, dass die Hybridisierungslösung unter dem Deckgläschen verdunstet, wurde zur Erhöhung der Luftfeuchtigkeit innerhalb der Hybridisierungskammer etwas H<sub>2</sub>O in die dafür vorgesehenen Vertiefungen pipettiert. Die Hybridisierung erfolgte üN (16 h) bei 64°C im Dunkeln.

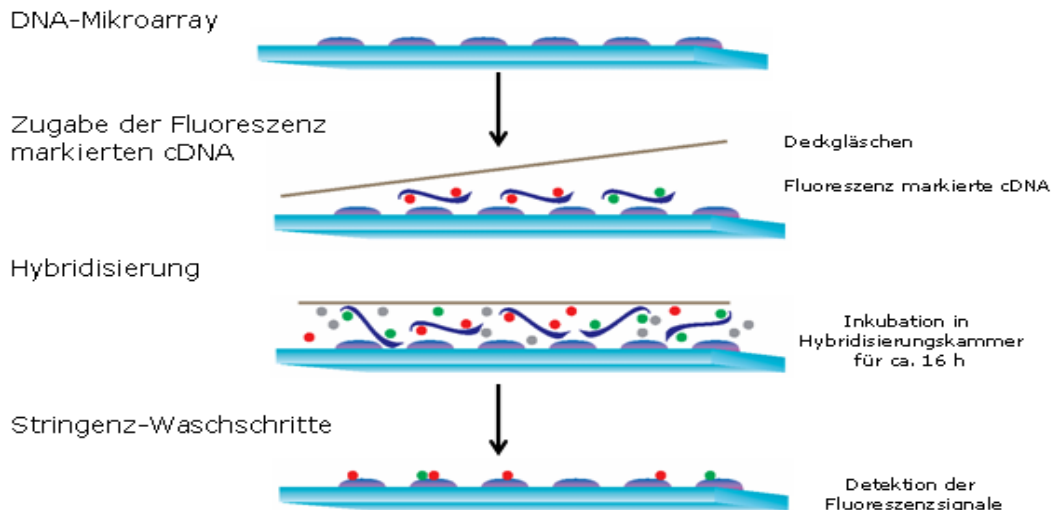


Abb. 87: Grafische Darstellung des Ablaufs einer Hybridisierung von Fluoreszenz markierter cDNA auf ein DNA-Mikroarray (nach Amersham Biosciences Handbuch). Beschreibung der Stringenz-Waschschritte siehe 5.4.3 Waschen der DNA Mikroarrays.

20x SSC-Puffer:

3 M NaCl	175,3 g
0,3 M Na <sub>3</sub> Citrat • 2H <sub>2</sub> O	88,2 g
ad 1 l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,0 einstellen, autoklavieren	

50x Denhardts-Lösung:

2% (w/v) Ficoll 400	10 g
2% (w/v) Polyvinylpyrrolidon	10 g
2% (w/v) BSA Fraktion V	10 g
ad 500 ml H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,0 einstellen, autoklavieren, Aliquots bei -20°C lagern	

### 11.4.3 Waschen der DNA Mikroarrays

Der Zweck der Waschschritte ist, die nicht gebundenen und nur locker anhaftenden markierten cDNAs (Targets) zu entfernen, um das sogenannte Hintergrundrauschen so klein wie möglich zu halten. Dabei ist darauf zu achten, daß die DNA-Mikroarray-Slides während des gesamten Waschvorgangs nicht trocknen. Deshalb wurden vorher je zwei Glasfärbekästen mit je 250 ml Waschlösungen I und II bereitgestellt. Nach der Hybridisierung wurden zum Waschen die Hybridisierungskammern auseinander gebaut und die Slides in die

Waschlösung I getaucht. Unter leichtem Schütteln wurden bei diesem Schritt die Deckgläschen von den Slides entfernt, denn die Deckgläschen sollten nicht von Hand vom Slide gelöst werden. Anschließend wurden die DNA-Mikroarrays unter Rühren jeweils zweimal 5 min in den Waschlösungen I und II gewaschen und sofort durch Zentrifugation für 5 min bei 1500 rpm in 50 ml Reaktionsgefäßen getrocknet. Bis zur Detektion, die direkt im Anschluss erfolgte, wurden die Mikroarrays bei Raumtemperatur in lichtgeschützten Objektträgerkästen gelagert, denn jede Exposition mit Licht führt zu einem Verlust der Signalintensität durch Bleichen der fluoreszenten Farbstoffe.

Waschlösung I:

0,5x SSC-Puffer	12,5 ml 20x SSC-Puffer
0,01% (v/v) SDS	500 µl 10%-iger SDS-Lösung
ad 500 ml H <sub>2</sub> O bidest.	

Waschlösung II:

0,06x SSC-Puffer	1,5 ml 20x SSC-Puffer
0,01% (v/v) SDS	500 µl 10%-iger SDS-Lösung
ad 500 ml H <sub>2</sub> O bidest.	

20x SSC-Puffer:

3 M NaCl	175,3 g
0,3 M Na <sub>3</sub> Citrat• 2H <sub>2</sub> O	88,2 g
ad 1 l H <sub>2</sub> O bidest.	
mit Salzsäure pH 7,0 einstellen, autoklavieren	

## 11.5 Bildanalyse und Rohdatenerhebung (Signal Detektion)

### 11.5.1 Bildanalyse

Die Quantifizierung der Fluoreszenzsignale der DNA-Mikroarray-Slides wurde mit dem GenePix 4000B Microarray Scanner der Firma Axon durchgeführt. Der Scanner erfasst simultan die Daten der zwei zur Anregung der Fluoreszenz



enthaltenen Laser (532 nm (17mW) und 635 nm (10mW)). Die Laserstrahlen regen die Fluoreszenz an einer bestimmten Position auf dem Array an und die Energie der Photonen des emittierten fluoreszenten Lichts wird in ein elektrisches Signal konvertiert, welches zu der Zahl der detektierten Photonen proportional ist. Das emittierte Licht geht dazu durch eine Anregungslinse, von wo es mit Hilfe einer Photomultiplieröhre detektiert wird. Der Photomultiplier verstärkt das analoge Signal von jedem Photon, welches anschließend in digitale Werte überführt wird. Das Helligkeitsmuster auf dem so entstehenden Digitalbild entspricht dann den Fluoreszenzintensitäten der hybridisierten cDNAs auf dem Array.

Die Signalausgabe der Photomultiplieröhre wird in 16-bit digitale Werte konvertiert, was bedeutet, dass die vorher kontinuierlichen analogen Signalintensitätswerte umgewandelt werden in diskrete Intensitätsniveaus, wobei eine digitale Auflösung von 16-bit  $2^{16} = 65536$  Niveaus entspricht.

Zur Detektion der Signale wurden die *H. sal.* R1 DNA-Mikroarrays mit der hybridisierten Fläche nach unten in den Scanner eingelegt und zunächst ein sogenannter „Prescan“ mit 40  $\mu\text{m}$  Auflösung durchgeführt. An Hand dieses Prescans wurde der auszuwertende Bereich auf dem DNA-Chip festgelegt und Laserstärke sowie Photomultiplerverstärkung optimiert. Dabei wurden die Einstellungen so gewählt, dass möglichst wenige Signale im Sättigungsbereich ( $2^{16}-1$ ) lagen und die Intensitäten in den beiden Laserbereichen ungefähr gleich waren. Mit diesen Einstellungen der Laserstärke und der Photomultiplerverstärkung wurden alle Slides, die zu einem Transkriptom-Versuch gehören, mit einer Auflösung von 10  $\mu\text{m}$  ausgelesen, damit die Vergleichbarkeit der DNA-Mikroarrays innerhalb eines Experimentes gegeben war.

Die detektierten Bilder wurden von dem Programm GenePix™ Pro 4.0 zunächst als Falschfarbenbilder wiedergegeben, wobei den Intensitätswerten für den Cy3-Farbstoff die Farbe Grün und den Werten für den Cy5-Farbstoff die Farbe Rot zugeordnet wurde. Computergestützt wurden die roten und grünen Bilder zu einem sog. Ratiobild übereinandergelegt. In diesem Bild entsprechen gelb erscheinende Punkte Genen, die unter den zu vergleichenden Bedingungen, dieselbe Menge an Transkript aufweisen, rot oder grün erscheinende Punkte dagegen Genen, die unter einer Bedingung mehr Transkript besitzen, als unter der anderen Bedingung.

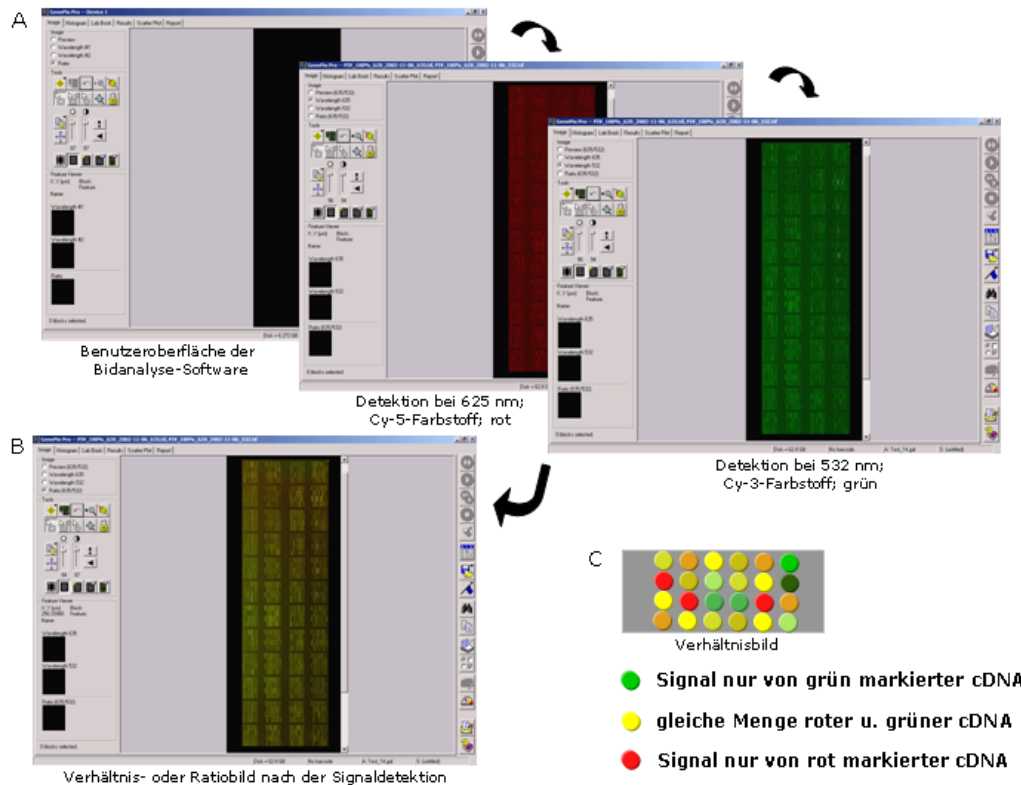


Abb. 88: Darstellung des Weges zum Erhalt eines Verhältnis- oder Ratiobildes bei der Bildanalyse mit GenePix™ Pro 4.0 anhand von Screenshots. A) Im Hintergrund ist die Benutzeroberfläche von GenePix™ Pro 4.0 dargestellt und davor Screenshots der Programmoberfläche mit den Ergebnissen der Cy5-(rot)- und Cy3-(grün)-Fluoreszenzdetektion. B) Screenshots des in silico generierten Verhältnisbildes aus den Ergebnissen der Cy5-(rot) und Cy3-(grün)-Signaldetektion. C) Schematische Darstellung eines Verhältnisbildes mit Erklärung zu den zu beobachtenden Signalen.

