

INAUGURAL-DISSERTATION

zur  
Erlangung der Doktorwürde  
der  
Naturwissenschaftlichen-Mathematischen Gesamtfakultät  
der  
Ruprecht - Karls - Universität  
Heidelberg

vorgelegt von

Diplom-Biologe

Boris Beckmann

aus:

Ratingen

Tag der mündlichen Prüfung: 05.02.2004

Studien der Genexpressionsänderung während der Entwicklung von  
*Drosophila melanogaster* mittels DNA-Microarrays

Gutachter: Prof. Dr. Frank Sauer

Prof. Dr. Richard Herrmann

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis.....</b>	<b>1</b>
<b>Abkürzungsverzeichnis.....</b>	<b>1</b>
<b>Zusammenfassung.....</b>	<b>3</b>
<b>Abstract.....</b>	<b>4</b>
<b>1 Einleitung.....</b>	<b>5</b>
<b>1.1 Die Fruchtfliege <i>Drosophila melanogaster</i> .....</b>	<b>5</b>
1.1.1 Der Modellorganismus.....	5
1.1.2 Der Lebenszyklus von <i>Drosophila melanogaster</i> .....	5
<b>1.2 Microarray-Technologie.....</b>	<b>10</b>
<b>1.3 Genexpressionsanalysen.....</b>	<b>12</b>
<b>1.4 Die Grundlage des Heidelberg FlyArray.....</b>	<b>15</b>
<b>2 Material und Methoden.....</b>	<b>17</b>
<b>2.1 Material.....</b>	<b>17</b>
2.1.1 Chemikalien und Reagenzien.....	17
2.1.2 Längenmarker für die Gelelektrophorese.....	18
2.1.3 Enzyme.....	18
2.1.4 Reagenziensätze („Kits“).....	19
2.1.5 <i>Drosophila melanogaster</i> genomische DNA und Gesamt-RNA.....	19
2.1.6 Bakterienstämme und Vektoren.....	19
2.1.7 Oligonukleotide.....	19
2.1.8 Lösungen und Puffer.....	21
2.1.9 Geräte.....	23
2.1.10 Verbrauchsmaterial.....	24
2.1.11 Materialien.....	24
2.1.12 Software und Internetadressen.....	25
<b>2.2 Methoden.....</b>	<b>26</b>
2.2.1 PCR-Amplifikation der Heidelberg Kollektion R1.....	26
Herstellung von Taq DNA-Polymerase.....	26
PCR-Amplifikation mit Gen-spezifischen Primern („First-step PCR“).....	26
PCR-Amplifikation mit universellen Primern („Second-step PCR“).....	27
Kontrolle der Heidelberg Kollektion R1 PCR-Amplifikationen.....	28
2.2.2 Herstellung heterogener Kontrollen für die Microarray-Analyse.....	28
PCR-Amplifikation der Kontrollen.....	28
Aufreinigung der PCR-Produkte.....	30
Herstellung externer Kontroll-mRNA durch <i>in vitro</i> Transkription.....	30
2.2.3 Bestimmung der RNA-Konzentration.....	30
2.2.4 RNA-Agarosegelelektrophorese (denaturierend).....	31
2.2.5 Bestimmung der DNA-Konzentration.....	31
2.2.6 Herstellung komplexer genomischer Kontrollen für die Microarray-Analyse.....	32
2.2.7 Herstellung von Positionsmarkern für die Microarray-Analyse.....	32
2.2.8 Herstellung der cDNA-Microarrays („Heidelberg FlyArray“).....	32
Vorbereitung der PCR-Produkte.....	32

Auftragung der DNA („Spotting“) auf Glasobjektträger.....	33
Nachbehandlung („Processing“) der cDNA-Microarrays.....	33
2.2.9 Fluoreszenzmarkierung von Nukleinsäuren.....	34
Fluoreszenzmarkierung von DNA.....	34
Fluoreszenzmarkierung von RNA durch cDNA-Erststrangsynthese (Reverse Transkription).....	34
2.2.10 Reinigung von markierten Nukleinsäuren.....	35
2.2.11 Photometrische Bestimmung der Einbaurate von Cy5/Cy3.....	35
2.2.12 Hybridisierung auf cDNA-Microarrays.....	36
2.2.13 Auswertung der Hybridisierungsdaten.....	37
Signaldetektion.....	37
Quantifizierung der gemessenen Signalintensitäten.....	37
Normalisierung und Datenanalyse mit M-CHiPS.....	38
Statistische Analyse der differentiellen Genexpression.....	40
Clusteranalysen der Transkriptionsprofile exprimierter Gene.....	41
Datenbankanwendung und Bioinformatik.....	42
<b>3 Ergebnisse.....</b>	<b>44</b>
<b>3.1 Optimierung der Eigenschaften für Microarrays mit hoher Dichte.....</b>	<b>44</b>
3.1.1 Ziel der Optimierungen.....	44
3.1.2 Verwendung unaufgereinigter PCR-Produkte.....	44
3.1.3 Erhöhung der DNA-Bindungskapazität auf Glasobjektträgern.....	47
3.1.4 Auswahl der geeigneten Glasobjektträger.....	49
<b>3.2 Aufbau des Microarray „Heidelberg FlyArray“.....</b>	<b>51</b>
3.2.1 PCR-Amplifikation der Heidelberg Kollektion R1.....	51
3.2.2 Herstellung der heterologen Kontrollen.....	53
3.2.3 Herstellung der cDNA-Microarrays.....	54
<b>3.3 Durchführung der Experimente.....</b>	<b>55</b>
3.3.1 Fluoreszenzmarkierung der RNA.....	55
3.3.2 Hybridisierungen auf den Microarrays.....	56
3.3.3 Signaldetektion und Quantifizierung der Hybridisierungsdaten.....	58
<b>3.4 Analyse der differentiellen Genexpression in der Entwicklung von <i>Drosophila melanogaster</i>.....</b>	<b>58</b>
3.4.1 Filterung und Auswertung der Daten mit M-CHiPS.....	58
3.4.2 Statistische Analyse der differentiellen Genexpression.....	61
3.4.3 Analyse der Genregulation mit Self-organizing maps und k-means - Clusteralgorithmen.....	62
Embryonalstadium.....	64
Larvenstadium.....	65
.....	69
Puppenstadium.....	70
Adultstadium .....	73
Sonstige Cluster.....	73
<b>4 Diskussion.....</b>	<b>76</b>
<b>4.1 Methoden zur Untersuchung der Genexpression.....</b>	<b>76</b>
4.1.1 Optimierung der DNA-Chip-Technologie.....	76
4.1.2 PCR-Amplifikation der Heidelberg Kollektion R1 und Konstruktion des Heidelberg FlyArray.....	78
4.1.3 Markierung der RNA.....	80

4.1.4	Hybridisierung der Microarrays.....	81
4.1.5	Auswertung der Expressionsdaten.....	82
<b>4.2</b>	<b>Untersuchung der Genexpression im Verlauf der Entwicklung von <i>Drosophila melanogaster</i>.....</b>	<b>83</b>
4.2.1	Genannotationen und Transkriptom von <i>Drosophila melanogaster</i> .....	83
4.2.2	Genexpression in den Entwicklungsstadien bei <i>Drosophila melanogaster</i> .....	85
<b>4.3</b>	<b>Ausblick.....</b>	<b>90</b>
<b>5</b>	<b>Danksagung.....</b>	<b>92</b>
<b>6</b>	<b>Eigene Publikationen.....</b>	<b>93</b>
<b>7</b>	<b>Literaturverzeichnis.....</b>	<b>94</b>

## Abkürzungsverzeichnis

aa-dUTP	Aminoallyl-2'-Desoxyuridin-5'-triphosphat
ATP	Adenosin-5'-triphosphat
bp	Basenpaare
BSA	Rinderserumalbumin
cDNA	komplementäre DNA
dATP	2'-Desoxyadenosin-5'-triphosphat
dCTP	2'-Desoxycytidin-5'-triphosphat
DEPC	Diethylpyrocarbonat
dGTP	2'-Desoxyguanosin-5'-triphosphat
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxyribonukleosid-5'-triphosphat
dsDNA	doppelsträngige DNA
dTTP	2'-Desoxythymidin-5'-triphosphat
dUTP	2'-Desoxyuridin-5'-triphosphat
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraessigsäure
g	Gramm
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranosid
h	Stunde
k	kilo (1000)
kb	Kilobasenpaare
l	Liter
M	Molar
m	milli ( $10^{-3}$ )
Mb	Megabasenpaare
min	Minuten
mRNA	messenger RNA
n	nano ( $10^{-9}$ )
OD	Optische Dichte
ORF	offener Leserahmen (open reading frame)
p	pico ( $10^{-12}$ )
PCR	Polymerase-Kettenreaktion (Polymerase Chain Reaction)
PP	Polypropylen
PS	Polysterol
RNA	Ribonukleinsäure
RNase	Ribonuklease

rpm	Umdrehung pro Minute
RT	Raumtemperatur / Reverse Transkription
SDS	Natriumdodecylsulfat
SSC	Saline Sodium Citrate
ssDNA	Einzelstrang DNA (single stranded DNA)
TAE	Tris-Acetat-EDTA
Taq	<i>Thermus aquaticus</i>
TE	Tris-EDTA
Tm	Schmelztemperatur
TNE	Tris-Natrium-EDTA
Tris	Tris-(hydroxymethyl)-aminomethan
U	Unit
UV	Ultraviolett
V	Volt
w/v	weight/volume
w/w	weight/weight
μ	Mikro (10 <sup>-6</sup> )

## Zusammenfassung

Die Fruchtfliege *Drosophila melanogaster* ist einer der am besten untersuchten Modellorganismen in der Entwicklungsbiologie. Die Kenntnis des Genoms und die Zahl der annotierten protein-kodierenden Gene von *Drosophila* sind jedoch Schwankungen unterworfen, da sie meist auf Computerberechnungen bereits bekannter Gene und nur teilweise auf experimentellen Nachweisen beruhen. Um das Transkriptom von *Drosophila* im Verlauf der Entwicklung zu untersuchen und neue potentielle Gene zu identifizieren wurde im Rahmen dieser Arbeit ein DNA-Microarray entwickelt, mit dem Veränderungen in der mRNA-Expression analysiert wurden. Auf der Grundlage einer neuentwickelten *ab initio* Genvorhersage im Rahmen einer Kooperation wurden spezifische Primer generiert, die eine Amplifikation der potentiellen Gene ermöglichten. Mit einem zweistufigen PCR-Protokoll konnten 20.948 (98%) PCR-Fragmente von 21.396 vorhergesagten Genen amplifiziert werden. Diese wurden als Basis für die Herstellung des Microarrays verwendet. So war es möglich die Vorteile computergestützter *in silico*- Berechnungen der möglichen Gene mit experimentellen Analysen zu kombinieren, um die Expression von Genen zu detektieren. Durch Optimierung der Protokolle und Methoden in der DNA-Chip-Konstruktion wurde ein hochdichter Microarray hergestellt, der nahezu das gesamte Transkriptom von *Drosophila* repräsentiert. Die Analyse differentiell exprimierter Gene erfolgte jeweils mit zwei RNA-Populationen die während der reversen Transkription mit unterschiedlichen Fluoreszenzfarbstoffen markiert und anschließend auf einem Microarray co-hybridisiert wurden. Es wurden Hybridisierungen mit der RNA aus 9 unterschiedlichen Entwicklungsstadien von *Drosophila melanogaster* durchgeführt, die den gesamten Lebenszyklus der Fliege umfassten. Nach der Quantifizierung der detektierten Signalintensitäten erfolgte die bioinformatische Datenanalyse mit unterschiedlichen Softwareprogrammen. Die Genexpressionsanalysen ergaben das 13.812 Gene im Verlauf der Entwicklung von *Drosophila* exprimiert werden, wovon ungefähr 2.500 bisher unbekanntes Genen entsprechen. Es konnte außerdem gezeigt werden, daß die Entwicklungsstadien spezifische Regulationen von Genaktivitäten aufweisen, die mit morphologischen Veränderungen während der Entwicklung korrelieren und daß Gengruppen ähnlicher biochemischer oder physiologischer Funktion gemeinsam reguliert werden.