### 11.5.2 Rohdatenerhebung

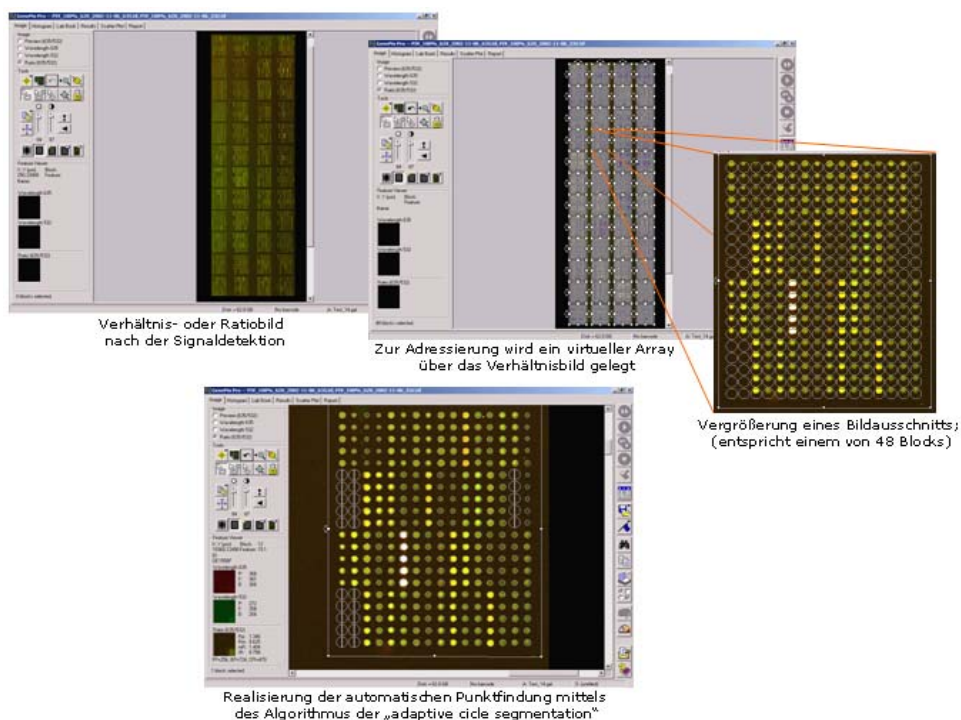
Zur Weiterverarbeitung der ausgelesenen Mikroarraybilder wurde mit Hilfe des Programms GenePix™ Pro 4.0 ein geometrisches Raster erstellt und über das Ratiobild gelegt. Dieser Prozeß kann in drei Teile eingeteilt werden:

- Rastererstellung („Addressing / Gridding“)
- Segmentierung/Aufteilung („Segmentation“)
- Intensitätsbestimmung („Intensity Extraction“)

Rastererstellung :

Bei diesem Prozess wurden den Punkten Koordinaten auf dem Array zugewiesen, d.h. es erfolgte eine Definition der Lage der Punkte im ausgelesenen Bild.

Das prinzipielle Layout, der Grundriss, des Arrays wurde durch das Spotten vorgegeben. Diese Informationen wurden genutzt, um eine sogenannte GAL-Datei („GenePixArrayList“) zu erstellen, mit deren Hilfe die Auswertesoftware (GenePix™ Pro 4.0) eine Array-Liste erstellt, was einem idealisierten Modell des Mikroarrays entspricht, wodurch die Zuordnung der Punkte zu den entsprechenden Gennamen ermöglicht wurde. Diese GAL-Datei wurde an Hand von Textdateien erstellt, die die Information der Konfiguration des Druckkopfes und der Belegung der Mikrotiterplatten beim Drucken beinhalten.



**Abb. 89:** Screenshots der Programmoberfläche GenePix™ Pro 4.0 zur Darstellung des Prozesses der Punktfindung („spot finding“). A) Im Hintergrund ist ein Verhältnisbild gezeigt und im Vordergrund ist ein Scan mit einem darübergelegten virtuellen Raster („GenePix Array List“) und einer Vergrößerung eines bestimmten Bereiches zu sehen. B) Screenshot des vergrößerten Bereichs aus A nach der Anwendung des

„adaptive circle segmentation“ Algorithmen der Software zur Punktfindung („spot finding“). Parallel zur Punktfindung findet die Adressierung jeder gefundenen Sonde statt.

### Aufteilung:

Die Segmentierung/Aufteilung, ermöglicht die Klassifizierung der Pixelwerte eines Punktes, dabei wurde das Bild durch eine sog. Punktmaske, die Teil des erstellten geometrischen Rasters ist, in verschiedene Teile eingeteilt und diesen Regionen unterschiedliche Merkmale wie Vordergrund oder Hintergrund zugewiesen. Diese Einteilung in Vorder- und Hintergrundwert, die nach der „adaptive circle“ Methode erfolgte, kann nach der automatischen Zuordnung durch das Programm manuell für einzelne Punkte auf dem Array angepaßt werden, wobei der Kreisdurchmesser des Punktes der Punktmaske variabel verändert werden kann.

### Intensitätsbestimmung:

Die Intensitätsbestimmung beinhaltet die Berechnung der Verhältnisse der roten und grünen Fluoreszenzintensitätswerte, sowie der Hintergrundwerte, dazu wurden die Pixelintensitäten innerhalb der Spotmaske summiert.

Die so ermittelten Hybridisierungsdaten wurden vom Programm in einer Ergebnistabelle zusammengefaßt. Ein Teil dieser Rohdaten wurde in einer neuen Ergebnistabelle zusammengestellt und für die anschließende statistische Auswertung als Tabulator getrennte Textdatei exportiert.

Zur Datensicherung wurden die erhaltenen Bilddaten auf eine CD-Rom und die ermittelten Rohdaten auf der Projektplatte „trans“ auf dem Server des MPI für Biochemie gespeichert.

## 11.6 Auswertung der Hybridisierungsdaten

Die Durchführung von Mikroarray-Experimenten und die Erhebung von numerischer Information ist nur der erste Schritt bei der Auswertung des Datenmaterials. DNA-Mikroarrays liefern gemäß ihres Hochdurchsatz-Anspruches extrem große Datenmengen. Insbesondere hier ist die Bioinformatik gefragt, zur Analyse geeignete Computerhilfsprogramme zur Verfügung zu stellen, die mit diesen großen Datenmengen umgehen können. Eine mögliche Lösung des Problems stellt die Benutzung von „R“ da, das große Datenmengen bewältigen kann und zur statistischen Mikroarray-Auswertung der Expressionsdaten von *H. sal.* R1 benutzt wurde.

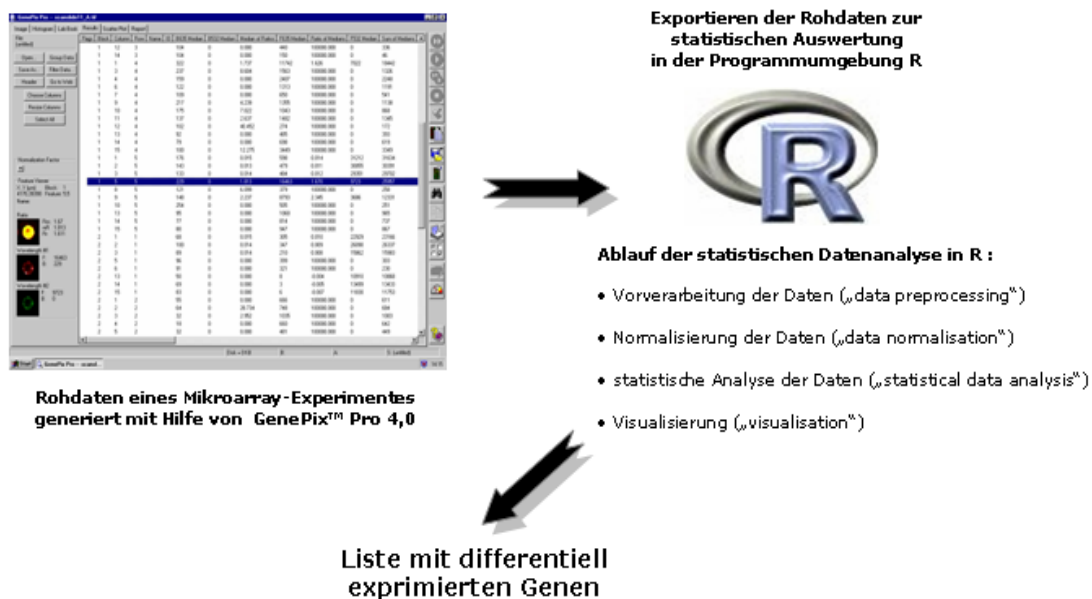


Abb. 90: Darstellung der Ablauforganisation der Datenanalyse eines Transkriptom-Experimentes. A Screenshot einer Ergebnistabelle nach der Bildanalyse. In Kombination mit den 16-bit Bild-Dateien der Cy5- u. Cy3-Signaldetektion stellen diese Tabellen die Rohwerte eines Mikroarray-Experimentes dar. B Die Inhalte der Ergebnistabellen werden als Textdateien exportiert und in die Programmumgebung R importiert. Die weitere statistische Analyse wird in R unter Verwendung des R/MAANOVA-Paketes und speziell entwickelter R-Funktionen durchgeführt (siehe 11.6.2).

### 11.6.1 Was ist R

R ist eine interpretierte Sprache und Entwicklungsumgebung für statistische (lineare und nicht-lineare Regression, klassische statistische Tests, Zeitreihen Analysen, Klassifizierung, Gruppierung, ...) Berechnungen und grafische Auswertetechniken. Es ist als „open source“ unter GNU GPL (General Public License)-Lizenz frei erhältlich und kann über das Internet unter [www.R-project.org](http://www.R-project.org) kostenlos bezogen werden. R ist der Sprache und Umgebung von S sehr ähnlich, das in den Bell Laboratories (früher AT&T, mittlerweile Lucent

Technologies) von John Chambers und Kollegen entwickelt wurde. Die kommerzielle Ausführung der Sprache S heißt S-PLUS und ist verfügbar über Insightful Corporation ([www.insightful.com](http://www.insightful.com)). Es gibt ein paar Unterschiede zwischen R und S, aber sehr viele der S-Befehle laufen in R ohne Änderung.

R wurde für Unix, Windows und Mac-OS von einem internationalen Entwicklerteam, dem sog. „R-Core“, entwickelt. Gestartet wurde das Projekt 1992 in der statistischen Abteilung der Universität von Auckland (Neuseeland) von Ross Ihaka und Robert Gentleman, auch bekannt als „R&R“.

### **11.6.2 Statistische Auswertung der Mikroarray-Daten von *H. sal.* R1**

Die statistische Auswertung erfolgte in Kooperation mit Dr. Gerhard Welzl (GSF-Forschungszentrum für Umwelt und Gesundheit, GmbH) und Dr. Jobst Landgrebe (Universität Göttingen). Für die Auswertung der *H.salinarum* Expressionsdaten wurden neben neuen R-Funktionen im Wesentlichen das R/MAANOVA-Paket benutzt ([www.jax.org/staff/churchill/labsite/software/anova/rmaanova/](http://www.jax.org/staff/churchill/labsite/software/anova/rmaanova/)). R/MAANOVA ist eine erweiterbare, interaktive R Umgebung für Mikroarray-Analysen und der Name MAANOVA steht für **MicroArray Analysis Of VA**riance. Der Ablauf der Analyse lässt sich in vier große Bereiche einteilen:

- Vorverarbeitung der Daten („Data Preprocessing“)
- Normalisierung der Daten („Data Normalization“)
- statistische Datenanalyse („Statistical Data Analysis“)
- Visualisierung der Daten („Visualization“)

### **11.6.3 Vorverarbeitung und Normalisierung der Daten**

Wie 11.4.2 beschrieben, wurde für jede Hybridisierung eine Datei erstellt, die nach der Analyse mit dem Programm GenePix™ Pro 4.0 in eine Textdatei überführt (exportiert) worden ist. Die Textdateien wiesen dabei immer folgenden Aufbau auf. In der Ersten Spalte stand der Genname (ID), dann Block, Spalte, Reihe (Angaben zur Lage auf dem Array), F532 Median, B532 Median, F635 Median, B635 Median und abschließend Flag. Dabei steht F für Feature und gibt

den Intensitätswert eines Punktes bei einer Wellenlänge an, B für Background / Hintergrundwert des Punktes und Flag ist ein Qualitätsmerkmal für den Punkt. Diese Tabulatorgetrennten Textdateien eines Experiments wurden für die Benutzung des R/MAANOVA-Pakets so zusammengefasst, dass in den ersten vier Spalten der Ergebnistabelle die allgemeinen Angaben über die Lage eines Gens auf dem Array stehen, wie ID, Block usw., in den nächsten fünf Spalten stehen die Angaben der Intensitätswerte und das Qualitätsmerkmal. Diese fünf Spalten wurden für jedes Experiment angehängt und das Resultat war eine Datei, bei der alle Wiederholungen eines Experiments enthalten waren. Diese zusammengefasste Datei, die der Ausgangspunkt der Analyse war, wurde als raw.data bezeichnet (siehe Abb. 30).