## Abstract

The fly *Drosophila melanogaster* is one of the best known model organism in current Developmental biology. Assuming to know the structure of the genome and the amount of annotated genes coding for proteins based on *in silico* predictions and partially experimental data only has to be treated with caution since it lacks experimental evidence.

To study the gene expression during the development of *Drosophila*, a new DNA-Microarray, the Heidelberg FlyArray, was developed to identify potential new genes.

A novel *ab initio* gene prediction allowed the generation of 20,948 (98%) amplicons in a two-step PCR-process based on 21,396 specific primer pairs calculated from the genomic sequence of *Drosophila*. Using these Amplicons for the construction of the microarray, we combined the advantages of *in silico* gene predictions with an experimental tool to monitor gene expression. By optimising protocols and methods for DNA-chip construction and handling, a high-density microarray was built covering nearly the whole transcriptome of *Drosophila*. For the analysis of differential expressed genes, total RNA taken from nine stages of *Drosophila* development, was reverse transcribed and indirect labeled with different Cy3/Cy5-fluorophores. The labeled cDNAs, representing the life cycle of *Drosophila* were co-hybridised onto the microarray. After detection and quantification of the signals, computational analysis on the data was performed. Using different filtering and statistical bioinformatic methods, we found 13,812 genes expressed throughout the development of *Drosophila* and identified nearly 2,500 novel genes. The application of various clustering methods on the expressed genes revealed transcriptional profiles of gene regulation specific for the stages and co-regulation of genes with similar function in biochemical pathways and physiology.

# 1 Einleitung

## 1.1 Die Fruchtfliege *Drosophila melanogaster*

### 1.1.1 Der Modellorganismus

*Drosophila melanogaster* („Gemeine Essigfliege“) gehört zur Familie der Taufliegen und spielt seit Anfang des letzten Jahrhunderts eine herausragende Rolle in der klassischen Genetik und modernen Entwicklungsbiologie. Aufgrund ihrer anspruchslosen Haltung im Labor und der hohen Reproduktionsrate eignet sich *Drosophila* vorzüglich für die Studie genetischer Mechanismen in der Entwicklung eines Organismus. Obwohl *Drosophila* zu den Wirbellosen zählt hat sie das Verständnis genetischer Grundlagen der Wirbeltierentwicklung enorm erweitert. Viele Schlüsselgene der Wirbeltiere (z.B. Mensch oder Maus) wie die an der Entwicklung beteiligten *Homöobox*-Gene wurden zuerst bei *Drosophila* entdeckt, und es konnten Mechanismen wie Achsendeterminierung, Identifizierung von Signalzyklen und die Transkriptionsregulation bei der Entstehung der Musterbildung auf molekularer Ebene aufgeklärt werden. Ein weiterer wichtiger Schritt für das Verständnis komplexer genetischer Systeme war die Sequenzierung des *Drosophila*-Genoms im Jahre 2000.

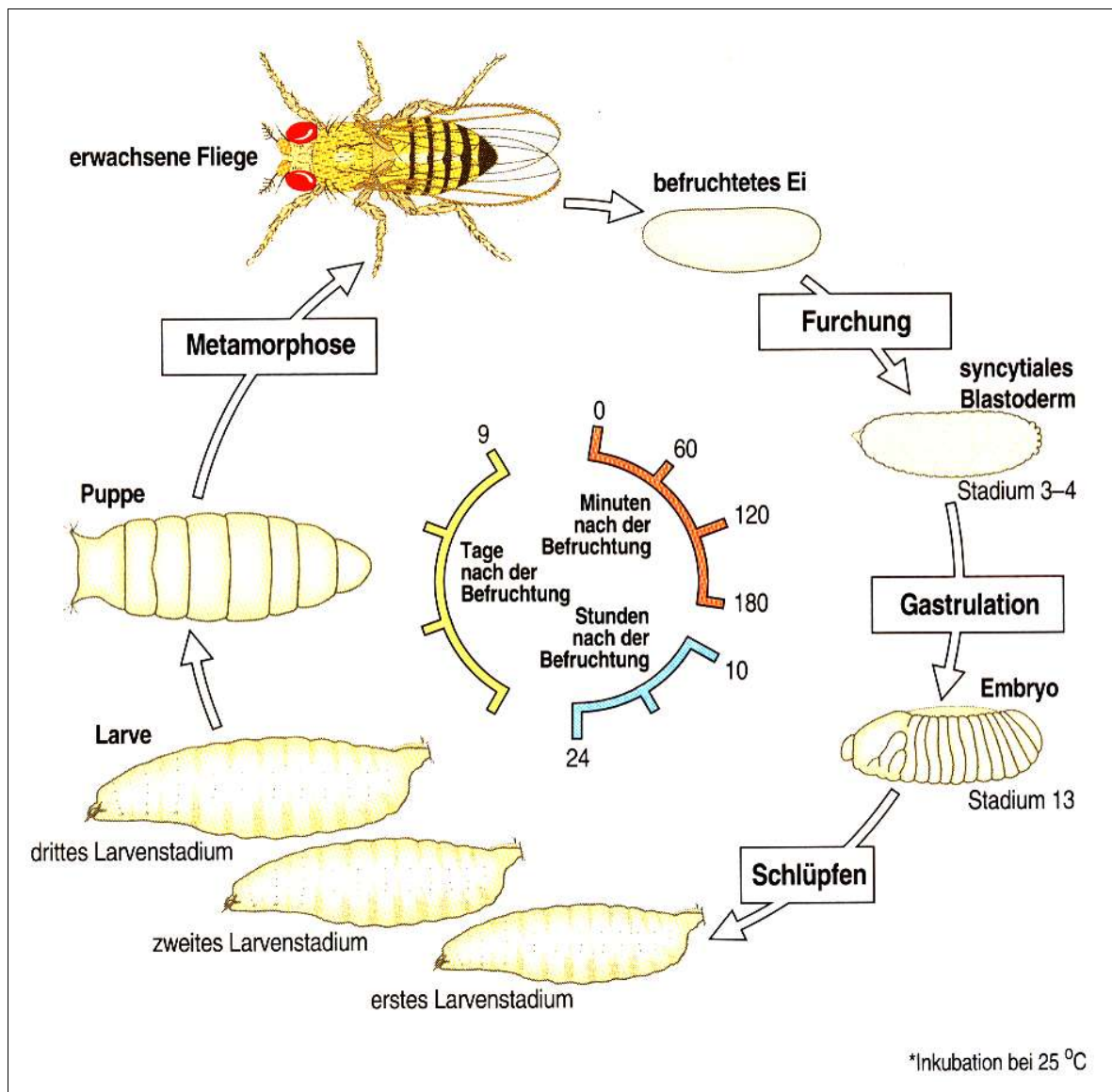
Das Genom von *Drosophila* hat eine Länge von 180 Mb, wovon ca. ein Drittel Heterochromatin ist. Die 120 Mb Euchromatin verteilen sich auf vier Chromosomen: zwei Autosomen, ein X-Chromosom und ein kleines Chromosom, das nur ungefähr 1 Mb Euchromatin enthält. Es wird angenommen, daß die Zahl der proteinkodierenden Gene ca. 14.000 beträgt [1]. Die Sequenz des Genoms liegt mittlerweile in der dritten Version mit fehlerbereinigten Sequenzinformationen vor [2].

### 1.1.2 Der Lebenszyklus von *Drosophila melanogaster*

Die Fruchtfliege *Drosophila melanogaster* ist aufgrund zahlreicher genetischer Studien und der Möglichkeit, genetische und chirurgische Manipulationen kombinieren zu können eines der am besten erforschten Entwicklungs-Modellsysteme [3,4]. Das Ei von *Drosophila* ist wurstförmig und an einem kleinen Fortsatz der äußeren Hülle, der Mikropyle, zu erkennen. Diese Hülle umschließt das ganze Ei. Spermien können durch diese Mikropyle in das anteriore Ende des Eies eindringen. Nach der Befruchtung erfolgt die Verschmelzung der Zellkerne von Eizelle und Spermium und der Zellkern der Zygote durchläuft eine Reihe

schneller mitotischer Teilungen. Im Gegensatz zu den meisten anderen Tierembryonen wird das Cytoplasma jedoch nicht geteilt, sondern es entsteht ein Syncytium bei dem sich viele Zellkerne in einem gemeinsamen Cytoplasma befinden. Das bedeutet, daß der Embryo während seiner frühen Entwicklung praktisch aus einer einzigen Zelle besteht. Erst nach Einsetzen der Segmentierung wird der Embryo vielzellig. Dieser Umstand ist ein entscheidender Unterschied und Vorteil für Manipulationen auf genetischer Ebene gegenüber anderen Organismen. Nach neun Teilungen wandern die Kerne an die Peripherie und bilden das syncytiale Blastoderm. Die vorläufige Zellbildung erfolgt durch eine Einstülpung der Membranen an der Oberfläche des Eies in welche die Zellkerne eingeschlossen werden. Die endgültigen Zellen werden nach 13 durchlaufenen Mitosen gebildet, wobei ca. 15 Zellen am Hinterende des Embryos verbleiben und sich zu den Polzellen weiterentwickeln, die später zu Keimzellen (für Spermien und Eier) werden.

Durch das Syncytium können Proteine, z.B. Transkriptionsfaktoren, während der ersten drei Stunden der Entwicklung zwischen den Zellkernen diffundieren und die Genaktivität durch Bildung von Transkriptionskaskaden beeinflussen. Die weibliche Fliege produziert mRNAs und Proteine im Ei, die das früheste Stadium der *Drosophila*-Entwicklung steuern, weshalb man sie als maternale Gene bezeichnet. Sie bilden einen Gradienten von maternalen Transkriptionsfaktoren, die entlang der antero-posterioren und der dorso-ventralen Körperachsen verlaufen, und aktivieren nach der Befruchtung in bestimmten Bereichen zygotische Gene.

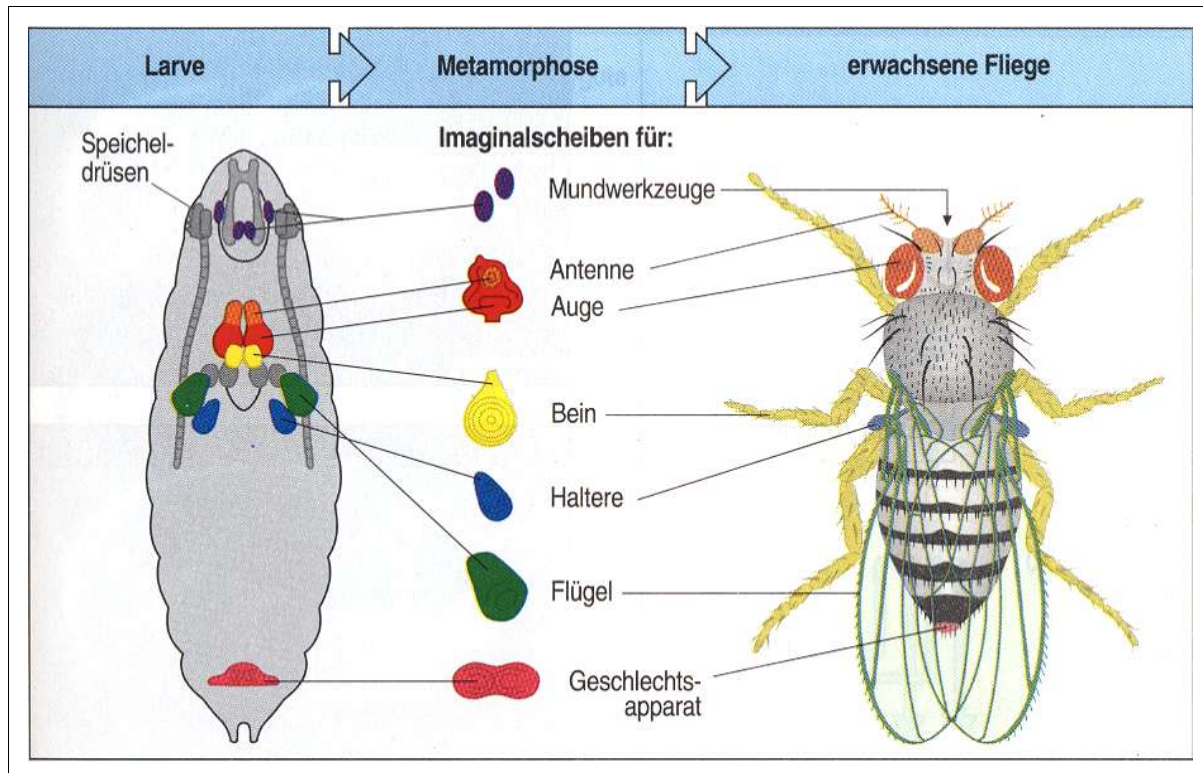


**Abb.1.1:** Lebenszyklus von *Drosophila melanogaster* aus [3].

Das entstehende Muster unterteilt den Embryo in unterschiedliche Regionen, die alle eine spezifische zygotische Genaktivität aufweisen. Insgesamt existieren 4 Hauptklassen von zygotischen Genen: Lückengene, Paarregelgene, Segmentpolaritätsgene und Selektorgene. Die Entwicklungsgene sind in einer festgelegten zeitlichen Reihenfolge aktiv, wodurch sich eine Art Hierarchie der Genaktivitäten entwickelt, bei der eine Gruppe von Genen für die Aktivierung einer anderen Gengruppe und daher für das nächste Entwicklungsstadium essentiell ist. Von einer einzigen epithelialen Schicht des zellulären Blastoderms stammen alle zukünftigen Gewebe ab. Der Mitteldarm des adulten Tieres wird aus dem vorderen und hinteren Bereich des Entoderms gebildet. Das zukünftige Mesoderm befindet sich in der am weitesten ventral gelegenen Region. Während der Gastrulation wandern entodermale und mesodermale Gewebe an ihre Positionen im Embryoinneren und das Ektoderm bildet die äußere Schicht. Die Gastrulation beginnt drei Stunden nach der Befruchtung, indem sich das

zukünftige Mesoderm in der Bauchregion einstülpt und entlang der ventralen Mittellinie eine Furche bildet. Ähnlich dem Neuralrohr der Wirbeltiere gelangen Mesodermzellen nach innen, wobei eine mesodermale Röhre entsteht. Die Zellen wandern nach dem Ablösen von der Oberfläche des Neuralrohres unterhalb des Ektoderms ins Innere, wo sie später Muskeln und andere Bindegewebe bilden. Im Gegensatz zu den Wirbeltieren verläuft bei Insekten wie bei allen Arthropoden der Hauptnervenstrang nicht auf der dorsalen, sondern auf der ventralen Seite. Ektodermale Zellen aus der Ventralregion, die das Nervensystem bilden, verlassen nach der Einstülpung des Mesoderms die Oberfläche und lagern sich zwischen dem Mesoderm und dem äußeren Ektoderm zu den Neuroblasten zusammen. Gleichzeitig entwickeln sich zwei Einstülpungen an beiden Seiten des zukünftigen anterioren und posterioren Mitteldarms, die nach innen wachsen und zum Entoderm des mittleren Darmes verschmelzen. Das Ektoderm wird dahinter an beiden Seiten nach innen gezogen und bildet den Vorder- und Enddarm, während sich die äußere Ektodermschicht zur Epidermis entwickelt. Das ventrale Blastoderm (Keimstreifen) umfaßt die wichtigste Rumpfreion und breitet sich während der Gastrulation aus, wodurch die hinteren Rumpfbereiche um das Hinterende herum auf die bisherige Dorsalseite geleitet werden. Nach der Embryonalentwicklung zieht sich der Keimstreifen wieder zurück. Bei einer Ausdehnung lassen sich erste Anzeichen der Segmentierung erkennen. Es zeigt sich dabei eine Reihe gleichmäßig angeordneter Vertiefungen, welche die Grenzen der Parasegmente darstellen, aus denen sich später die Segmente der Larve und der adulten Fliege entwickeln. Die Parasegmente und Segmente sind gegeneinander verschoben, wobei ein Segment jeweils aus dem hinteren Teil des einen und dem vorderen des folgenden Parasegmentes besteht. Insgesamt besteht der Embryo aus 14 Parasegmenten, von denen drei an den Mundpartien des Kopfes, drei an den Thorax- und acht an den Abdominalsegmenten beteiligt sind. Der nächste große Entwicklungsschritt ist die Larve, die etwa 24 Stunden nach der Befruchtung schlüpft und deren unterschiedliche Körperbereiche bereits mehrere Stunden vorher ausgebildet wurden. Der Kopf der Larve besteht aus einem vorderen Bereich, dem Akron und dem am weitesten hinten liegendem, der als Telson bezeichnet wird. Aufgrund von besonderen Merkmalen in der Cuticula, die durch Ausstülpungen der Epidermis zustande kommen, wird der dazwischen liegende Teil in drei Thorax- und acht Abdominalsegmente unterschieden. Jedes Segment besitzt charakteristische cuticuläre Strukturen wie z.B. Dentrikelstreifen auf seiner Ventralseite. Die Larve frißt und nimmt an Körpergröße zu, worauf Häutungen stattfinden und die Cuticula abgeworfen wird. Das geschieht zweimal, wobei jedes Stadium als Larvenstadium (Instar) bezeichnet wird. Neben spezifischen Organen besitzt die Larve weder Flügel noch Beine und auch weitere Organe erscheinen erst nach dem dritten Larvenstadium durch die Metamorphose, die hormonell reguliert wird. Die Organe und Gliedmaßen der Puppe sind jedoch bereits in der Larve als Imaginalscheiben angelegt. Dies sind Strukturen aus zukünftigen Epidermalzellen, die aus

dem zellulären Blastoderm stammen und jeweils aus ungefähr 40 Zellen bestehen. Es bilden sich daraus gefaltete Epithelsäckchen, die während des gesamten Larvenstadiums mitwachsen. Imaginalscheiben existieren für jedes der sechs Beine, für die beiden Flügel und Haltere, für den Genitalapparat, die Augen, die Antennen und die Kopfstrukturen der adulten Fliege. Nach dem Puppenstadium, das durch weiteres Wachstum gekennzeichnet ist, schlüpft nach ca. neun Tagen die erwachsene Fliege.



**Abb.1.2:** Die Metamorphose. Aus den Imaginalscheiben der Larve entstehen im Verlauf der Metamorphose adulte Strukturen wie Gliedmaßen und Organe in der erwachsenen Fliege, aus [3].

## 1.2 Microarray-Technologie

Der DNA-Chip ist allgemein ein Überbegriff für die Anwendung bei der Nukleinsäuremoleküle in einem Raster auf einen Träger aufgebracht werden. Man unterscheidet je nach Trägermaterial, Nukleinsäureart und Mechanismus der Immobilisierung hauptsächlich drei Klassen von DNA-Chips: „Macroarrays“, „Microarrays“ und „Oligo-Arrays“ [5,6]. Um Missverständnissen vorzubeugen, wurde die Terminologie für die DNA-Chip-Technologie in Übereinstimmung mit einer Konvention [7] wie folgt definiert: die auf dem Chip immobilisierte DNA wird als Sonde („probe“) bezeichnet auf die die freibewegliche Probe („target“) in Form markierter DNA oder RNA hybridisiert wird. Macroarrays wurden von Southern (1975) im Rahmen neuer Hybridisierungstechniken entwickelt. Sie bestehen aus Membranen wie Nitrocellulose, Nylon oder Polypropylen, auf die Sonden in Form von PCR-Produkten (ca. 200 – 3000 bp), Plasmid-Klonen (ca. 2000 – 10.000 bp) oder lange Oligonukleotiden (50 – 70 bp) [8] mechanisch oder mit der Hand aufgebracht werden. Die Bindung der DNA erfolgt elektrostatisch über die positiv geladene Oberfläche der Membranen. So befinden sich dann in jedem dieser Auftragspunkte („Spots“) identische Nukleinsäuremoleküle die viele Kopien einer bestimmten Sequenz tragen und damit ein Gen repräsentieren. Aufgrund des porösen Trägermaterials wird eine größere Menge an Sondenmaterial für den Auftrag erforderlich, wobei Raster mit einer Dichte von nicht mehr als 100 Sonden pro  $\text{cm}^2$  erstellt werden können, was bei einer Membrangröße von 22x22 cm ungefähr 36.000 Sonden entspricht. Durch die Größe und die Porösität der Membranen erfolgt die Hybridisierung in relativ großen Volumina von bis zu 20 ml, wodurch die Probe stark verdünnt wird. Die Markierung der Proben erfolgt meist radioaktiv oder chemilumineszent. Microarrays werden meist aus festen Trägermaterialien, wie Glas hergestellt. Als Sonden werden auch hier PCR-Produkte oder lange Oligonukleotide, in selteneren Fällen auch Plasmid-Klone verwendet. Die DNA-Sonden werden mit Hilfe von speziellen Robotern („Arraying device“) mechanisch in einem definierten, hochdichten Raster auf das Glas aufgebracht (siehe Abb. 1.3).



**Abb.1.3:** Ein Microarray-Roboter beim Auftragen der Sonde. Der Ausschnitt zeigt den Kopf mit den Nadeln („pins“) die gelöste DNA-Sonden in einem Flüssigkeitsreservoir enthalten und durch Kontakt mit der Oberfläche auf die Glasobjekträger auftragen („spotting“).