Der Ablauf der Datenanalyse wird im folgenden schematisch dargestellt. In den Abbildungen bedeuten jeweils Ellipsen, dass dort die Dateien enthalten sind, auf die eine Funktion, die in den Kästen stehen, angewandt wurde. Durch die Pfeile wird dargestellt, welche Datei mit welcher Funktion bearbeitet wurde.

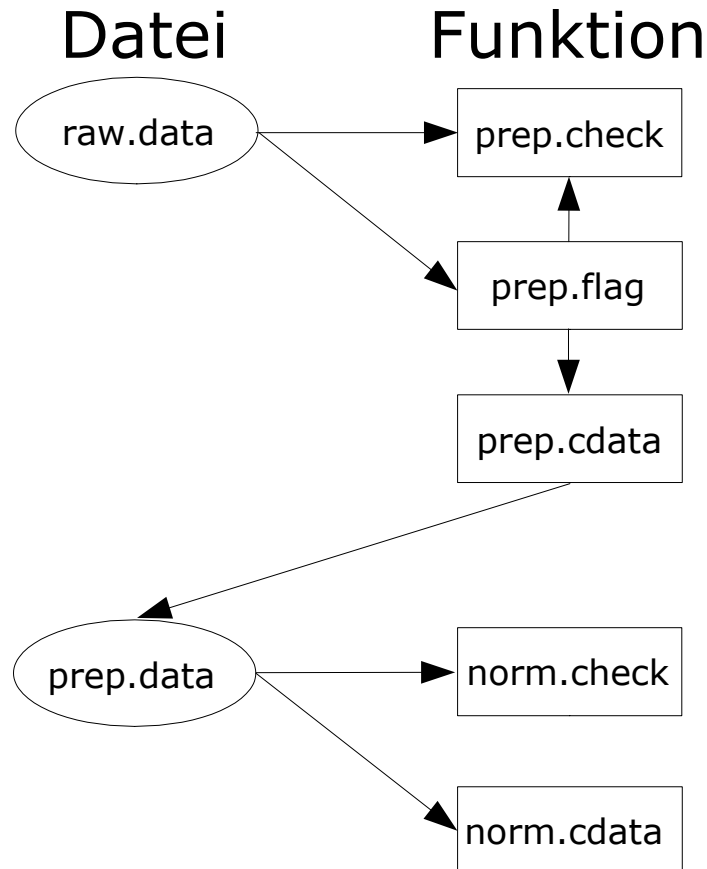


Abb. 91: Schematische Darstellung des Ablaufs der Vorverarbeitung und Normalisierung in R.

Beim Preprocessing wurde auf die Datei `raw.data` zum Einen die Funktion `prep.check` und zum Anderen die Funktion `prep.flag` angewandt, und anschließend die Funktion `prep.cdata`. Beim Durchlaufen dieser Schritte wurde die Qualität der Daten überprüft was z.B. den Ausschluss von Punkten mit niedrigen Intensitätswerten, die Hintergrundkorrektur und das Filtern der Daten mit einem Flag beinhaltet.

Das Ergebnis dieses Preprocessings ist die Datei `prep.data`, die jetzt Hintergrundkorrigierte Werte enthält und bei der Gene markiert wurden, deren Intensitätswerte nicht von den Hintergrundwerten zu unterscheiden waren. Der nächste sehr wichtige Schritt bei der Auswertung ist die Normalisierung. Durch die Normalisierung sollen systematische Effekte, die nicht auf Grund des Experiments entstanden sind, erkannt und entfernt werden. Zu diesen systematischen Effekten zählen unter Anderem räumliche Unterschiede der Intensitäten auf einem Array oder Unterschiede zwischen verschiedenen Arrays aufgrund verschiedener Hybridisierungseffizienz oder hoher Hintergrundwerte. Bei der Normalisierung wurden die Intensitätswerte als erstes in logarithmische



Werte transformiert, wodurch die Varianz in den Werten stabilisiert wurde und dann das Ergebnis mit der Funktion `norm.check` durch die Erstellung verschiedener Grafiken überprüft. Mit Hilfe der Funktion `norm.cdata` wurde dann die Normalisierung der Daten durchgeführt, so dass alle Ergebnisse der verschiedenen Arrays untereinander vergleichbar waren. Als Resultat der Normalisierung erhielt man die Datei `norm.data`, die die Grundlage der statistischen Auswertung ist.

#### 11.6.4 Statistische Datenanalyse

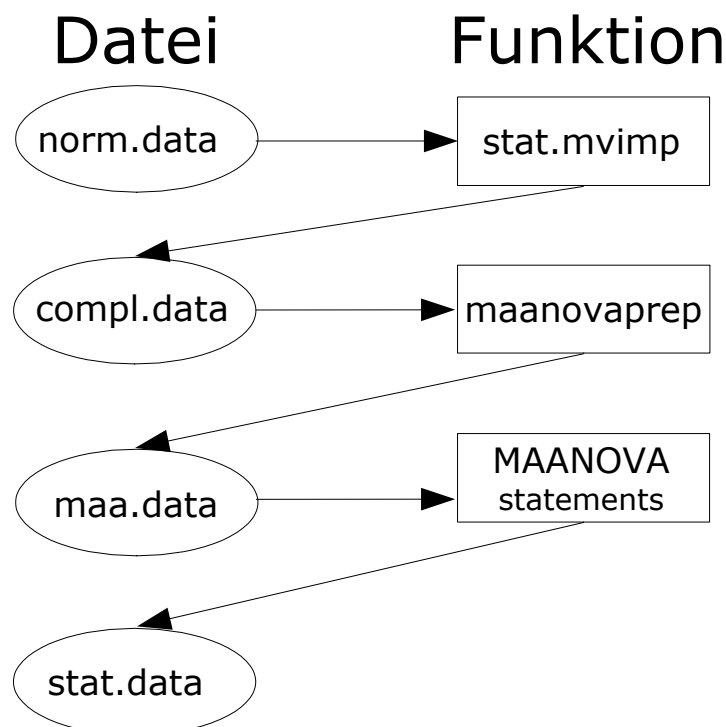


Abb. 92: Schematische Darstellung der statistischen Datenanalyse in R.

Zur statistischen Auswertung der Daten mit R/MAANOVA benötigt man einen vollständigen Datensatz. Das bedeutet, dass für jedes Gen und jede Wiederholung ein Wert in der Datenmatrix enthalten sein muss. In der Datenmatrix `norm.data` gibt es Gene, für die nicht überall ein Eintrag vorhanden ist, sei es durch Markierung im Preprocessing Schritt oder durch eine fehlende Sonde auf dem Chip. Mit der Funktion `stat.mvimp` wurden diese fehlenden Werte ersetzt und dabei die Datei `compl.data` erstellt, die nun einen vollständigen Datensatz darstellt, mit dem weiter gearbeitet werden kann. Diese Datei `compl.data` wurde durch die Funktion `maanovaprep` so umgeformt, dass als Ergebnis die Datei `maa.data` entstanden ist, mit der es jetzt möglich war eine „mixed effect“ ANOVA mit der Funktion `MAANOVA` durchzuführen. Anschließend

erhielt man eine Datei mit dem Namen `stat.data`, die alle statistischen Kennwerte enthält.

### 11.6.5 Visualisierung der Daten

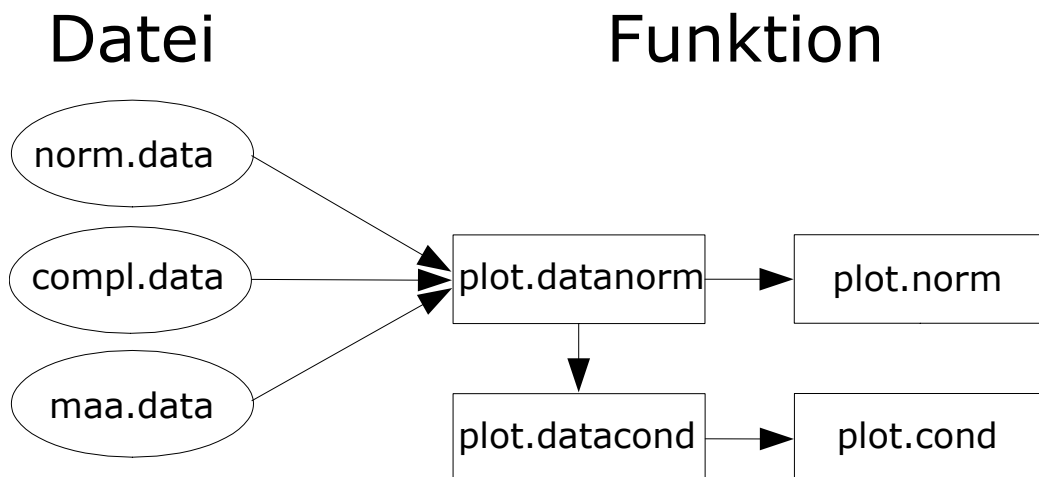


Abb. 93: Schematische Darstellung der Visualisierung in R.

Als abschließenden Punkt der statistischen Auswertung kann man die grafische Darstellung der Ergebnisse ansehen. Dazu wurde mit Hilfe der Funktion `plot.datanorm` eine Datei erzeugt, die eine Kombination aus den drei Dateien `norm.data`, `compl.data` und `maa.data` darstellt. Diese Datei ist Ausgangspunkt für die grafische Ausgabe, die mit der Funktion `plot.norm` erfolgen kann.

Die Datei ist allerdings nicht nur Ausgangspunkt für die grafische Ausgabe, sie kann auch benutzt werden, um eine Datenreduktion zu erzeugen, die notwendig ist für weitere statistische Methoden, deren Ergebnisse mit der Funktion `plot.cond` grafisch dargestellt werden können. Alle Schritte der Datenanalyse in dem Programm R sind durch die Dokumentation festgehalten und können in einer History aufgerufen und dadurch nachvollzogen werden. Die Auswertungen sind auf dem Server im MPI für Biochemie gespeichert und gesichert.



## 12. Material

### 12.1 Biologisches Material

#### 12.1.1 Bakterienstämme

DH5 $\alpha$	<i>supE44, _lac169,(_80lacZ, _M15), hsdR17,recA1, endA1, gyrA96, thi-1,relA1</i>
<i>Halobacterium salinarum</i> R1	Stoeckenius und Kunau, 1968

#### 12.1.2 Antikörper

Anti-Digoxigenin-Alkalische-Phosphatase F<sub>ab</sub>- Roche Diagnostics, Mannheim  
Fragment

#### 12.1.3 Enzyme und Proteine

BSA Fraktion V	Serva, Heidelberg
DNA-Poymerasen	
TaKaRa LaTaq	BioWhittacker, Rockland, USA
Hot-Star Taq DNA-Polymerase	Qiagen, Hilden
Phusion DNA-Polymerase	Biometra, Göttingen
Proteinase K	Roche Diagnostics, Mannheim
Restriktionsendonukleasen	New England Biolabs, Schwalbach
Reversetranskriptasen	
M-MLV Reverse Transkriptase RnaseH(-) Point Mutant	Promega, Mannheim
Superscript III	Invitrogen, Karlsruhe
Ribonuklease Inhibitor	MBI-Fermentas, St. Leon-Rot

T4-DNA-Ligase	New England Biolabs, Schwalbach
T7-RNA-Polymerase	MBI Fermentas, St. Leon-Rot
Lysozym	Sigma, Deisenhofen

### 12.1.4 Standardgrößenmarker für Proteine und Nukleinsäuren

Referenzproteine für SDS-Page  
(Protein-Längenstandards)

DNA-Längenstandards

1 kb DNA-Leiter	Gibco BRL Life Technologies, Karlsruhe
25 bp DNA-Leiter	Invitrogen, Karlsruhe
100 bp DNA-Leiter	MBI Fermentas, St. Leon-Rot
pBR322 DNA/AluI	MBI Fermentas, St. Leon-Rot
pUC19 DNA/Mspl	MBI Fermentas, St. Leon-Rot
DNA-Molecularweight-Marker VII DIG markiert	Roche Diagnostics, Mannheim

RNA-Längenstandards

RNA 6000 Leiter	Ambion, Huntingdon
RNA-Ladder High range	MBI Fermentas, St. Leon-Rot
RNA-Molecularweight Marker I DIG markiert	Roche Diagnostics, Mannheim

### 12.1.5 Synthetische Oligonukleotide

Alle Oligonukleotide, die für dieses Projekt benutzt wurden, sind von der Firma Metabion (Martinsried) synthetisiert worden.