Die Glasoberfläche wird vor dem DNA-Auftrag so modifiziert, daß die Nukleinsäuremoleküle entweder kovalent (z.B. über Aldehydgruppen) oder über Ladungswechselwirkungen (z.B. über Aminogruppen) an der Oberfläche fixiert werden können. Im Gegensatz zu den porösen Membranen sind aufgrund der soliden, undurchlässigen Trägereigenschaften des Glases wesentlich geringere Mengen an Sondenmaterial für den Auftrag nötig. Eine Miniaturisierung der Microarrays wird aus diesem Grund und durch die Verwendung von fluoreszenzmarkierten Hybridisierungsproben, die entweder mit hochauflösenden CCD-Kameras oder mit konfokalen Lasermikroskopen detektiert werden können, erreicht. Durch die Präzision der robotergesteuerten Fixierung der DNA werden Raster von einer Dichte mit bis zu 2000 Sonden pro  $\text{cm}^2$  auf dem Microarray ermöglicht. Normalerweise werden Glasobjekträger mit dem Standardformat 25x75 mm verwendet. Für eine Hybridisierung sind nur sehr kleine Volumina von 10 – 50  $\mu\text{l}$ , meist unter einem Deckglas nötig, wodurch eine hohe lokale Probenkonzentration und folglich eine verstärkte Sensitivität erreicht wird.

Die dritte Klasse, die sogenannten Oligo-Arrays stellen eine Abwandlung der Microarrays dar und unterscheiden sich dadurch, daß die Sonden in Form von kurzen Oligonukleotiden *in situ* auf der Oberfläche des Chips synthetisiert werden. Die Oligonukleotide werden dabei photolithographisch oder chemisch synthetisiert. Dabei werden die Oligonukleotide pro Zyklus jeweils um ein Nukleotid verlängert [9,10]. Durch das photolithographische Verfahren werden derzeit die höchsten Rasterdichten mit bis zu 400.000 Oligonukleotiden auf einer Fläche von 1,6  $\text{cm}^2$  erreicht (Affymetrix). Die Spezifität für ein Gen ist aufgrund der kurzen Oligonukleotide nicht immer gewährleistet, weshalb meistens mehrere verschiedene Oligonukleotide für ein Gen ausgewählt werden, die die gesamte Gensequenz abdecken.

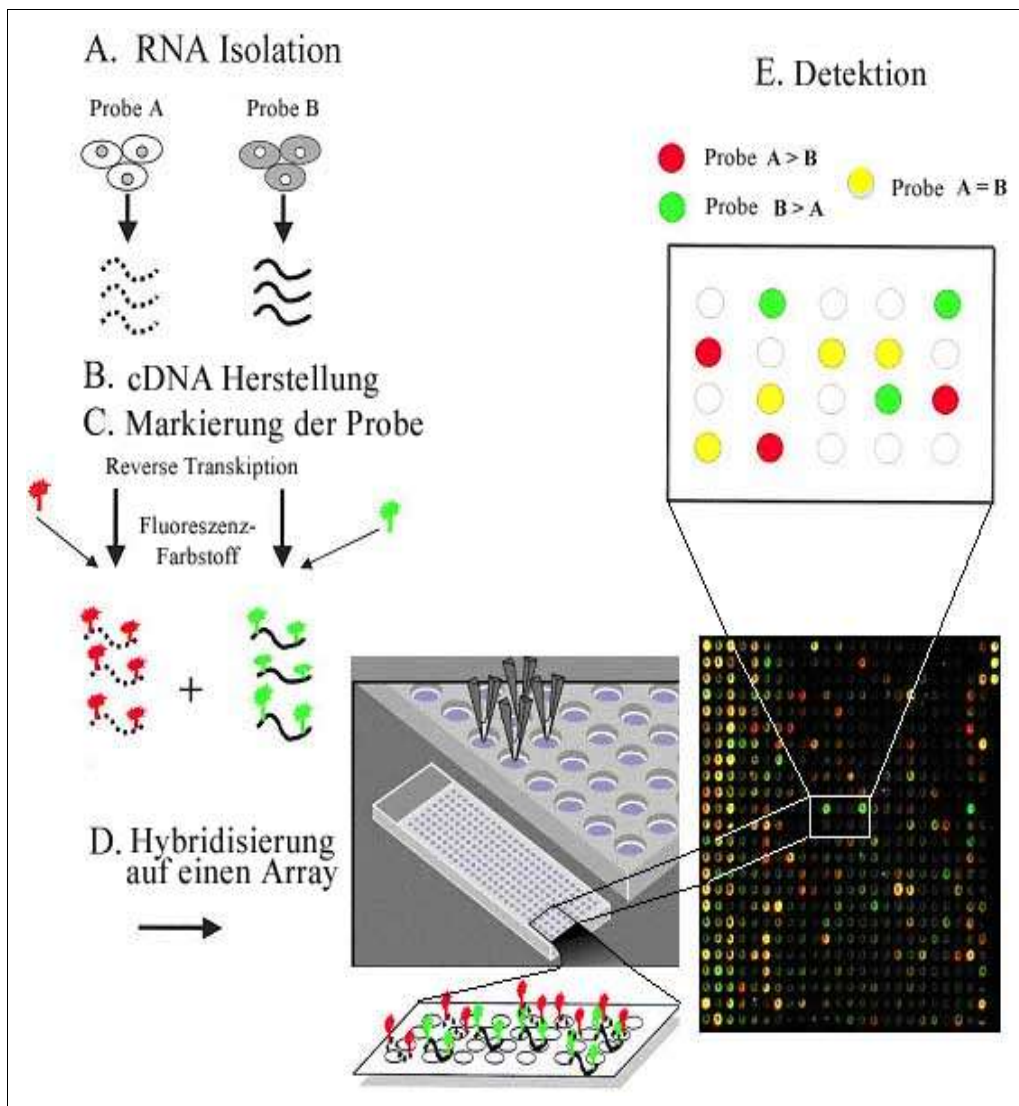


### 1.3 Genexpressionsanalysen

Nach der Sequenzierung ganzer Genome erforderte die darauffolgende Phase der Genomforschung, die „funktionelle Genomanalyse“, die Entwicklung effizienter Techniken um die Expression mehrerer tausend Gene parallel zu identifizieren. Über die Sequenzdaten hinaus ist es von besonderem Interesse Methoden zu entwickeln, die es erlauben Gene nicht nur individuell zu untersuchen, sondern auch das genregulatorische Netzwerk ganzer Organismen zu erfassen [11].

Eine der entwickelten Hochdurchsatztechniken ist die Analyse der Genexpression mittels der DNA-Chip-Technologie, die auf der Hybridisierung von mRNA oder Gesamt-RNA auf ein geordnetes Raster von immobilisierten Nukleinsäuremolekülen basiert, welche mit den Sequenzen der zu analysierenden Gene korrelieren. Bei älteren Techniken wie dem „Northern Blot“ können nur einzelne Gene analysiert werden, hingegen ist es mit der DNA-Chip-Technologie möglich die Expression mehrerer tausend Gene gleichzeitig zu untersuchen [12,13]. Die verwendeten Microarrays basieren zumeist auf cDNA-Bibliotheken bei denen die klonierten cDNA-Fragmente mittels PCR amplifiziert und anschließend auf den Glasobjektträger aufgebracht werden. Bei der Verwendung von cDNA-Bibliotheken liegt der Vorteil für die Untersuchungen darin, daß weder die Sequenz, noch nähere Informationen über einzelne Gene vorliegen müssen. Der Nachteil ist, daß zwischen eventuell vorhandenen Spleiss-Varianten nicht differenziert werden kann. Bei bekannter Sequenz des Genoms können spezifische Primerpaare für eine gezielte Amplifikation oder lange Oligonukleotide (50 – 70 bp) zur Herstellung der Microarrays verwendet werden. Die aus dem ganzen Organismus-, oder aus Gewebe isolierte RNA wird für die Hybridisierung als komplexe Probe markiert. In der zu untersuchenden biologischen Probe sind alle vorhandenen mRNA -Spezies entsprechend ihrer Transkriptmenge repräsentiert.

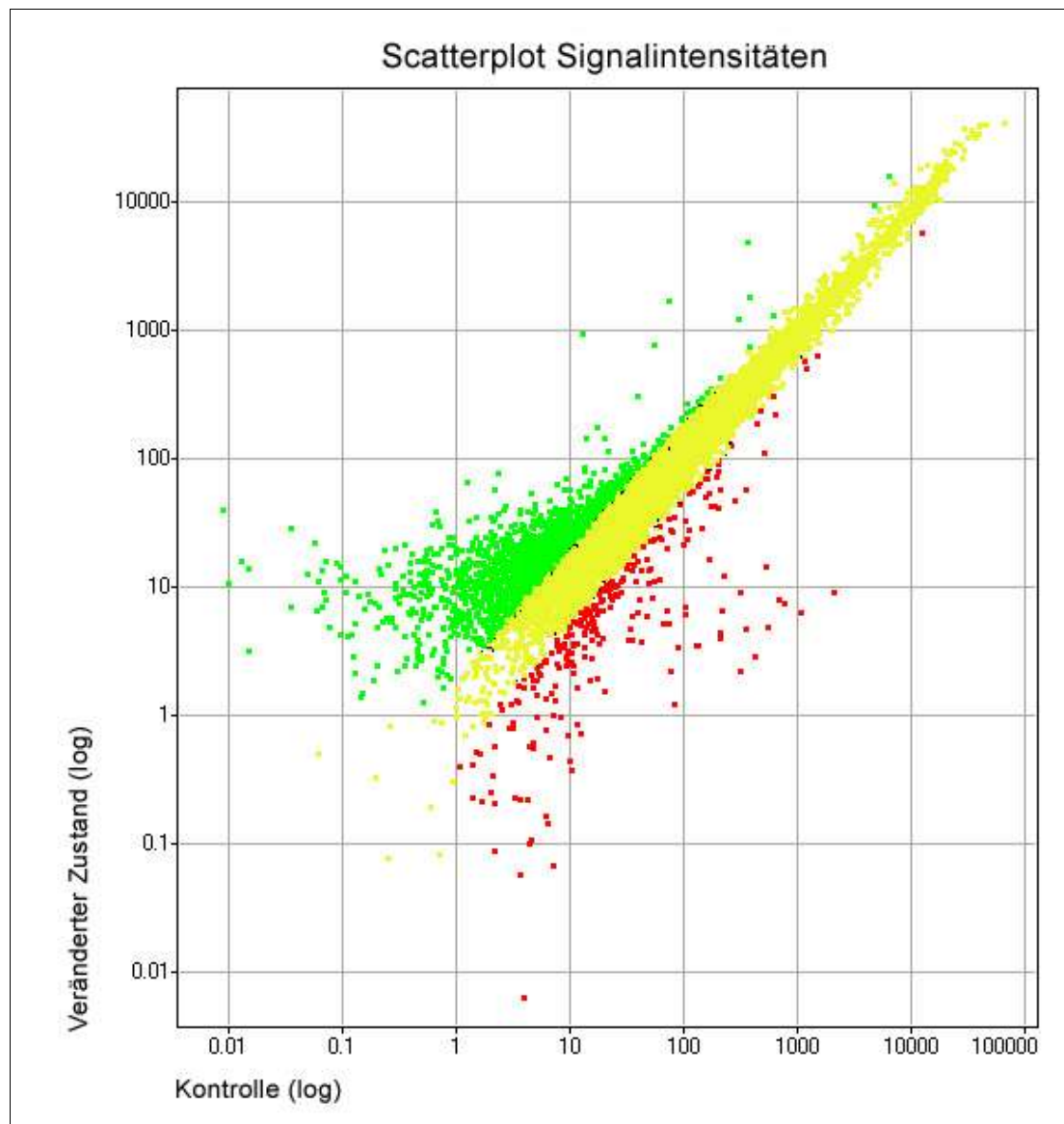
Je höher die Expression eines Gens ist, desto größer ist sein Anteil in der markierten RNA und dementsprechend stärker ist das zugehörige Hybridisierungssignal auf dem DNA-Chip [14]. Verwendet man für die Markierung zweier Proben unterschiedliche Fluoreszenzfarbstoffe, die aufgrund ihrer Excitations- und Emissionsspektren unabhängig voneinander detektiert werden können, kann man die beiden Proben in einer kompetitiven Hybridisierung einsetzen. Zu diesem Zweck werden die beiden mRNA-Spezies, z.B. aus verschiedenen Zelltypen oder Entwicklungsstadien mit unterschiedlichen Farbstoffen markiert und gleichzeitig auf einem Microarray hybridisiert. In einer einzelnen Hybridisierung können somit zwei unterschiedliche mRNA-Spezies direkt quantitativ verglichen werden. An die Hybridisierung der markierten Proben mit korrelierenden Sonden schließt sich eine Detektion und Quantifizierung der erhaltenen Signalintensitäten sowie eine ausführliche computergetützte Datenanalyse an [15,16].