## Primer für die quantitative RT-PCR:

OE-Nummer	Gen	Funktions- klasse	Forward Primer	Reverse Primer
OE1710R	korB	CIM	5' - AACCCGGACCTCGAAGTGAT - 3'	5' - CCGTAGATGCGGTTGTCCAT - 3'
OE1808F	crtI3	LIP	5' - AAACCCCTTCGCGGACATCT - 3'	5' - TACTCGAAGACGCGTTTGGAC - 3'
OE1956F	nuoCD	EM	5' - CGACCACCTGTCTGTGTCA - 3'	5' - GCACGATGATCGAGACCTCC - 3'
OE2225F	dmsB	EM	5' - AGAACAAACGTCGCCGATGGG - 3'	5' - TCGGCTGGTAGGTCATTTTCG - 3'
OE2868R	sdhC	CIM	5' - GTTCACCCACATCTCGGTCC - 3'	5' - GGCCAGACTCTCCAATCCAC - 3'
OE3073R	-	MIS		
OE3093R	crtB1	LIP	5' - GAGGAGGTCACCACGTTTCATC - 3'	5' - GACGACCCACGGAGGTACTC - 3'
OE3100F	blp	MIS	5' - TCGGCACCTGCGAAGACATC - 3'	5' - CGGAGACCCGGTACGTGAT - 3'
OE3101R	bat	MIS	5' - CGGAAGAGCCGATCATCTACAT - 3'	5' - CATCTGGAACCCGACGTAGTGT - 3'
OE3102R	brp	MIS	5' - GGGCTCTACCTGTTGGGTATG - 3'	5' - GCAGGTTCTCGAAATGCTCG - 3'
OE3106F	bop	EM	5' - TCGGACCCAGATGCAAAGAA - 3'	5' - ACGGTACCATTGTGAGGCCAT - 3'
OE3107F	OE3107F	HY	5' - TTTCCAGCGTCGCTCAACA - 3'	5' - CAAAACGTCATGCCGAACG - 3'
OE3136F	OE3136F	CHY	5' - CCGATGAATGTGATCTCCCGT - 3'	5' - TGTACTCGCTGACGTGTTTCCC - 3'
OE3376F	crtB2	LIP	5' - TACGTCTCTACGCGTTCTTCC - 3'	5' - TGAACGTGTGGATGTCCCTCGT - 3'
OE3381R	crtI1	LIP	5' - TGGAGAAAACCGACCAGATCG - 3'	5' - TTGTCTTTGAAGAAGATGCGG - 3'
OE3468R	crtI2	LIP	5' - ACTTCTACGAGCTCACC CGC - 3'	5' - AGGACTCGAACAGCTCGTGG - 3'
OE3554F	carB	AA	5' - GGTGGCCAAACGACAAGTTC - 3'	5' - TCAGCAGGCTCTCCTCGAAG - 3'
OE3556R	carA	AA	5' - GCGACACCGAGAAAATGGACT - 3'	5' - ATAGCCGTGTTCTGTGTGGTC - 3'
OE3980R	blh	CHY	5' - GCCGAAAACCTCCTGTTGT - 3'	5' - CGAGTGCCAGACGAGAAG - 3'
OE3983R	crtY	LIP	5' - CTCACGTTCTGGCTGTGCG - 3'	5' - TCGTGTATGACAGCCCAAC - 3'
OE4217R	fer1	EM	5' - ATGGACATGCAGCAGATCCTC - 3'	5' - TGGCGTTGTAGACGATCTTGAC - 3'
OE4496R	fer4	EM	5' - AGCGAAGTAGACACCCACAAG - 3'	5' - AGATGTCGAAGTCAACCGCCAC - 3'
OE4612F	hly	CP	5' - GTGAGGAAACGGCTCAACA - 3'	5' - GATTGTGTCACGCGATTCGA - 3'
OE5205R	arcB	AA	5' - TTCGAGCAGCAGAGATGATG - 3'	5' - ATCGTGAACATGTCGGTGAGC - 3'
OE5206R	arcC	AA	5' - TCGACATCGTTGAGGCAGAA - 3'	5' - AGCACGATCCTGTGCGATGAC - 3'
OE5208R	arcA	AA	5' - AGGCCATCACGAACTTCCTGA - 3'	5' - ATCGTAGTCTCCTGGATCCCA - 3'
OE5209R	arcR	REG	5' - CAAGTCGGAATCGTCGAACAA - 3'	5' - TCCGAGCCGGTAGAGATCAGTA - 3'
OE7065F	cydA1	EM	5' - ACGTACGGACTAACGAGCAGC - 3'	5' - GGAAGCTCATTGGGATACCG - 3'

## Primer zur Herstellung der Deletionsmutanten:

R1Δ3107F:

US-OE3107-1.for

5'-GCG CGG ATC CGC TTC GTG TGG TGG GCG AT-3'

US-OE3107-1.rev

5'-CGG GCC CTG CGT GTG TCG TTG AAA AGC CGC G-3'

DS-OE3107-1.for

5'-CAA CGA CAC ACG CAG GGC CCG CTG GCG AGC C-3'

DS-OE3107-1.rev

5'-CGC GTC TAG AGC GTC CTT GCA CTT CGA CAG CAT-3'

R1Δ3136F:

USdel3136-1.for

5'-GCG CGG ATC CCG ACG CCG ACG CGG TCC CCG-3'

Usdel3136-1.rev

5'-GCC CGT GAG TGT ATC TCA TCG GAA TAC CTA CCC G-3'

Dsdel3136-1.for

5'-TGA GAT ACA CTC ACG GGC CAG GTT CTCCGA G-3'

Dsdel3136-1.rev

5'-CGC GTC TAG ACA GCC CTG ATC ACT GCG TTC -3'

Primer für in vitro Expression (RTS-100 System):

C-terminale AviTag DNA:

Avi-C-3136-for

5'-actttaagaaggagatataccATGAGTCAGTCGCAGCTGGC-3'

Avi-C-3136-rev

5'-gaagatgtcggtcaggccccGGGTCCGCAAACGGCTGTGCG-3'

N-terminale AviTag DNA:

Avi-N-3136-for

5'-cgcttaattaaacatagaccAGTCAGTCGCAGCTGGCAGC-3'

Avi-N-3136-rev

5'-ttagttagttaccggatcccGGGTCCGCAAACGGCTGGTG-3'

Standardprimer:

Random Hexamer Primer

Promega, Mannheim

## 12.2 Chemikalien

Agarose (SeaKem® LE)

BioWhittaker, Rockland, USA

Ampicillin

Roth, Karlsruhe

$\alpha$ -<sup>32</sup>P dATP

Amersham Biosciences, Freiburg

Bacto™ Agar

Difco Laboratories, Detroit, USA

Bacto™ Hefeextrakt

Difco Laboratories, Detroit, USA

Bacto™ Trypton

Difco Laboratories, Detroit, USA

Bacto™ Pepton

Difco Laboratories, Detroit, USA

Bäckerhefe t-RNA

Roche Diagnostics, Mannheim

Blocking Reagent

Roche Diagnostics, Mannheim

CDP-Star detektion reagent

Amersham Biosciences, Freiburg

Desoxynukleosidtriphosphate	Peqlab, Erlangen
1,2-Dichlorethan	Merck, Darmstadt
Diethyl-Pyroc carbonat (DEPC)	Sigma, Deisenhofen
DIG Easy Hyb	Roche Diagnostics, Mannheim
Digoxigenin-11-UTP	Roche Diagnostics, Mannheim
Dimethylsulfoxid (DMSO)	Merck, Darmstadt
Dinatriumhydrogenphosphat	Merck, Darmstadt
Ficoll 400	Sigma, Deisenhofen
Isopropanol	Merck, Darmstadt
LUMI-Light Reagenz	Roche Diagnostics, Mannheim
MOPS	Biomol, Hamburg
1-Methylimidazol	Fluka, ?
peqGOLD RNAPure	Peqlab, Erlangen
RNaseZap	Ambion, Huntingdon
ROTI PhenolChloroform	Roth, Karlsruhe
D(+)-Saccharose	Roth, Karlsruhe
Siedesalz (NaCl)	Staatliche Saline, Bad Reichenhall
Succinanhydrid	Sigma, Deisenhofen
SYBR-Green II RNA-Gel Stain	MoBiTec, Göttingen

### 12.3 Kits

ABI Prism Big Dye™ Terminator Cycle Sequencing Ready Reaction	Applied Biosystems, Foster City, USA
Cy Scribe First-strand cDNA labelling	Amersham Biosciences, Freiburg
DECAprime II-Nothern Max	Ambion, Huntingdon
DNAfree	Ambion, Huntingdon
DNA 1000 LabChip® Kit	Agilent, Karlsruhe
DNA 7500 LabChip® Kit	Agilent, Karlsruhe



RNA 6000 Nano LabChip® Kit	Agilent, Karlsruhe
NucleoSpin RNA L	Macheray&Nagel, Düren
QIAprep Spin Plasmidextraction Kit	Qiagen, Hilden
QIAquick PCR-Purification Kit	Qiagen, Hilden
RNA 6000 Nano LabChip® Kit	Agilent, Karlsruhe
RNase-free DNase Set	Qiagen, Hilden
RNeasy (Mini/Midi) Kit	Qiagen, Hilden
TOPO TA Cloning® Kit	Invitrogen, Karlsruhe
TURBO DNafree	Ambion, Huntingdon

## 12.4 Sonstige Materialien

CMT-GAPS II Coated Slides (Barcode)	Corning, Schiphol, NL
Färbekästen (Glas) mit Einsatz	Roth, Karlsruhe
Genomic-tip G500/G-Säulen	Qiagen, Hilden
Hybond-N Membran	Amersham Biosciences, Freiburg
Hybri-Slip HS	Sigma, Deisenhofen
Hybridisierungskammern	Corning, Schiphol, NL
Immobilon™-P-Transfermembran	Millipore, Eschborn
MicroCon YM 30	Millipore, Eschborn
3MM Chr Chromatography Paper	Whatman, Maidstone, UK
MultiScreen 96 PCR	Millipore, Eschborn
Novex® 4-20% TBE Gel	Invitrogen, Karlsruhe
Novex® High Density TBE Sample Buffer	Invitrogen, Karlsruhe
Poly-Prep-L-Lysin Slides	Sigma, Deisenhofen

## 12.5 Geräte

Analysenwaage Mettler Toledo HL52	Mettler-Toledo GmbH, Giessen
-----------------------------------	------------------------------

Agarosegel-Elektrophoresekamern	Zentralwerkstatt, Max-Planck-Institut für Biochemie
Agilent Bioanalyser 2100	Agilent, Karlsruhe
Bio-Print Photodocumentation mit BIO-CAPT V.97 Software	Vilber Lourmat, Marne-La-Vallee, FR
ChipWriter Pro™	Virtek, Waterloo, Ontario, CA
Entwicklermaschine X-OMAT M 35	Kodak, Stuttgart
FIREBOY plus	INTEGRA Biosciences, Fernwald
GenePix 4000B	Axon, Foster City, USA
GeneAmp PCR-System 9700	Perkin Elmer, Shelton, USA
Gilson Pipetten P2, P10, P20, P100, P200, P1000	Abimed, Langenfeld
Hybridisierungsöfen Biometra OV5	Biometra, Göttingen
Klett-Summerson Photoelectric Colorimeter	Klett Manufacturing CO., INC., New York, USA
Laborwaage Mettler Toledo PB1502	Mettler-Toledo GmbH, Giessen
Lambda 2 UV/VIS-Spektrometer	Perkin Elmer, Überlingen
Mini-Wipptisch WT 16	Biometra, Göttingen
Mikrowelle NN-5556/5506	Panasonic, Hamburg
Multi Cycler PTC-200 MJ Research	Biozym, Hessisch Oldendorf
MultiScreen Vakuum Manifold	Millipore, Eschborn
Optima LE80-K Ultrazentrifuge	Beckmann, Unterschleißheim
Power Supply ECPS	Amersham Biosciences, Freiburg
Reinstwasseranlage MilliQ <sub>PLUS</sub>	Millipore, Eschborn
Sorvall RC 5C Plus Zentrifuge	Beckmann, Unterschleißheim
Speed Vac™	Bachhofer, Reutlingen
Thermoblock BT3	Grant/Boekel, Cambridge, UK
Tischzentrifuge Centrifuge 5415C	Eppendorf, Hamburg
Tischzentrifuge Centrifuge 5417R	Eppendorf, Hamburg
TL-100 Ultrazentrifuge	Beckmann, Unterschleißheim

Ultrospec 3000 UV/VIS-Spektrometer	Amersham Biosciences, Freiburg
UV-Stratalinker™ 2400	Stratagene, Amsterdam, NL
Video Copy Processor	Mitsubishi, Ratingen
Vortex	Bender&Hobein, Zürich, CH
XCell <i>SureLock</i> ™ Mini-Cell mit Blot Module Kit	Invitrogen, Karlsruhe

## **12.6 Computerprogramme, Datenbanken und Internet-Seiten**

GeneAmp 5700 SDS Software	Applied Biosystems, Foster City, USA
GenePixPro 4.0	Axon Instruments, Union City, USA
PrimeArray	
Primer Express 2.0	Applied Biosystems, Foster City, USA

### 13. Zusammenfassung

Basierend auf der Genomsequenz von *H. sal.* R1 wurde ein genspezifischer Gesamt-Genom-DNA-Mikroarray konstruiert. Hierzu wurden für jeden ORF des Genoms ORF-spezifische Oligonukleotide abgeleitet und zur spezifischen Amplifizierung der Genabschnitte mittels PCR eingesetzt. Nach Amplifizierung und Reinigung der Genabschnitte deckten die Produkte über 97% des halobakteriellen Genoms ab. Zur Konstruktion des Gesamt-Genom-DNA-Mikroarrays wurde jede spezifische Gensonde in fünffacher Wiederholung auf den DNA-Array aufgebracht. Auf diese Weise wurde ein DNA-Mikroarray erstellt, der mit einer Gesamtzahl von 13545 genspezifischen Sonden, das bisher dichteste Raster eines archaealen DNA-Mikroarrays aufweist.

Durch parallele genomweite Genexpressionsanalyse in *H. sal.* R1, wurde der Vergleich zwischen aerobem und phototrophem Wachstum in drei umfassenden DNA-Mikroarrayexperimenten gezogen. Die Mikroarrayexperimente wurden mit dem so genannten „common reference“ Experimentdesign durchgeführt, bei dem eine Mischung aller RNA-Proben eines Experiments als Referenz bei den Hybridisierungen dient. Als weitere Voraussetzung zur späteren statistischen Datenanalyse wurden die Transkriptomexperimente alle vier- bis fünfmal in unabhängigen Experimenten wiederholt. Die Wahl der Referenz und die Anzahl der unabhängigen biologischen Replikate haben die Basis geschaffen, die erhobenen Expressionsdaten mit Hilfe des R/MAANOVA Paktes der flexiblen und leistungsstarken statistischen Programmoberfläche R auszuwerten.

Eine leistungsstarke und flexible Programmoberfläche zur Datenanalyse war unerlässlich, denn mit steigender Komplexität eines Transkriptomexperiments, steigt auch die Anzahl der notwendigen Wiederholungen und damit einhergehend die Gesamtzahl der auszuwertenden Datenpunkte. Für ein durchgeführtes Zeitreihenexperiment vom Wechsel aerobes zu phototrophem Wachstum mit sechs Zeitpunkten, fallen ca. 384.000 Datenpunkte an, für deren Vorder- und Hintergrundwerte die statistischen Kennwerte berechnet werden mussten.

Ein solcher statistischer Kennwert ist der so genannte p-Wert, der die Signifikanz eines Ergebnisses widerspiegelt. Auf der Basis dieser signifikanten p-Werte ist

eine Liste von 242 Kandidatengenomen erstellt worden, die als differentiell exprimiert angesehen werden.

Ein Anteil von 54.5% dieser differentiell exprimierten Gene weist kein homologes Protein oder Funktion auf. Diese Tatsache, birgt die Chance sowohl die Existenz dieser ORFs, als auch ihre Funktion aufzuklären.

In diesem Zusammenhang wurden für die hypothetischen ORFs OE3107F und OE3136F Deletionsmutanten hergestellt und näher charakterisiert. Dabei wurde festgestellt, dass die Deletionsmutanten R1 $\Delta$ 3107 und R1 $\Delta$ 3136 im Vergleich zum WT-Stamm *H. sal.* R1 deutliche Unterschiede in ihrer Pigmentzusammensetzung aufweisen. Beide Deletionsstämme weisen z.B. einen geringeren Gehalt an Bakteriorhodopsin auf.

Somit hat die neu etablierte Methode der DNA-Mikroarray basierten Genexpressionsanalyse dazu beigetragen, zwei bisher unbekannte Kandidaten der Regulation der Expression des Bakteriorhodopsins in *H. sal.* R1 zu identifizieren. Durch zellfreie *in vitro* Expression des Gens OE3136F wurde ein möglicher Ansatz zur näheren Charakterisierung und Funktionsaufklärung aufgezeigt.

Neben der Herstellung von Deletionsmutanten und deren Charakterisierung, wurde durch die Anwendung einer weiteren Datenanalyse mittels PCA (Hauptkomponentenanalyse) und dem Ansatz die erhobenen Transkriptomdaten auf Stoffwechselwegen abzubilden, zwei weitere denkbare Wege aufgezeigt, aus den ermittelten Expressionsdaten mehr Informationen zu erhalten.

Die Ergebnisse aller Transkriptomexperimente für *H. sal.* R1 stimmen mit den Ergebnissen früherer Arbeiten überein und durch unabhängige Methoden wie RT-PCR, Northern-Blot-Analysen und Proteomvergleich konnten die Resultate der Expressionsanalysen eindeutig verifiziert werden.

Die Konstruktion und Herstellung der *H. sal.* R1 Gesamtgenom-Mikroarrays und Ausarbeitung eines Standardprotokolls zur Versuchsdurchführung, bilden die

Grundlage aller Transkriptomexperimente der Arbeitsgruppe. Daneben ermöglicht die Schaffung einer bioinformatischen Infrastruktur zur statistisch signifikanten Auswertung der DNA-Mikroarray-Hybridisierungsergebnisse die Erstellung einer Transkriptomdatenbank, die durch Anknüpfung an die bereits vorhandene HaloLex-Datenbank jedem Nutzer für weitere Mikroarray-Experimente mit anderer Fragestellung in leichter Form zur Verfügung steht.

Abschließend kann gesagt werden, dass die DNA-Mikroarray basierende Transkriptomanalyse von *H. sal.* R1 dazu beigetragen hat das Wissen über den Prozess der Anpassung an das phototrophe Wachstum zu erweitern.

Die in der Arbeit erhobenen Daten bilden die Grundlage einer Datensammlung, die es zu einem späteren Zeitpunkt ermöglichen wird, über viele verschiedene Experimente hinweg neue Co-Regulationen von Genen zu erfassen und damit neue Gene und Verknüpfungen zwischen Stoffwechselwegen schnell und einfach zu detektieren.

Die vorliegende Arbeit kann als Ausgangspunkt für genomweite funktionelle Charakterisierung haloarchaealer Genexpression und ihrer Regulation angesehen werden. Dieser Punkt ist im Hinblick auf die wachsende Bedeutung der Systembiologie von entscheidender Wichtigkeit, denn nur auf der Basis von soliden experimentellen Ergebnissen können Modelle aufgestellt und verbessert werden.



## 14. Abkürzungsverzeichnis

ANOVA	Analysis of Variance
APS	Ammoniumpersulfat
AS	Aminosäuren
BCA	Bicinchoninsäure
BLAST	Basic Local Alignment Search Tool
BR	Bakteriorhodopsin
BRE	Transcription Factor B Recognition Element
CMT-GAPS	Corning Microarray Technologie $\gamma$ -Aminosilan
CTAB	N-Cetyl-N,N,N-trimethylammoniumbromid
DEPC	Diethylpyrocarbonat
DIG	Digoxigenin
DMSO	Dimethylsulfoxid
ds	Doppelstrang
EDTA	Ethylendiammintetraessigsäure
FDR	False Discovery Rate
FWER	Family-wise Error Rate
HTH	Helix-Turn-Helix
kbp	Kilobasenpaare
kDa	Kilodalton
LOWESS	Locally Weighted Scatterplot Smoothing
MIAME	Minimum Information about a Microarray Experiment
OD	optische Dichte bei bestimmter Wellenlänge
ORF	Open Reading Frame
PA	Polyacrylamid
PAGE	Polyacrylamid-Gelelektrophorese
PCR	Polymerase Kettenreaktion



PCA	Principal Component Analysis
pH	umgekehrter dekadischer Logarithmus der Protonenkonzentration
rpm	Rounds Per Minute
RT	Raumtemperatur oder Reverse Transkriptase (je nach Kontext)
RTS	Rapid Translation System
SDS	Sodiumdodecylsulfat
SVP	Singular Value Decomposition
TEMED	N,N,N,N-Tetramethylamin-N-oxid
ün	über Nacht

## 15. Literatur

- Albers, S.V., Koning, S.M., Konings, W.N. and Driessen, A.J.M. (2004) Insights into ABC transport in archaea. *Journal of Bioenergetics and Biomembranes*, **36**, 5-15.
- Albers, S.V., Van de Vossenberg, J., Driessen, A.J.M. and Konings, W.N. (2001) Bioenergetics and solute uptake under extreme conditions. *Extremophiles*, **5**, 285-294.
- Alter, O., Brown, P.O. and Botstein, D. (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10101-10106.
- Alter, O., Brown, P.O. and Botstein, D. (2003) Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 3351-3356.
- Arfin, S.M., Long, A.D., Ito, E.T., Toller, L., Riehle, M.M., Paegle, E.S. and Hatfield, G.W. (2000) Global gene expression profiling in Escherichia coli K12 - The effects of integration host factor. *Journal of Biological Chemistry*, **275**, 29672-29684.
- Bairoch, A., Boeckmann, B., Ferro, S. and Gasteiger, E. (2004) Swiss-Prot: Juggling between evolution and stability. *Briefings in Bioinformatics*, **5**, 39-55.
- Baliga, N.S. and DasSarma, S. (2000) Saturation mutagenesis of the haloarchaeal bop gene promoter: identification of DNA supercoiling sensitivity sites and absence of TFB recognition element and UAS enhancer activity. *Molecular Microbiology*, **36**, 1175-1183.
- Baliga, N.S., Goo, Y.A., Ng, W.V., Hood, L., Daniels, C.J. and DasSarma, S. (2000) Is gene expression in Halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors? *Molecular Microbiology*, **36**, 1184-1185.
- Baliga, N.S., Kennedy, S.P., Ng, W.V., Hood, L. and DasSarma, S. (2001) Genomic and genetic dissection of an archaeal regulon. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 2521-2525.