**Abb.1.4:** Ablauf der Genexpressionsanalyse mit Microarrays. A) Zwei unterschiedliche RNA-Proben werden enzymatisch mit unterschiedlichen Farbstoffen markiert (B+C). Die Proben werden gemeinsam auf einen Microarray hybridisiert (D). Es folgt die Detektion der Signalintensitäten beider Farbstoffe mittels eines Lasers und die Generierung von Falschfarbenbildern durch Überlagerung beider gescannten Farbkanäle (E).

Zunächst werden die Daten normalisiert [17], um die Hybridisierungen direkt miteinander vergleichbar zu machen und Unterschiede zwischen einzelnen Hybridisierungen auszugleichen. Gründe für die Ungleichheit können ungleiche Ausgangsmengen von RNA oder cDNA, unterschiedliche Effizienz der Markierungsreaktion, Unterschiede im Hybridisierungsverhalten oder abweichendes Signalverhalten bei der Detektion sein. Da die Ergebnisse der Experimente statistisch validiert werden müssen erfolgt meist ein mehrmalige Wiederholung der kompetitiven Hybridisierungen und daher auch eine Normalisierung zwischen verschiedenen Microarrays [18]. Werden in einem Experiment zwei Zustände miteinander verglichen, z.B. Kontrolle und veränderter Zustand, können die Resultate der Datenanalyse in einem zweidimensionalen Koordinatensystem aufgetragen werden. Dabei werden die normalisierten Signalintensitäten des veränderten Zustandes gegen die normalisierten Signalintensitäten des Kontrollzustandes aufgetragen. Der daraus

resultierende Faktor für die Induktion bzw. Repression eines Gens ergibt sich aus dem Verhältnis der Signalintensitäten (veränderter Zustand / Kontrollzustand) [19, 20].



**Abb.1.5:** Schema der Genexpression. Auftrag der Signalintensitäten zweier RNA-Proben. Grüne Punkte entsprechen Genen, die im veränderten Zustand stärker exprimiert werden als im Kontrollzustand. Gelbe Punkte zeigen gleich stark exprimierte Gene. Rote Punkte repräsentieren Gene, die im veränderten Zustand schwächer exprimiert werden als in der Kontrolle.

Die Gene, die in beiden Bedingungen gleich stark exprimiert werden, besitzen die gleichen Signalintensitäten und liegen in dem Koordinatensystem auf der Ursprungsgeraden mit einer Steigung von eins. Gene, die in dem veränderten Zustand der Zelle stärker exprimiert werden als in der Kontrollbedingung liegen oberhalb; Gene, die schwächer exprimiert werden, unterhalb dieser Geraden. Für die Normalisierung und nachfolgende Kontrolle der Hybridisierungsqualität wurde die Software M-CHIPS [21] eingesetzt, die in der Abteilung Funktionelle Genomanalyse am DKFZ entwickelt wurde. Die Software nutzt die Korrespondenzanalyse [22], eine Projektionsmethode, die die Verhältnisse zwischen den analysierten Genen und den Hybridisierungen darstellt. Die Gene, die in einer Bedingung

herauf- oder herunterreguliert werden, sind von besonderem Interesse, da sie die Veränderung der Genexpression repräsentieren. Um Muster in der Genexpression unter verschiedenen Bedingungen und gleichmäßig regulierte Gruppen von Genen zu identifizieren, werden sogenannte Clusteranalysen verwendet, die ihren Ursprung meist in der statistischen Datenanalyse haben. Zu diesen Methoden, Genexpressionsdaten zu analysieren, gehören das hierarchische Clustering [23], k-means Clustering [24], Self-Organizing Maps [25] oder die Principal Component Analyse [26].

## 1.4 Die Grundlage des Heidelberg FlyArray

Die Zahl der Gene des Genoms von *Drosophila melanogaster* wurde in einer ersten Annotation des „Berkeley *Drosophila* Genome Project“ (BDGP) auf 14.000 geschätzt [1]. Einige Vergleiche dieser Annotation mit bekannten Proteinsequenzen [27], *in silico* Daten [28] sowie neuen „Expressed Sequence Tags“ (ESTs) [29,30] ließ jedoch an der Genauigkeit dieser vorhergesagten Genzahl zweifeln. Die Annotationen des BDGP basieren im wesentlichen auf EST-Daten und Homologiekriterien aus *ab initio* Genvorhersagen von Genie [31] und GENSCAN [32], weshalb hauptsächlich bereits bekannte Gene, oder Gene mit großer Ähnlichkeit zu bekannten Proteinen entdeckt wurden. Gene, die für Proteine unbekannter Funktion kodieren, können daher „übersehen“ werden. Eine Möglichkeit dies zu umgehen, ist der Genom-Vergleich mit verwandten Spezies wie *Anopheles gambiae* [33] und *Drosophila pseudoobscura* [34] unter der Voraussetzung daß genügend Gene zwischen den Arten im Laufe der Evolution konserviert wurden.

Um die Limitierungen der Genvorhersagen zu umgehen, sollte ein Microarray, der nahezu das gesamte Transkriptom von *Drosophila* abdeckt, entwickelt werden. Zu diesem Zweck wurde zuerst eine aus der BDGP *Drosophila* Genomannotation Version 2.0, der BDGP cDNA Kollektion Version 1.0 [35] und aus *ab initio* Genvorhersagen der Fgenesh-Software [36] kombinierte Genannotationsliste erstellt. Die Fgenesh-Software benutzt unter anderem einen weniger stringenten Algorithmus für die Definition der 5'- und 3'-Termini eines Exons als die obengenannten Algorithmen, weshalb der vorhergesagte Genatz aus 20.622 Genen bestand. Dieser Satz wurde mit den BDGP Annotationen kombiniert unter der Voraussetzung, daß Genvorhersagen die eine Überlappung der Exonsequenzen von über 30% aufweisen, das gleiche Gen repräsentieren. Die endgültige Genliste wurde als „Heidelberg Prediction“ (HDC) bezeichnet und enthielt 21.396 potentielle Gene, wovon 7.464 Gene einzig aus der Fgenesh-Vorhersage entstammten und nicht in der BDGP Annotation vorhanden waren. Daß die Gene der Heidelberg-Vorhersage nur zu einem geringen Teil konservierte Gene darstellen, zeigte ein Vergleich mit der Sequenz von *Drosophila pseudoobscura*, wobei nur ca. 13% der Gene eine Sequenzübereinstimmung von mehr als 50% aufwiesen [37]. Dies bedeutet, daß die Gene der Heidelberg-Vorhersage kaum

Homologien mit konservierten DNA-Sequenzen oder Proteinmotiven zeigen, was umgekehrt bedeuten könnte, daß die Gene für *Drosophila melanogaster*-spezifische neuartige Proteine-, oder bisher nicht entdeckte Proteinfamilien kodieren könnten. Auf Grundlage der Heidelberg prediction wurden daher mit Hilfe der GenomePride-Software [38] spezifische PCR-Primer für 21.306 potentielle Gene konstruiert. Diese Software erlaubt das Design spezifischer Primer für das komplette Transkriptom eines Organismus durch eine Strategie, die in zwei Phasen gegliedert ist: Zuerst wird in der Exonstruktur eines Gens die optimale Zielregion zwischen den Exons ermittelt und ein Amplikon von benutzerdefinierter Größe, in unserem Fall 500 bp, berechnet. In einem zweiten Schritt wird die optimale Position beider Primer für das Amplikon ermittelt, wobei die 5'- und 3'- terminalen Regionen ausgeschlossen werden, so das die PCR-Amplifikation an zentraler Position innerhalb des selektierten Exons erfolgen kann. Auf diese Weise wurde die Heidelberg Kollektion R1 generiert, die je ein Primerpaar für die 21.306 Gene beinhaltet. Auf Grundlage dessen, sollten die PCR-Amplifikationen durchgeführt werden, um einen Microarray zu konstruieren der mögliche neue Gene identifiziert und Genexpressionsanalysen in der Entwicklung von *Drosophila melanogaster* ermöglicht.

## 2 Material und Methoden

### 2.1 Material

#### 2.1.1 Chemikalien und Reagenzien

Laborchemikalien, allgemein (Reinheitsgrad: höchst möglich)	Merck, Darmstadt
	Sigma, Deisenhofen
	Serva, Heidelberg
	Fluka, Neu-Ulm
	Roth, Karlsruhe
Agarose	Sigma, Deisenhofen
Agarose, niedrigschmelzend	Biozym, Oldendorf
Aminoallyl-dUTP	Sigma, Deisenhofen
Ampicillin (50 mg/ml)	Serva, Heidelberg
Bacto Tryptone	Difco, Detroit, MI, USA
Bacto Yeast Extract	Difco, Detroit, MI, USA
Betain, Monohydrat	Sigma, Deisenhofen
Cy3-monoreaktiver Ester	Amersham Biosciences, Freiburg
Cy5-monoreaktiver Ester	Amersham Biosciences, Freiburg
Desoxy-Ribonukleotid-5'-Triphosphate (je 100 mM)	MBI Fermentas, St. Leon-Rot
Diethylpyrocarbonat (DEPC)	Roth, Karlsruhe
Isopropyl- $\beta$ -D-thiogalactopyranosid (IPTG)	MBI Fermentas, St. Leon-Rot
Trizma <sup>®</sup> Base	Sigma, Deisenhofen
Trizma <sup>®</sup> Hydrochlorid	Sigma, Deisenhofen
Glucose	Merck, Darmstadt
Ethylendiamintetraessigsäure (EDTA)	Merck, Darmstadt
Dithiothreitol (DTT)	Merck, Darmstadt
Phenylmethylsulfonylfluorid (PMSF)	Sigma, Deisenhofen
Nonidet P-40	Sigma, Deisenhofen
Tween 20	Gerbu Biotechnik GmbH, Gaiberg
Glycerin 87% (v/v)	Fluka, Deisenhofen
Essigsäure	Merck, Darmstadt
Adenosine 5'-Triphosphat (100 $\mu$ M)	MBI Fermentas, St. Leon-Rot
Cytidine 5'-Triphosphat (100 $\mu$ M)	MBI Fermentas, St. Leon-Rot

Guanosine 5'-Triphosphat (100 µM)	MBI Fermentas, St. Leon-Rot
Uridine 5'-Triphosphat (100 µM)	MBI Fermentas, St. Leon-Rot
Hoechst 33258	Sigma, Deisenhofen
Kalbsthymus DNA, genomisch	Sigma, Deisenhofen
Ethanol	Riedel-de Haen, Seelze
Fischsperma-DNA, DNA MB grade (10 mg/ml)	Roche Diagnostics, Mannheim
Natriumdodecylsulfat (SDS)	Gerbu Biotechnik GmbH, Gaiberg
Fetales Kälberserum („BSA“)	Sigma, Deisenhofen
Hexamer-Primer pd(N) <sub>6</sub> (45 U)	Amersham Biosciences, Freiburg
FluoroLink™ Cy3-dCTP	Amersham Biosciences, Freiburg
FluoroLink™ Cy5-dCTP	Amersham Biosciences, Freiburg
Dimethylsulfoxid (DMSO)	Fluka, Deisenhofen
Kresolrot (o-Kresolsulfonphthalein)	Sigma, Deisenhofen

### 2.1.2 Längenmarker für die Gelelektrophorese

GeneRuler™ 100 bp DNA Marker	MBI Fermentas, St. Leon-Rot
GeneRuler™ 1kb DNA Marker	MBI Fermentas, St. Leon-Rot
RNA Ladder	MBI Fermentas, St. Leon-Rot

### 2.1.3 Enzyme

<i>Taq</i> DNA-Polymerase (5 U/µl)	Qiagen, Hilden
<i>Taq</i> DNA-Polymerase, selbstgemacht	
Lysozym (28262)	Serva Electrophoresis GmbH, Heidelberg
T3 RNA-Polymerase (20 U/µl)	MBI Fermentas, St. Leon-Rot
DNase I (10 U/µl)	Roche Diagnostics, Mannheim
<i>Mbo</i> I (10 U/µl)	MBI Fermentas, St. Leon-Rot
Klenow-Fragment, 3'-5' Exo <sup>-</sup>	NEB, Schwalbach
SuperScript® II Reverse Transkriptase	Life Technologies, Karlsruhe

### 2.1.4 Reagenziensätze („Kits“)

Dynabeads® Oligo(dT) <sub>25</sub>	Dynal AS, Oslo, Norwegen
RiboGreen® RNA-Quantifizierungs Kit	Molecular Probes, Eugene, OR, USA
QIAquick® PCR Purification Kit	Qiagen, Hilden

### 2.1.5 *Drosophila melanogaster* genomische DNA und Gesamt-RNA

Stamm: Oregon <sup>R</sup> , wildtyp	ZMBH Universität Heidelberg, Heidelberg
--------------------------------------	---

### 2.1.6 Bakterienstämme und Vektoren

pTP4 <i>E.coli</i> DH5α Stamm coll. Nr.81	J. D. Hoheisel, DKFZ Heidelberg
<i>Escherichia coli</i> , Stamm XL1-Blue	Stratagene, Heidelberg
pBluescript® II SK+ mit poly(A/T) <sub>50</sub> Insert	M. Scheideler, DKFZ Heidelberg

### 2.1.7 Oligonukleotide

#### ***Amplifikation der Heidelberg***

#### ***Collection V1***

Gen-spezifischer Primersatz, Eurogentec, Seraign, Belgien  
 21.306 Primerpaare (100µM), Primersequenzen als Tabelle auf CD  
 gelöst in TE-Puffer

Universeller Primersatz, 10 Interactiva, Ulm  
 Primer (1universeller + 9  
 unterschiedliche Primer)  
 (100µM), gelöst in TE-Puffer



Uni-Tag	5'-TGG CGC CCC TAG ATG
Tag-1	5'-CGA CGC CCG CTG ATA
Tag-2	5'-TAG GTC TAG CCC CGC
Tag-3	5'-GGC GTC TGA CGC TAG
Tag-4	5'-CGC ATG TAG CCT GCC
Tag-5	5'-TAG CCT CCC TAG CGC
Tag-6	5'-AGC CCC TAA CGT GCG
Tag-7	5'-AGC TAG CCC GGG TGA
Tag-8	5'-TGA CCC GGG TAG CCT
Tag-9	5'-TGA GGT ACG CGT GGG

### **Amplifikation externer**

#### **Kontrollen: *A.thaliana* cDNA**

#### ***A.thaliana* universelle Primer**

T7 forward	5'-TAA TAC GAC TCA CTA TAG GG
M13 reverse	5'-TCA CAC AGG AAA CAG CTA TGA C

### ***A.thaliana* spezifische**

#### **Primer**

arab1 forward	5'-GGA ATT CCA GCT GAC CAC CAT GGA TAT TAG CCG GCG TGA AAG T
arab1 reverse	5'-GAT CCC CGG GAA TTG CCA TGT CCC TAA ACC GAG TCT TCG AT
arab2 forward	5'-GGA ATT CCA GCT GAC CAC CAT GCG GTT ACA CAT ATC CGG TTT C
arab2 reverse	5'-GAT CCC CGG GAA TTG CCA TGT TGA TGT TTG TAG ACG CCC A
arab3 forward	5'-GGA ATT CCA GCT GAC CAC CAT GCA CGC GTC CGA TTT TTT T
arab3 reverse	5'-GAT CCC CGG GAA TTG CCA TGC ACA ACG GAA ATG AAT AAC CC
arab4 forward	5'-GGA ATT CCA GCT GAC CAC CAT GCG TCC GAA ATC AAG TGC TAA
arab4 reverse	5'-GAT CCC CGG GAA TTG CCA TGA GCG CAA CGC AAT TAA TGT G
arab5 forward	5'-GGA ATT CCA GCT GAC CAC CAT GCG TCC GGT GAG AGA AGA AAG A
arab5 reverse	5'-GAT CCC CGG GAA TTG CCA TGA CTG CTT CGA TGC CTT CTC TT
arab6 forward	5'-GGA ATT CCA GCT GAC CAC CAT GCG GAA CTA GTC TCA CCA GAC A
arab6 reverse	5'-GAT CCC CGG GAA TTG CCA TGG CTG ATT TCG AAA CTC AGA GA
arab7 forward	5'-GGA ATT CCA GCT GAC CAC CAT GTC CGT TTT CTC TTG GGC TAC T
arab7 reverse	5'-GAT CCC CGG GAA TTG CCA TGC AGT TTG CTT GGT GAC AAG T
arab8 forward	5'-GGA ATT CCA GCT GAC CAC CAT GTC TCC TCA CGC TTA TAG AGC A
arab8 reverse	5'-GAT CCC CGG GAA TTG CCA TGT AAC TAG CTT CTC GCA GCT TG

## 2.1.8 Lösungen und Puffer

### Allgemein

2YT-Medium	1,6% (w/w)	Bacto Tryptone
	1% (w/w)	Bacto Yeast Extract
	0,5% (w/w)	NaCl
	96,9% (w/w)	H <sub>2</sub> O autoklaviert
Puffer A	50 mM	Tris-HCl, pH 8,0
	50 mM	Glucose
	1 mM	EDTA, pH 8,0
Prälyse-Puffer +	4 mg/ml	Puffer A Lysozym
Lyse-Puffer	10 mM	Tris-HCl, pH 8,0
	50 mM	KCl
	1 mM	EDTA, pH 8,0
	1 mM	PMSF
	0,5% (v/v)	Tween 20
	0,5% (v/v)	Nonidet P40
Dialyse-Puffer	50 mM	Tris-HCl, pH 8,0
	50 mM	KCl
	0,1 mM	EDTA, pH 8,0
	1 mM	DTT
	0,5 mM	PMSF
	50% (v/v)	Glycerin
10x PCR-Puffer	0,1 mM	Tris-HCl, pH 8,3
	0,5 mM	KCl
DEPC-Wasser	0,1% (v/v)	DEPC inkubiert über Nacht und autoklaviert
TE-Puffer	100 mM	Tris-HCl, pH 8,0
	10 mM	EDTA, pH 8,0
6x DNA-Probenpuffer	0,25%	Bromphenolblau

	0,25%	Xylencyanol
	30%	Glycerin in H <sub>2</sub> O
50x TAE-Puffer	0,4 M	Tris Base
	0,4 M	Essigsäure
	20 mM	EDTA, pH 8,0
10x HMFM-Medium	3 mM	MgSO <sub>4</sub> x 7 H <sub>2</sub> O
	15 mM	Tri-Natriumcitrat x 2 H <sub>2</sub> O
	70 mM	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
	55% (w/v)	Glyzerin ad 800 ml H <sub>2</sub> O, autoklaviert
	270 mM	KH <sub>2</sub> PO <sub>4</sub>
	130 mM	K <sub>2</sub> HPO <sub>4</sub> ad 200 ml H <sub>2</sub> O, autoklaviert
	beide Lösungen	vereinigen
10x TNE-Puffer	100 mM	Tris-HCl, pH 7,4
	10 mM	EDTA
	2 M	NaCl
20x SSC	3 M	NaCl
	300 mM	Natriumcitrat, pH 7,0
Hexamer-Primer pd(N) <sub>6</sub>	1,0 mg/ml	gelöst in TE-Puffer
<b>DNA Microarrays</b>		
NaPO <sub>4</sub> -Puffer	600 mM	Na <sub>2</sub> HPO <sub>4</sub>
	600 mM	NaH <sub>2</sub> PO <sub>4</sub> eingestellt auf pH 8,5
1x Spotting solution	3x	SSC
	150 mM	NaPO <sub>4</sub> -Puffer
	1,5 M	Betain (sterilfiltriert)
Test Spotting solution (1x)		1x Spotting solution
+	200 ng/μl	Fischsperma-DNA (geschert)
+	0,1 mM	Kresolrot

Rinsing Solution 1	0,1% (v/v)	SDS
Prähybridisierungs-Puffer	5x	SSC
	0,1% (v/v)	SDS
	1% (w/v)	BSA
Phosphat-Puffer	1 M	K <sub>2</sub> HPO <sub>4</sub>
	1 M	KH <sub>2</sub> PO <sub>4</sub> eingestellt auf pH 8,5
Phosphatwasch-Puffer	5 mM	Phosphat-Puffer
	80% (v/v)	Ethanol
Phosphatelutions-Puffer	4 mM	Phosphat-Puffer
Waschlösung A	2x	SSC
	0,2%	SDS
Waschlösung B	2x	SSC
Waschlösung C	0,2x	SSC