- Ball, C., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Parkinson, H., Quackenbush, J., Ringwald, M., Sansone, S.A., Sherlock, G., Spellman, P., Stoeckert, C., Tateno, Y., Taylor, R., White, J. and Winegarden, N. (2004) Microarray data standards: An open letter. *Environmental Health Perspectives*, **112**, A666-A667.
- Ball, C., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Parkinson, H., Quackenbush, J., Ringwald, M., Sansone, S.A., Sherlock, G., Spellman, P., Stoeckert, C., Tateno, Y., Taylor, R., White, J. and Winegarden, N. (2004) An open letter on microarray data from the MGED Society. *Microbiology-Sgm*, **150**, 3522-3524.
- Ball, C.A., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Parkinson, H., Quackenbush, J., Ringwald, M., Sansone, S.A., Sherlock, G., Spellman, P., Stoeckert, C., Tateno, Y., Taylor, R., White, J. and Winegarden, N. (2004) Submission of microarray data to public repositories. *Plos Biology*, **2**, 1276-1277.
- Ball, C.A., Sherlock, G., Parkinson, H., Rocca-Serra, P., Brooksbank, C., Causton, H.C., Cavalieri, D., Gaasterland, T., Hingamp, P., Holstege, F., Ringwald, M., Spellman, P., Stoeckert, C.J., Stewart, J.E., Taylor, R., Brazma, A. and Quackenbush, J. (2002) Standards for Microarray data. *Science*, **298**, 539-539.
- Bell, S.D. (2005) Archaeal transcriptional regulation - variation on a bacterial theme? *Trends in Microbiology*, **13**, 262-265.
- Bell, S.D. and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Current Opinion in Microbiology*, **4**, 208-213.
- Bell, S.D., Magill, C.P. and Jackson, S.P. (2001) Basal and regulated transcription in Archaea. *Biochemical Society Transactions*, **29**, 392-395.
- Benes, V. and Muckenthaler, M. (2003) Standardization of protocols in cDNA microarray analysis. *Trends in Biochemical Sciences*, **28**, 244-249.
- Betlach, M.C., Shand, R.F. and Leong, D.M. (1989) Regulation of the Bacterio-opsin Gene of a Halophilic Archaeobacterium. *Canadian Journal of Microbiology*, **35**, 134-140.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and

- Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, **31**, 365-370.
- Bolstad, B.M., Collin, F., Simpson, K.M., Irizarry, R.A. and Speed, T.P. (2004) Experimental design and low-level analysis of microarray data. *DNA Arrays in Neurobiology, Vol 60*, **60**, 25-+.
- Boucher, Y., Kamekura, M. and Doolittle, W.F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Molecular Microbiology*, **52**, 515-527.
- Brazma, A., Sarkans, U., Robinson, A., Vilo, J., Vingron, M., Hoheisel, J. and Fellenberg, K. (2002) Microarray data representation, annotation and storage. *Advances in Biochemical Engineering Biotechnology. Chip technology*, 113-139.
- Brazma, A. and Vilo, J. (2001) Gene expression data analysis. *Microbes and Infection*, **3**, 823-829.
- Bretz, F., Landgrebe, J. and Brunner, E. (2005) Design and analysis of two-color microarray experiments using linear models. *Methods of Information in Medicine*, **44**, 423-430.
- Bretz, F., Landgrebe, J. and Brunner, E. (2005) Multiplicity issues in microarray experiments. *Methods of Information in Medicine*, **44**, 431-437.
- Brown, P.O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet.*, **21**, 33-7.
- Chapman, S., Schenk, P., Kazan, K. and Manners, J. (2002) Using biplots to interpret gene expression patterns in plants. *Bioinformatics*, **18**, 202-204.
- Cheng, C., Pounds, S.B., Boyett, J.M., Pei, D., Kuo, M.L. and Roussel, M.F. (2004) Statistical significance threshold criteria for analysis of microarray gene expression data. *Stat Appl Genet Mol Biol.*, **3**, Article36. Epub 2004 Dec 19.
- Chomczynski, P. and Sacchi, N. (1987) Single-Step Method of Rna Isolation by Acid Guanidinium Thiocyanate Phenol Chloroform Extraction. *Analytical Biochemistry*, **162**, 156-159.
- Churchill, G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genetics*, **32**, 490-495.
- Churchill, G.A. (2004) Using ANOVA to analyze microarray data. *Biotechniques*, **37**, 173-+.

- Cui, X.G., Hwang, J.T.G., Qiu, J., Blades, N.J. and Churchill, G.A. (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*, **6**, 59-75.
- Cui, X.Q. and Churchill, G.A. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology*, **4**.
- Curtis, R.K., Oresic, M. and Vidal-Puig, A. (2005) Pathways to the analysis of microarray data. *Trends in Biotechnology*, **23**, 429-435.
- Daffertshofer, A., Lamoth, C.J.C., Meijer, O.G. and Beek, P.J. (2004) PCA in studying coordination and variability: a tutorial. *Clinical Biomechanics*, **19**, 415-428.
- Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. and Conklin, B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genetics*, **31**, 19-20.
- Danon, A. and Caplan, S.R. (1977) Co<sub>2</sub> Fixation by Halobacterium-Halobium. *Febs Letters*, **74**, 255-258.
- Danson, M.J. (1989) Central Metabolism of the Archaeobacteria - an Overview. *Canadian Journal of Microbiology*, **35**, 58-64.
- Danson, M.J. and Hough, D.W. (1992) The Enzymology of Archaeobacterial Pathways of Central Metabolism. *Biochemical Society Symposium*, 7-21.
- Dassarma, S., Berquist, B.R., Coker, J.A., Dassarma, P. and Muller, J.A. (2006) Post-genomics of the model haloarchaeon Halobacterium sp. NRC-1. *Saline Systems.*, **2**, 3.
- DasSarma, S., Kennedy, S.P., Berquist, B., Ng, W.V., Baliga, N.S., Spudich, J.L., Krebs, M.P., Eisen, J.A., Johnson, C.H. and Hood, L. (2001) Genomic perspective on the photobiology of Halobacterium species NRC-1, a phototrophic, phototactic, and UV-tolerant haloarchaeon. *Photosynthesis Research*, **70**, 3-17.
- de Haan, J.R., Wehrens, R., Bauerschmidt, S., Piek, E., van Schaik, R.C. and Buydens, L.M. (2006) Interpretation of ANOVA models for microarray data using PCA. *Bioinformatics*, **14**, 14.
- DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y.D., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics*, **14**, 457-460.

- Derisi, J.L. and Iyer, V.R. (1999) Genomics and array technology. *Current Opinion in Oncology*, **11**, 76-79.
- DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680-686.
- Deshpande, A. and Sonar, S. (1999) Bacterioopsin-triggered retinal biosynthesis is inhibited by bacteriorhodopsin formation in *Halobacterium salinarium*. *Journal of Biological Chemistry*, **274**, 23535-23540.
- Diehl, F., Grahlmann, S., Beier, M. and Hoheisel, J.D. (2001) Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. *Nucleic Acids Res.*, **29**, E38.
- Donders, A.R., van der Heijden, G.J., Stijnen, T. and Moons, K.G. (2006) Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.*, **59**, 1087-91. Epub 2006 Jul 11.
- Dudoit, S., Gendeman, R.C. and Quackenbush, J. (2003) Open source software for the analysis of microarray data. *Biotechniques*, 45-51.
- Dufva, M. (2005) Fabrication of high quality microarrays. *Biomolecular Engineering*, **22**, 173-184.
- Duggan, D.J., Bittner, M., Chen, Y.D., Meltzer, P. and Trent, J.M. (1999) Expression profiling using cDNA microarrays. *Nature Genetics*, **21**, 10-14.
- Eisen, M.B. and Brown, P.O. (1999) DNA arrays for analysis of gene expression. *Methods Enzymol.*, **303**, 179-205.
- Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14863-14868.
- El-Sayed, W.S.M., Takaichi, S., Saida, H., Kamekura, M., Abu-Shady, M., Seki, H. and Kuwabara, T. (2002) Effects of light and low oxygen tension on pigment biosynthesis in *Halobacterium salinarum*, revealed by a novel method to quantify both retinal and carotenoids. *Plant and Cell Physiology*, **43**, 379-383.
- Elvidge, G. (2006) Microarray expression technology: from start to finish. *Pharmacogenomics*, **7**, 123-134.

- Fathallah-Shaykh, H.M. (2005) Microarrays - Applications and pitfalls. *Archives of Neurology*, **62**, 1669-1672.
- Fellenberg, K., Hauser, N.C., Brors, B., Hoheisel, J.D. and Vingron, M. (2002) Microarray data warehouse allowing for inclusion of experiment annotations in statistical analysis. *Bioinformatics*, **18**, 423-433.
- Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T. and Solas, D. (1991) Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science*, **251**, 767-773.
- Frank, H.A. and Cogdell, R.J. (1996) Carotenoids in photosynthesis. *Photochemistry and Photobiology*, **63**, 257-264.
- Fraser, N.J., Hashimoto, H. and Cogdell, R.J. (2001) Carotenoids and bacterial photosynthesis: The story so far. *Photosynthesis Research*, **70**, 249-256.
- Freeman, W.M., Robertson, D.J. and Vrana, K.E. (2000) Fundamentals of DNA hybridization arrays for gene expression analysis. *Biotechniques*, **29**, 1042-+.
- Gabriel, K.R. and Odoroff, C.L. (1990) Biplots in Biomedical-Research. *Statistics in Medicine*, **9**, 469-485.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D. and Brown, P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, **11**, 4241-4257.
- Geiduschek, E.P. and Ouhammouch, M. (2005) Archaeal transcription and its regulators. *Molecular Microbiology*, **56**, 1397-1407.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y.C., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H. and Zhang, J.H. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, **5**.
- Gropp, F. and Betlach, M.C. (1994) The Bat Gene of Halobacterium-Halobium Encodes a Trans-Acting Oxygen Inducibility Factor. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 5475-5479.

- Gropp, F., Gropp, R. and Betlach, M.C. (1995) Effects of Upstream Deletions on Light-Regulated and Oxygen-Regulated Bacterioopsin Gene-Expression in Halobacterium-Halobium. *Molecular Microbiology*, **16**, 357-364.
- Gurtner, C., Tu, E., Jamshidi, N., Haigis, R.W., Onofrey, T.J., Edman, C.F., Sosnowski, R., Wallace, B. and Heller, M.J. (2002) Microelectronic array devices and techniques for electric field enhanced DNA hybridization in low-conductance buffers. *Electrophoresis*, **23**, 1543-1550.
- Hartmann, R. and Oesterhelt, D. (1977) Bacteriorhodopsin-Mediated Photophosphorylation in Halobacterium-Halobium. *European Journal of Biochemistry*, **77**, 325-335.
- Hartmann, R., Sickinger, H. and Oesterhelt, D. (1980) Anaerobic Growth of Halobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 3821-3825.
- Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N. and Quackenbush, J. (2000) A concise guide to cDNA microarray analysis. *Biotechniques*, **29**, 548-+.
- Heller, M.J. (2002) DNA microarray technology: Devices, systems, and applications. *Annual Review of Biomedical Engineering*, **4**, 129-153.
- Heller, M.J., Forster, A.H. and Tu, E. (2000) Active microelectronic chip devices which utilize controlled electrophoretic fields for multiplex DNA hybridization and other genomic applications. *Electrophoresis*, **21**, 157-164.
- Herzel, H., Beule, D., Kielbasa, S., Korbel, J., Sers, C., Malik, A., Eickhoff, H., Lehrach, H. and Schuchhardt, J. (2001) Extracting information from cDNA arrays. *Chaos*, **11**, 98-107.
- Hoheisel, J.D. (2006) Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics*, **7**, 200-210.
- Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G.N., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S.C., Hoffman, E., Jedlicka, A.E., Kawasaki, E., Martinez-Murillo, F., Morsberger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A., Wilson, M., Yang, Y.Q., Ye, S.Q. and Yu, W. (2005) Multiple-laboratory comparison of microarray platforms. *Nature Methods*, **2**, 345-349.