### Kommerzielle Lösungen

RNA-Ladepuffer	Sigma, Deisenhofen
10x Qiagen PCR-Puffer (inkl. 15 mM MgCl <sub>2</sub> )	Qiagen, Hilden
10x Restriktionspuffer R+	MBI Fermentas, St. Leon-Rot
10x Klenow-Puffer	NEB, Schwalbach
5x Erststrang-Puffer	Life Technologies, Karlsruhe
Slide Hyb-buffer#1	Ambion, Huntingdon, UK

### 2.1.9 Geräte

Zentrifuge „J-6B“	Beckman Coulter, Unterschleissheim
-------------------	------------------------------------

Thermocycler PTC-200	MJ Research Inc., Waltham, MA, USA
Elektrophoresenetzgerät E835	Hofer, San Francisco, CA, USA
Geldokumentationssystem Gel Doc 1000	BioRad, München
Multiscreen PCR-Filter Vakuumsystem	Millipore, Eschborn
Fluorometer FLUOstar Galaxy	BMG LabTechnologies, Offenburg
Zentrifuge „Biofuge pico“	Heraeus, Hanau
Zentrifuge „Megafuge 1.0R“	Heraeus, Hanau
Zentrifuge „Sorvall® Super T21“	Kendro Laboratory Products, Hanau
Zentrifuge „2K15“	Sigma Laborzentrifugen, Osterode
MicroGrid II Array-Roboter	BioRobotics, Cambridge, UK
UV-Crosslinker UVC 500	Hofer, San Francisco, CA, USA
Hybridisierungsofen, rotierend	H. Saur Laborbedarf, Reutlingen
Vakuumkonzentrator	H. Saur Laborbedarf, Reutlingen
Photometer Ultraspec 2000	Amersham Biosciences, Freiburg
ScanArray 5000	Perkin Elmer, Boston, MA, USA

### 2.1.10 Verbrauchsmaterial

Dialyseschlauch Servapor® (44145.01)	Serva Electrophoresis GmbH, Heidelberg
PCR-Platte, 96-well, PP	Steinbrenner, Neckargemünd
Multiplate™ 96, V-bottom, PP	Nalge Nunc, Rochester, NY, USA
Multiscreen Manu 30 PCR-Filterplatte, 96-well	Millipore, Eschborn
ELISA 8-well F-Strips, schwarz	Greiner, Solingen
384er Microtiterplate with lid, V-bottom, PS	Genetix, Christchurch, UK
UVette® 220-1600 nm	Eppendorf, Hamburg
QMT Amino Slides-Objektträger	Quantifoil, Jena über PeqLab, Erlangen
Deckgläser LifterSlips® (22 x 60 mm)	Menzel Gläser, Braunschweig
Poly-L-Lysin-Objektträger, selbstgemacht	DKFZ, Heidelberg

### 2.1.11 Materialien

Elektrophoresekammer, horizontal, groß	PeqLab Biotechnologie, Erlangen
--	---------------------------------

Elektrophoresekammer, horizontal, mini	Renner, Dannstadt
12-Kanal-Pipette Biohit Proline®	Biohit, Helsinki, Finnland
Kanüle, Microlance™, 1/2“, 0,9x40mm	Becton Dickinson, Franklin Lakes, NJ, USA
SMP3 Microarray-Nadeln	TeleChem, Sunnyvale, CA, USA
Atlas® Hybridisierungskammer	BD Biosciences Clontech, Palo Alto, CA, USA
Hybridisierungskammern	TeleChem, Sunnyvale, CA, USA

## 2.1.12 Software und Internetadressen

J-Express Pro™	<a href="http://www.molmine.com">www.molmine.com</a>	Molmine AS, Bergen, Norwegen
R statistical language 1.7.1	<a href="http://www.r-project.org">www.r-project.org</a>	
GenomePride®	<a href="http://pride.molgen.mpg.de">pride.molgen.mpg.de</a>	MPI für molekulare Genetik, Berlin
MCHiPS	<a href="http://www.mchips.org">www.mchips.org</a>	DKFZ, Heidelberg
GNU/Linux Debian 3.0	<a href="http://www.debian.org">www.debian.org</a>	
PostgreSQL 7.3.1	<a href="http://www.postgresql.org">www.postgresql.org</a>	
Python 2.2	<a href="http://www.python.org">www.python.org</a>	
Tomcat 4.1	<a href="http://jakarta.apache.org/tomcat">jakarta.apache.org/tomcat</a>	
Java Server Pages™ 1.2	<a href="http://java.sun.com/products/jsp">java.sun.com/products/jsp</a>	Sun Microsystems, Santa Clara, CA, USA
Java 2™ 1.4	<a href="http://java.sun.com">java.sun.com</a>	Sun Microsystems, Santa Clara, CA, USA
GenePix Pro™ 4.0	<a href="http://www.axon.com">www.axon.com</a>	Axon Instruments, Union City, CA, USA
Matlab® 6.0	<a href="http://www.mathworks.com">www.mathworks.com</a>	MathWorks, Natick, MA, USA
R/maanova 0.91	<a href="http://www.jax.org/staff/churchill/absite/software/anova/rmaanova">www.jax.org/staff/churchill/absite/software/anova/rmaanova</a>	Jackson Lab., Bar Harbor, Mn, USA
Flybase	<a href="http://www.flybase.org">www.flybase.org</a>	
HeidelbergFlyArray	<a href="http://HDFlyArray.zmbh.uni-heidelberg.de/cgi-bin/gbrowse">HDFlyArray.zmbh.uni-heidelberg.de/cgi-bin/gbrowse</a>	
Gene Ontology™ Consortium	<a href="http://www.geneontology.org">www.geneontology.org</a>	
Gene Expression Omnibus (GEO)	<a href="http://www.ncbi.nlm.nih.gov/geo">www.ncbi.nlm.nih.gov/geo</a>	

## 2.2 Methoden

### 2.2.1 PCR-Amplifikation der Heidelberg Kollektion R1

#### **Herstellung von Taq DNA-Polymerase**

Die Herstellung rekombinanter *Taq* DNA-Polymerase erfolgte durch Animpfen von 5 ml 2YT-Medium (mit 100 µg/ml Ampicillin) mit einer Kolonie des pTP4 *E. coli DH5 α* Stammes und Inkubation über Nacht bei 37°C in einem Schüttler. Je 500 µl der Kultur wurden in einen 2 l Schüttelkolben mit 1 l 2YT-Medium (mit 80 µg/ml Ampicillin) überführt und bei 37°C und 200 rpm bis zu einem OD<sub>600</sub>-Wert von ca. 0,8 geschüttelt. Zur Induktion der *Taq* DNA-Polymerase-Expression wurden je Kolben 625 µl 20% (w/v) IPTG zugegeben. Daraufhin erfolgte eine weitere Inkubation der Zellen für 12 h bei 37°C und 200 rpm. Die Suspension wurde zum Ernten der Zellen für 10 min bei 4500 rpm (Beckman J6-B, 4°C) zentrifugiert. Der Überstand wurde verworfen und das Zellpellet in 50 ml Puffer A resuspendiert. Die Zellen wurden erneut für 10 min bei 4500 rpm und 4°C zentrifugiert und das Zellpellet anschließend in 25 ml Prälyse-Puffer resuspendiert und für 15 min bei RT stehengelassen. Nach der Zugabe von 30 ml Lyse-Puffer wurde die Suspension für 1 h bei 75°C im Schüttler lysiert. Die Lösung wurde dann in Corex-Gläsern bei 27000x g (15000 rpm) für 10 min zentrifugiert. Das klare Lysat (~240 ml) wurde in einen Erlenmeyerkolben gegeben und über 1 h unter ständigem Rühren bei RT wurden die Proteine durch Zugabe von 72 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (zerstoßen) ausgefällt. Die Suspension wurde bis zur vollständigen Lösung des (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> für weitere 30 min gerührt. Danach erfolgte eine Zentrifugation bei 15000 rpm für 10 min bei 4°C. Das Pellet wurde in Puffer A (1 ml / 5 ml Lysat) resuspendiert. Anschließend wurde die Proteinlösung (ca. 20 ml) in zwei Dialyseschläuche (Serva) gefüllt und in 2x 12 h in 1 l Dialyse-Puffer bei 4°C unter Rühren inkubiert. Alle 8 h wurde der Dialyse-Puffer ausgetauscht und abschließend wurde die Proteinlösung 1:1 mit Dialyse-Puffer verdünnt und bei -70°C aufbewahrt.

#### **PCR-Amplifikation mit Gen-spezifischen Primern („First-step PCR“)**

Die mit Hilfe der Software GenomePride errechneten Oligonukleotidsequenzen wurden von der Firma Eurogentec synthetisiert und mit einer Konzentration von 100 µM pro well in 96-well Platten ausgeliefert. Alle Primersequenzen beinhalteten neben den spezifischen Sequenzen einen universellen von 10 unterschiedlichen Tag-Sequenzen von je 15 bp Länge an ihrem 5'-Terminus. Aus den jeweiligen 100 µM für Forward- und Reverse-Primer pro Amplikon wurde ein 2,5 µM Primer-Mix hergestellt. Die PCR-Amplifikation [39,40] mit genomischer *Drosophila*-DNA erfolgte in 96-well PCR-Platten in je 50 µl Reaktionsvolumen.

Reagenzien	1x50 µl	Endkonzentration
10x PCR-Puffer	5 µl	1x
25 mM MgCl <sub>2</sub>	4 µl	2 mM
dNTPs (je 2,5 mM)	0,8 µl	40 µM
genomischeDNA (1µg/µl)	0,1 µl	100 ng
Taq-Polymerase	0,4 µl	
H <sub>2</sub> O	31,7 µl	

**Tab. 1: Schema für PCR-Amplifikation mit spezifischen Primern**

Der jeweilige Reaktionsansatz wurde vorgelegt und dazu wurden jeweils 8 µl des spezifischen Primer-Paares (2,5 µM) gegeben. Das PCR-Programm wurde mit einer ersten Denaturierung bei 94°C für 5 Minuten gestartet, dann folgten 15 Zyklen mit einer Denaturierungsphase bei 94°C für 30 s, eine Annealingphase bei 65°C für 30 s mit einer Temperaturverringerng um -1°C pro Zyklus, eine Elongationsphase bei 72°C für 1 min 30 s. Nachfolgend starteten 20 Zyklen mit einer Denaturierungsphase bei 94°C für 30 s, eine Annealingphase bei 50°C für 30 s, eine Elongationsphase bei 72°C für 1 min 30 s mit einer Zeiterhöhung um +5 s pro Zyklus. Dann erfolgte eine Elongationsphase bei 72°C für 10 min und abschließend ein Kühlphase bei 4°C für 10 min.

### **PCR-Amplifikation mit universellen Primern („Second-step PCR“)**

Die zweite PCR diente zur Amplifikation des ersten PCR-Produktes. Um Kreuzkontaminationen zu verhindern wurden die 10 universellen Primer (100µM) so miteinander gemischt, daß alle 9 Wells einer PCR-Platte fortlaufend unterschiedliche Primerkombinationen (10 µM) beinhalteten. Die PCR-Reaktion erfolgte in 96-well Platten zu je 100 µl Reaktionsvolumen. Es wurden je 1 µl der ersten Gen-spezifischen PCR-Produkte und 4 µl des universellen Primergemisches (10 µM jeder) vorgelegt.