- Johansson, P. and Hakkinen, J. (2006) Improving missing value imputation of microarray data by using spot quality weights. *BMC Bioinformatics.*, **7**, 306.
- Jolliffe, I.T. and Morgan, B.J. (1992) Principal component analysis and exploratory factor analysis. *Stat Methods Med Res.*, **1**, 69-95.
- Kennedy, S.P., Ng, W.V., Salzberg, S.L., Hood, L. and DasSarma, S. (2001) Understanding the adaptation of Halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Research*, **11**, 1641-1650.
- Kerr, M.K. (2003) Design considerations for efficient and effective microarray studies. *Biometrics*, **59**, 822-828.
- Kerr, M.K. (2003) Experimental design to make the most of microarray studies. *Methods Mol Biol.*, **224**, 137-47.
- Kerr, M.K. and Churchill, G.A. (2001) Experimental design for gene expression microarrays. *Biostatistics.*, **2**, 183-201.
- Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genetical Research*, **77**, 123-128.
- Kerr, M.K., Martin, M. and Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, **7**, 819-837.
- Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F. and Oesterhelt, D. (2005) The membrane proteome of Halobacterium salinarum. *Proteomics*, **5**, 180-197.
- Koch, M.K. and Oesterhelt, D. (2005) MpcT is the transducer for membrane potential changes in Halobacterium salinarum. *Molecular Microbiology*, **55**, 1681-1694.
- Kolbe, M., Besir, H., Essen, L.O. and Oesterhelt, D. (2000) Structure of the light-driven chloride pump halorhodopsin at 1.8 angstrom resolution. *Science*, **288**, 1390-1396.
- Krajewski, P. and Bocianowski, J. (2002) Statistical methods for microarray assays. *Journal of Applied Genetics*, **43**, 269-278.
- Krzanowski, W.J. (2004) Biplots for multifactorial analysis of distance. *Biometrics*, **60**, 517-524.
- Kyrpides, N.C. and Ouzounis, C.A. (1999) Transcription in Archaea. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 8545-8550.

- Landgrebe, J., Bretz, F. and Brunner, E. (2004) Efficient two-sample designs for microarray experiments with biological replications. *In Silico Biology*, **4**, 461-470.
- Landgrebe, J., Welzl, G., Metz, T., van Gaalen, M.M., Ropers, H., Wurst, W. and Holsboer, F. (2002) Molecular characterisation of antidepressant effects in the mouse brain using gene expression profiling. *Journal of Psychiatric Research*, **36**, 119-129.
- Landgrebe, J., Wurst, W. and Welzl, G. (2002) Permutation-validated principal components analysis of microarray data. *Genome Biol.*, **3**, RESEARCH0019. Epub 2002 Mar 22.
- Langer, D., Hain, J., Thuriaux, P. and Zillig, W. (1995) Transcription in Archaea: Similarity to that in Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 5768-5772.
- Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R. and Quackenbush, J. (2005) Independence and reproducibility across microarray platforms. *Nature Methods*, **2**, 337-343.
- Leung, Y.F. and Cavalieri, D. (2003) Fundamentals of cDNA microarray data analysis. *Trends in Genetics*, **19**, 649-659.
- Lilley, D.M.J. (2001) The ribosome functions as a ribozyme. *ChemBiochem*, **2**, 31-35.
- Lilley, D.M.J. (2003) The origins of RNA catalysis in ribozymes. *Trends in Biochemical Sciences*, **28**, 495-501.
- Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genetics*, **21**, 20-24.
- Lipshutz, R.J., Morris, D., Chee, M., Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R. and Fodor, S.P.A. (1995) Using Oligonucleotide Probe Arrays to Access Genetic Diversity. *Biotechniques*, **19**, 442-447.
- Liu, L., Hawkins, D.M., Ghosh, S. and Young, S.S. (2003) Robust singular value decomposition analysis of microarray data. *Proc Natl Acad Sci U S A.*, **100**, 13167-72. Epub 2003 Oct 27.
- Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827-836.

- Lou, X.J., Schena, M., Horrigan, F.T., Lawn, R.M. and Davis, R.W. (2001) Expression monitoring using cDNA microarrays. A general protocol. *Methods Mol Biol.*, **175**, 323-40.
- Madigan, M.T. and Oren, A. (1999) Thermophilic and halophilic extremophiles. *Current Opinion in Microbiology*, **2**, 265-269.
- Maskos, U. and Southern, E.M. (1992) Oligonucleotide Hybridisations on Glass Supports a Novel Linker for Oligonucleotide Synthesis and Hybridisation Properties of Oligonucleotides Synthesised in-Situ. *Nucleic Acids Research*, **20**, 1679-1684.
- Muller, J.A. and DasSarma, S. (2005) Genomic analysis of anaerobic respiration in the archaeon Halobacterium sp strain NRC-1: Dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *Journal of Bacteriology*, **187**, 1659-1667.
- Muller, V. and Oren, A. (2003) Metabolism of chloride in halophilic prokaryotes. *Extremophiles*, **7**, 261-266.
- Nadon, R. and Shoemaker, J. (2002) Statistical issues with microarrays: processing and analysis. *Trends in Genetics*, **18**, 265-271.
- Ness, S.A. (2006) Basic microarray analysis - Strategies for successful experiments. *Methods in Molecular Biology*, 13-33.
- Neugebauer, D.C., Zingsheim, H.P. and Oesterhelt, D. (1983) Biogenesis of purple membrane in halobacteria. *Methods Enzymol.*, **97**, 218-26.
- Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroder, M., Spudich, J.L., Jung, K.H., Alam, M., Freitas, T., Hou, S.B., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, R., Riley, M., Hood, L. and DasSarma, S. (2000) Genome sequence of Halobacterium species NRC-1. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 12176-12181.
- Oesterhelt, D. (1975) The Purple Membrane of Halobacterium-Halobium a New System for Light Energy Conversion. *Wolstenholme, G.E.W. And David W. Fitzsimons (Ed.). Ciba Foundation Symposium, No. 31. Energy*

- Transformation in Biological Systems. London, England, July 2-4, 1974. X+416p. Illus. Associated Scientific Publishers: Amsterdam, Netherlands; American Elsevier: New York, N.Y., U.S.A. Isbn 90-219-4034-5; Isbn 0-444-15164-8, 147-167.*
- Oesterhelt, D. (1976) Bacterio Rhod Opsin as an Example of a Light Driven Proton Pump. *Angewandte Chemie International Edition in English*, **15**, 17-24.
- Oesterhelt, D. (1976) Bacteriorhodopsin as a Light-Driven Ion-Exchanger. *Febs Letters*, **64**, 20-22.
- Oesterhelt, D. (1998) The structure and mechanism of the family of retinal proteins from halophilic archaea. *Current Opinion in Structural Biology*, **8**, 489-500.
- Oesterhelt, D. (2004) Photosynthese und andere Energieumwandlungsprozesse in Zellen. *Nova Acta Leopoldina*, **NF 91**, 17-39.
- Oesterhelt, D. and Krippahl, G. (1973) Light Inhibition of Respiration in Halobacterium-Halobium. *Febs Letters*, **36**, 72-76.
- Oesterhelt, D. and Stoeckenius, W. (1973) Functions of a New Photo Receptor Membrane. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 2853-2857.
- Oesterhelt, D. and Stoeckenius, W. (1974) Isolation of the Cell Membrane of Halobacterium-Halobium and Its Fractionation into Red and Purple Membrane. *Fleischer, Sidney and Lester Packer (Ed.). Methods in Enzymology, Vol. 31. Biomembranes, Part A. Xx+889p. Illus. Academic Press: New York, N.Y., U.S.A.; London, England. Isbn 0-12-181894-2, 667-678.*
- Oren, A. (1999) Bioenergetic aspects of halophilism. *Microbiology and Molecular Biology Reviews*, **63**, 334-+.
- Ouhammouch, M. (2004) Transcriptional regulation in Archaea. *Current Opinion in Genetics & Development*, **14**, 133-138.
- Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M. and Geiduschek, E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5097-5102.

- Peck, R.F., Echavarri-Erasun, C., Johnson, E.A., Ng, W.V., Kennedy, S.P., Hood, L., DasSarma, S. and Krebs, M.P. (2001) brp and blh are required for synthesis of the retinal cofactor of bacteriorhodopsin in *Halobacterium salinarum*. *J Biol Chem.*, **276**, 5739-44. Epub 2000 Nov 22.
- Peck, R.F., Johnson, E.A. and Krebs, M.P. (2002) Identification of a lycopene beta-cyclase required for bacteriorhodopsin biogenesis in the Archaeon *Halobacterium salinarum*. *Journal of Bacteriology*, **184**, 2889-2897.
- Pfeifer, F.A., Boyer, H.W. and Betlach, M.C. (1985) Restoration of Bacterioopsin Gene-Expression in a Revertant of *Halobacterium-Halobium*. *Journal of Bacteriology*, **164**, 414-420.
- Pounds, S.B. (2006) Estimation and control of multiple testing error rates for microarray studies. *Brief Bioinform.*, **7**, 25-36.
- Quackenbush, J. (2003) Microarrays - Guilt by association. *Science*, **302**, 240-241.
- Quackenbush, J. (2003) Open-source software accelerates bioinformatics. *Genome Biol.*, **4**, 336. Epub 2003 Sep 1.
- Quackenbush, J. (2004) Data standards for 'omic' science. *Nature Biotechnology*, **22**, 613-614.
- Quackenbush, J. (2006) Standardizing the standards. *Mol Syst Biol.*, **2**, 2006.0010. Epub 2006 Feb 21.
- Raddatz, G., Dehio, M., Meyer, T.F. and Dehio, C. (2001) PrimeArray: genome-scale primer design for DNA-microarray construction. *Bioinformatics*, **17**, 98-99.
- Raychaudhuri, S., Chang, J.T., Imam, F. and Altman, R.B. (2003) The computational analysis of scientific literature to define and recognize gene expression clusters. *Nucleic Acids Research*, **31**, 4553-4560.
- Rudolph, J., Nordmann, B., Storch, K.-F., Gruenberg, H., Rodewald, K. and Oesterhelt, D. (1996) A family in halobacterial transducer proteins. *FEMS Microbiology Letters*, **139**, 161-168.
- Rudolph, J. and Oesterhelt, D. (1995) Chemotaxis and phototaxis require a CheA histidine kinase in the archaeon *Halobacterium salinarum*. *EMBO (European Molecular Biology Organization) Journal*, **14**, 667-673.
- Ruepp, A., Muller, H.N., Lottspeich, F. and Soppa, J. (1995) Catabolic Ornithine Transcarbamylase of *Halobacterium-Halobium (Salinarium)* - Purification,

- Characterization, Sequence Determination, and Evolution. *Journal of Bacteriology*, **177**, 1129-1136.
- Ruepp, A. and Soppa, J. (1996) Fermentative arginine degradation in *Halobacterium salinarium* (formerly *Halobacterium halobium*): Genes, gene products, and transcripts of the arcRACB gene cluster. *Journal of Bacteriology*, **178**, 4942-4947.
- Scheel, I., Aldrin, M., Glad, I.K., Sorum, R., Lyng, H. and Frigessi, A. (2005) The influence of missing value imputation on detection of differentially expressed genes from microarray data. *Bioinformatics.*, **21**, 4272-9. Epub 2005 Oct 10.
- Scheel, J., von Brevern, M.C., Horlein, A., Fischer, A., Schneider, A. and Bach, A. (2002) Yellow pages to the transcriptome. *Pharmacogenomics*, **3**, 791-807.
- Schena, M. (1996) Genome analysis with gene expression microarrays. *Bioessays*, **18**, 427-431.
- Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E. and Davis, R.W. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends in Biotechnology*, **16**, 301-306.
- Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H. and Herzel, H. (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Res.*, **28**, E47.
- Shand, R.F. and Betlach, M.C. (1991) Expression of the bop gene cluster of *Halobacterium halobium* is induced by low oxygen tension and by light. *J Bacteriol.*, **173**, 4692-9.
- Shand, R.F. and Betlach, M.C. (1991) Expression of the Bop Gene-Cluster of *Halobacterium-Halobium* Is Induced by Low Oxygen-Tension and by Light. *Journal of Bacteriology*, **173**, 4692-4699.
- Shand, R.F. and Betlach, M.C. (1994) Bop Gene-Cluster Expression in Bacteriorhodopsin-Overproducing Mutants of *Halobacterium-Halobium*. *Journal of Bacteriology*, **176**, 1655-1660.
- Sharov, V., Kwong, K.Y., Frank, B., Chen, E., Hasseman, J., Gaspard, R., Yu, Y., Yang, I. and Quackenbush, J. (2004) The limits of log-ratios. *Bmc Biotechnology*, **4**.

- Shih, S.C. and Smith, L.E.H. (2005) Quantitative multi-gene transcriptional profiling using real-time PCR with a master template. *Experimental and Molecular Pathology*, **79**, 14-22.
- Soppa, J. (1999) Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Molecular Microbiology*, **31**, 1589-1592.
- Soppa, J. (1999) Transcription initiation in Archaea: facts, factors and future aspects. *Molecular Microbiology*, **31**, 1295-1305.
- Soppa, J. (2001) Basal and regulated transcription in Archaea. *Archaea: Ancient Microbes, Extreme Environments, and the Origin of Life*, **50**, 171-217.
- Soppa, J. (2005) From replication to cultivation: hot news from Haloarchaea. *Current Opinion in Microbiology*, **8**, 737-744.
- Soppa, J. (2006) From genomes to function: haloarchaea as model organisms. *Microbiology-Sgm*, **152**, 585-590.
- Southern, E.M. (1975) Detection of Specific Sequences among DNA Fragments Separated by Gel-Electrophoresis. *Journal of Molecular Biology*, **98**, 503-&.
- Southern, E.M. (1992) Genome mapping: cDNA approaches. *Curr Opin Genet Dev.*, **2**, 412-6.
- Southern, E.M. (1996) DNA chips: Analysing sequence by hybridization to oligonucleotides on a large scale. *Trends in Genetics*, **12**, 110-115.
- Southern, E.M. (2001) DNA microarrays. History and overview. *Methods Mol Biol.*, **170**, 1-15.
- Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, **9**, 3273-3297.
- Stahler, P., Beier, M., Gao, X. and Hoheisel, J.D. (2006) Another side of genomics: Synthetic biology as a means for the exploitation of whole-genome sequence information. *J Biotechnol.*, **124**, 206-12. Epub 2006 Jan 24.
- Stears, R.L., Martinsky, T. and Schena, M. (2003) Trends in microarray analysis. *Nature Medicine*, **9**, 140-145.

- Storey, J.D., Xiao, W.Z., Leek, J.T., Tompkins, R.G. and Davis, R.W. (2005) Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12837-12842.
- Sturn, A., Quackenbush, J. and Trajanoski, Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics*, **18**, 207-208.
- Szabo, A., Perou, C.M., Karaca, M., Perreard, L., Quackenbush, J.F. and Bernard, P.S. (2004) Statistical modeling for selecting housekeeper genes. *Genome Biology*, **5**.
- Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. (2001) Quantitative assessment of DNA microarrays - Comparison with Northern blot analyses. *Genomics*, **71**, 34-39.
- Taylor, J., Tibshirani, R. and Efron, B. (2005) The 'miss rate' for the analysis of gene expression data. *Biostatistics*, **6**, 111-117.
- Tebbe, A., Klein, C., Bisle, B., Siedler, F., Scheffer, B., Garcia-Rizo, C., Wolfertz, J., Hickmann, V., Pfeiffer, F. and Oesterhelt, D. (2005) Analysis of the cytosolic proteome of *Halobacterium salinarum* and its implication for genome annotation. *Proteomics*, **5**, 168-179.
- Tibshirani, R. (2006) A simple method for assessing sample sizes in microarray experiments. *Bmc Bioinformatics*, **7**.
- Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5116-5121.
- Werner, F., Eloranta, J.J. and Weinzierl, R.O. (2000) Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Res.*, **28**, 4299-305.
- Winzler, E.A., Schena, M. and Davis, R.W. (1999) Fluorescence-based expression monitoring using microarrays. *Expression of Recombinant Genes in Eukaryotic Systems*, **306**, 3-+.
- Woese, C.R. (2001) Translation: In retrospect and prospect. *Rna-a Publication of the Rna Society*, **7**, 1055-1067.
- Woese, C.R. (2004) The archaeal concept and the world it lives in: a retrospective. *Photosynthesis Research*, **80**, 362-372.



- Woese, C.R. (2004) A new biology for a new century. *Microbiology and Molecular Biology Reviews*, **68**, 173-+.
- Woese, C.R. and Fox, G.E. (1977) Phylogenetic Structure of Prokaryotic Domain - Primary Kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5088-5090.
- Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a Natural System of Organisms - Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 4576-4579.
- Wolf, D., Gray, C.P. and de Saizieu, A. (2000) Visualising gene expression in its metabolic context. *Brief Bioinform.*, **1**, 297-304.
- Woodson, J.D., Peck, R.F., Krebs, M.P. and Escalante-Semerena, J.C. (2003) The cobY gene of the archaeon Halobacterium sp strain NRC-1 is required for de novo cobamide synthesis. *Journal of Bacteriology*, **185**, 311-316.
- Wu, X. and Dewey, T.G. (2006) From microarray to biological networks - Analysis of gene expression profiles. *Methods in Molecular Biology*, 35-48.
- Yang, M.C.K., Yang, J.J., McIndoe, R.A. and She, J.X. (2003) Microarray experimental design: power and sample size considerations. *Physiological Genomics*, **16**, 24-28.
- Yang, Y.H., Buckley, M.J. and Speed, T.P. (2001) Analysis of cDNA microarray images. *Brief Bioinform.*, **2**, 341-9.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation - art. no. e15. *Nucleic Acids Research*, **30**, E15-E15.
- Ye, R.W., Wang, T., Bedzyk, L. and Croker, K.M. (2001) Applications of DNA microarrays in microbial systems. *Journal of Microbiological Methods*, **47**, 257-272.
- Yeung, K.Y., Haynor, D.R. and Ruzzo, W.L. (2001) Validating clustering for gene expression data. *Bioinformatics*, **17**, 309-318.
- Yeung, K.Y. and Ruzzo, W.L. (2001) Principal component analysis for clustering gene expression data. *Bioinformatics*, **17**, 763-774.

- Zaigler, A., Schuster, S.C. and Soppa, J. (2003) Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon *Haloferax volcanii*. *Molecular Microbiology*, **48**, 1089-1105.
- Zhang, J., Carey, V. and Gentleman, R. (2003) An extensible application for assembling annotation for genomic data. *Bioinformatics (Oxford)*, **19**, 155-156.
- Zhu, T. (2003) Global analysis of gene expression using GeneChip microarrays. *Current Opinion in Plant Biology*, **6**, 418-425.



## 16. Danksagung

Ich bedanke mich bei allen Personen, die mich bei dieser Arbeit unterstützt haben.

Mein besonderer Dank gilt:

Meinem Doktorvater Prof. Dr. D. Oesterhelt für die Möglichkeit diese Arbeit in seinem Arbeitskreis anfertigen zu können. Ich bedanke mich für seine großzügige Unterstützung, seine außerordentlichen Kenntnisse, zahlreichen Anregungen, sein Interesse an dieser Arbeit und die vielen Freiheiten zur eigenen Gestaltung des Themas. Vor allem aber für seine Geduld besonders im letzten Jahr.

Dr. Gerhard Welzl (GSF-Forschungszentrum f. Umwelt u. Gesundheit, GmbH Neuherberg) und Dr. Jobst Landgrebe (Universität Göttingen) für die Programmierung der Auswertungsprogramme der Datenanalyse und deren Geduld. Sie haben einen erheblichen Anteil am Gelingen dieser Arbeit beigetragen, denn ohne deren Hilfe hätten die roten, grünen und gelben Punkte wohl für immer ihr Geheimnis für sich behalten.

Dr. Markus Panhuysen und Dr. Alexander Zaigler für die erfolgreiche Kooperation und Zusammenarbeit bei der Herstellung der DNA-Mikroarrays.

Der Bioinformatikgruppe der „ersten“ Stunde mit Friedhelm Pfeiffer, Michaela Falb, Jan Wolfertz, Volker Hickmann und Carolina Garcia-Rizo.

Meinen Praktikanten Anne Scholz, Lasse Peters und Markus Schwarzwälder für ihre super Hilfe im Labor.

Allen meinen Kollegen und Mitarbeitern des Arbeitskreises Membranbiochemie für ihre Hilfsbereitschaft und das angenehme Arbeitsklima am Institut. Mit euch hat nicht nur das Arbeiten und Diskutieren sehr viel Spaß gemacht, sondern auch die tollen Ausflüge, Ski-Tage, Retreats und Feste.

Ich möchte die herausgreifen, denen ich in besonderer Weise zu Dank verpflichtet bin:

Da wäre die Labortruppe des Oe-Labors mit Bettina, Brigitte, Susanne, Andy, Jörg, Christian, Flo und allen anderen momentanen und ehemaligen Kollegen. Vielen, vielen Dank für die hilfsbereite, offene Arbeitsatmosphäre und tatkräftige Unterstützung. Ohne euch wäre manches Experiment nicht möglich gewesen und die Frühstücks- und Kaffeepausen bei weitem kürzer und langweiliger. Besonders

schön ist, dass Bettina, Brigitte und Flo mich „schon“ in der Schweiz besucht haben.

Michaela Falb für ihre Hilfe, Motivation, tollen Diskussionen und so mancher Tasse Tee.

Wilfried Staudinger für seine Hilfe, die (nicht ganz so regelmäßigen) Lauftreffs, Cinema Oe-Abende und die vielen, vielen kleinen und großen anderen Dinge. Aber ganz besonders für das Asyl in Martinsried in der Nach-Instituts-Zeit.

Matthias Schlesner nicht nur als tollen Kollegen, sondern auch als Ruderpartner. Mit ihm und Bernhard Döhla habe ich so manche Stunde auf dem Starnberger See verbracht.

Nicht vergessen sind die „Guten Geister“ Christoph Opawsky, Walter Gmelin und Hüseyin Besir, die mir nicht nur das Leben mit dem Computer erleichtert (ermöglicht) haben.

Es war eine tolle Zeit bei euch in Martinsried!

Schlussendlich möchte ich meiner Familie danken: Andrea, Rainer, Erik, Anneke, ganz neu Till-Fridtjoff, Tante Liesel, Omi Frieda und dem Rest der Bande. Ohne euch wäre meine Welt wesentlich ärmer.

Meinen Eltern Marianne und Günter werde ich wohl niemals genug danken können für ihr Vertrauen, ihre Zuneigung, ihre bedingungslose Unterstützung, ihrem Rückhalt und der Tatsache, dass ich immer mit ihrer Hilfe rechnen kann. Ich hoffe diesen Luxus noch lange „genießen“ zu können.

## 17. Lebenslauf

<b>Persönliche Daten</b>	Jens Twellmeyer geboren am 26.06.1971
	in Bramsche, Niedersachsen deutsch ledig
<b>Schulbildung</b>	08/78 – 06/82 Grundschule Achmer
	08/82 – 06/84 Orientierungsstufe Bramsche
	07/84 – 06/91 Greselius Gymnasium Bramsche; Abschluß: Allgemeine Hochschulreife
<b>Zivildienst</b>	07/91 – 09/92 Staatl. Medizinaluntersuchungsamt Osnabrück
<b>Universität</b>	10/92 – 03/99 Universität Bielefeld; Studiengang: Dipl.-Chemie
	08/98 – 03/99 Diplomarbeit bei Prof. H. Tschesche : „Untersuchungen zur Proteolyse von Integrinen“
	04/99 – 09/99 Wissenschaftlicher Angestellter in der Arbeitsgruppe von Prof. H. Tschesche
<b>Promotion</b>	10/99 – 05/01 Klinische Forschergruppe der Frauenklinik der TU München, Prof. Schmitt : „Entwicklung von synthe- tischen und natürlichen Inhibitoren des Urokinase Plasminogen aktivatorsystems“
	06/01 – 10/05 Max-Planck-Institut für Biochemie, Prof. D. Oesterhelt : „Transkriptom- Analyse von <i>Halobacterium salinarum</i> “
<b>Auslandsaufenthalt</b>	03/97 – 06/97 ERASMUS-Stipendium an der Dublin City University, Irland
<b>Anstellung</b>	Seit 10/05 Sigma-Aldrich Chemie GmbH Buchs (SG) Schweiz