Reagenzien	1x 100 µl	Endkonzentration
10x Qiagen PCR-Puffer	10 µl	1x
dNTPs (je 2,5 mM)	4 µl	100 µM
Qiagen Taq-Polymerase (5 U/µl)	0,4 µl	0,02 U/µl
H <sub>2</sub> O	80,6 µl	

**Tab. 2: Schema für PCR-Amplifikation mit universellen Primern**

Das PCR-Programm wurde mit einer ersten Denaturierung bei 94°C für 5 min initiiert. Es folgten 15 Zyklen mit einer Denaturierungsphase bei 94°C für 30 s, eine Annealingphase bei 60°C für 30 s, dann eine Elongationsphase bei 72°C für 1 min 30 s. Die nächsten 20 Zyklen bestanden aus einer Denaturierungsphase bei 94°C für 30 s, einer



Annealingphase bei 60°C für 30 s und einer Elongationsphase bei 72°C für 1 min 30 s mit einer Zeiterhöhung von +5 s pro Zyklus. Danach erfolgte eine Elongationsphase bei 72°C für 10 min und eine abschließende Kühlphase bei 4°C für 10 min.

### **Kontrolle der Heidelberg Kollektion R1 PCR-Amplifikationen**

Alle PCR-Reaktionen wurden mit Hilfe eines Elektrophoresesystems von PeqLab Biotechnologie GmbH durchgeführt. Es war aufgrund der Gelkammergröße möglich, 2x 96-PCR, das entspricht 192 Proben, auf ein Gel aufzutragen. Es wurde jeweils ein 2%iges Agarosegel (Agarose, niedrighschmelzend von Biozym Diagnostik) verwendet. Es wurden jeweils 2 µl der Proben mit 6x Ladepuffer versetzt und mit einer 12-Kanal Pipette (Biohit) aufgetragen. Zusätzlich wurde zur Kontrolle der Fragmentgrößen pro Gel 4 x 1kb DNA-Längenstandard (MBI Fermentas) mitaufgetragen. Die Probenaufftrennung erfolgte in 1x TAE-Puffer bei 10 V/cm für 25 min. Die Qualität der Fragmente wurde aufgrund ihrer zu erwartenden Größe beurteilt. Fehlende Amplifikate, falsche Größen oder Mehrfachbanden wurden wiederholt, wenn die jeweilige 96-well PCR-Platte weniger als 90% positive Fragmente aufwies. Von jeder Gelelektrophorese wurde ein Foto gemacht und zu Dokumentationszwecken in Laborjournalhefte eingeklebt.

## **2.2.2 Herstellung heterogener Kontrollen für die Microarray-Analyse**

### **PCR-Amplifikation der Kontrollen**

Als heterogene Kontrollen unterschiedlicher Organismen, die auf den Microarray aufgetragen werden sollten, dienten 8 cDNA-Fragmente von *Arabidopsis thaliana*. Diese 8 selektierten cDNAs stellen pflanzentypische Gene dar und zeigten bei anderen Microarray-Hybridisierungen mit tierischen Proben keine Kreuzhybridisierungseffekte. Die PCR-Amplifikationen wurden direkt mit transformierten *E. coli* XL1-Blue Stämmen, die unterschiedlich große cDNA-Inserts in einem pBluescript® II SK+ mit poly(A/T)<sub>50</sub> Vektor (Vektorkonstrukte von Sonja Bastuck) tragen, durchgeführt. Durch die Kombination des Vektors mit zwei verschiedenen Primertypen und nachfolgender PCR konnten unterschiedliche Amplifikate, einmal als Probe für den Microarray und andererseits als Vorlage für eine mRNA-Synthese, generiert werden.

Kontrolle	cDNA	Primer Name	Neuer Primer Name	Größe (bp) Amplifikate für Array	Größe (bp) Amplifikate für mRNA
<i>A.thaliana</i> 6	F10E12T7	EST_P_21265	arab 1	1281	359
<i>A.thaliana</i> 7	F10F2T7	EST_P_21275	arab 2	1820	346
<i>A.thaliana</i> 9	240I17T7	EST_P_20661	arab 3	888	334
<i>A.thaliana</i> 18	240J12T7	EST_P_20665	arab 4	597	515
<i>A.thaliana</i> 19	240J16T7	EST_P_20666	arab 5	1520	491
<i>A.thaliana</i> 23	240K2T7	EST_P_20931	arab 6	881	427
<i>A.thaliana</i> 28	240L16T7	EST_P_20935	arab 7	1875	446
<i>A.thaliana</i> 30	240L23T7	EST_P_20937	arab 8	1147	431

**Tab. 3: *A.thaliana* cDNAs für die PCR-Amplifikation von Kontrollen**

Die Tabelle 3 zeigt die als Kontrolle ausgesuchten cDNAs. Für die Verwendung als Kontrollsystem wurden neue Primernamen eingeführt. Die PCR-Amplifikation erfolgte in 100 µl Reaktionsvolumen. Je 2 µl der spezifischen Forward- und Reverse Primer (10 µM) wurden für die PCR eingesetzt um Amplifikate zu generieren, die als Probe für den Microarray dienten. 2 µl des universellen Primer T7 (Forward) und 2 µl M13 (Reverse) Primer (je 10 µM) wurden verwendet um PCR-Fragmente zu amplifizieren, aus denen mittels *in-vitro* Transkription mRNA synthetisiert wurde.

Reagenzien	1x 100 µl	Endkonzentration
10x PCR-Puffer	10 µl	1x
25 mM MgCl <sub>2</sub>	6 µl	1,5 mM
dNTPs (je 2,5 mM)	8 µl	200 µM
Taq-Polymerase	0,1 µl	
H <sub>2</sub> O	72,9 µl / 70,9 µl	

**Tab. 4: Schema für die PCR-Amplifikation der Kontrollen**

Das Reaktionsgemisch wurde direkt mit je 1 µl des betreffenden Bakterienklones inokuliert und wie folgt amplifiziert: eine Denaturierungsphase bei 94°C für 3 min, 35 Zyklen bestehend aus einer Denaturierung bei 94°C für 1 min, einem Annealing bei 53°C für 30 s, einer Elongation bei 72°C für 1 min. Darauf folgte eine Elongationsphase bei 72°C für 10 min und eine nachfolgende Abkühlungsphase bei 4°C für 10 min. Nach der Aufreinigung der PCR-Amplifikate (siehe 2.2.2) wurden je 50 µl in 384er-Platten (V-bottom, Genetix) pipettiert und getrocknet. Anschließend wurden die Kontrollen in je 10 µl 1x Spotting-solution resuspendiert. Zusätzlich wurden PCR-Amplifikate der Kontrollen gereinigt, ihre Konzentration fluorimetrisch bestimmt und Lösungen mit den Konzentrationen 50 ng/µl, 200 ng/µl und 500 ng/µl in 1x Spotting-solution hergestellt.

Die Bakterienklone wurden zur Aufbewahrung in eine Dauerkultur bestehend aus 1x HMFM-

Medium mit Ampicillin (50 µg/ml) übertragen.

### **Aufreinigung der PCR-Produkte**

Die amplifizierten PCR-Produkte wurden mit 96-well Multiscreen PCR-Filterplatten (Millipore) aufgereinigt, um überschüssige Primer und nicht-inkorporierte Nukleotide zu entfernen. Dazu wurde der PCR-Ansatz in die Filterplatte pipettiert und mit Hilfe des Vakuum Filtrationsystems (Millipore) wurden die Proben getrocknet (800 mbar, 10 min). Daraufhin wurden die Proben 2x mit 30 µl DEPC-Wasser gewaschen, in 60 µl DEPC-Wasser aufgenommen und für 15 min auf einem Schüttler inkubiert. Das gereinigte PCR-Produkt wurde in eine 96-well Mikrotiterplatte (Nunc) überführt und bei RT über Nacht getrocknet. Der Erfolg der PCR wurde durch eine Agarosegelelektrophorese überprüft.

### **Herstellung externer Kontroll-mRNA durch *in vitro* Transkription**

Die PCR-Produkte der *A.thaliana* cDNA, die mit den universellen M13 und T7 Primern amplifiziert wurden, dienen zur Herstellung der externen Kontroll-mRNA durch die *in vitro* Transkription. Es wurden jeweils 500 ng der aufgereinigten PCR-Produkte (Messung der DNA-Konzentration siehe 2.2.5) mit 1x Transkriptionspuffer (MBI Fermentas), 10 µl NTPs (je 10 mM) und 1,5 µl T3-RNA-Polymerase (MBI Fermentas) zusammengegeben und bei 37°C für 2 h inkubiert. Es folgte eine Deaktivierung des Enzyms bei 70°C und eine anschließende Spaltung der DNA durch 1 µl DNase (10 U/µl) (Roche) bei 37°C für 15 min.

Die synthetisch hergestellte mRNA wurde mit Oligo (dT)<sub>25</sub>-Magnetpartikeln (Dynabeads®, Dynal) aufgereinigt. Dazu wurden 2,5 mg der Magnetpartikel in 200 µl Dynal-Bindungspuffer resuspendiert und mit 200 µl der RNA-Lösung gemischt, die vorher bei 65°C für 2 min erhitzt wurde. Die Lösung wurde für 5 min unter Schütteln inkubiert und in einem Magnetseparator getrennt. Die Magnetpartikel wurden zweimal mit jeweils 400 µl Dynal-Waschpuffer mit Hilfe des Separators gewaschen. Die Partikel wurden dann in 20 µl Dynal-Elutionspuffer resuspendiert und bei 85°C für 2 min inkubiert. Anschließend erfolgte die Trennung der eluierten RNA-Lösung von den Magnetpartikeln durch den Separator und die RNA wurde bei -80°C aufbewahrt.

### **2.2.3 Bestimmung der RNA-Konzentration**

Die Konzentration von RNA wurde fluorometrisch mit Hilfe des RiboGreen® RNA-Quantifizierungs Kit ermittelt. Dazu wurde die RiboGreen-Konzentration für den Probenpuffer im oberen („high range buffer“) Meßbereich (20 ng/ml – 1 µg/ml) verwendet.

Der RiboGreen-Farbstoff wurde dazu im Verhältnis 1:200 in TE-Puffer verdünnt. Als Standard diente eine ribosomale RNA mit einer Konzentration von 2 µg/µl. Die Verdünnung erfolgte für verschiedene Konzentrationen, wie in Tabelle 5 dargestellt. Die Proben wurden entsprechend mit TE-Puffer und RiboGreen verdünnt (1 µl Probe / 200 µl high range buffer). Die Verdünnungen wurden in ELISA 8-well F-Strips, schwarz (Greiner) angesetzt und gemischt. Es folgte eine Inkubation der Proben bei RT für 5 min unter Lichtabschluß. Die Fluoreszenzmessung erfolgte mittels eines FLUOstar Galaxy Fluorometer (BMG) bei einer Anregungswellenlänge von 480 nm und einer Emissionswellenlänge von 520 nm.

<b>20 ng/ml – 1 µg/ml RNA</b>	<b>high range buffer</b>		
TE (µl)	2 µg/ml RNA (µl)	RiboGreen (µl)	RNA (ng/ml)
0	100	100	1000
50	50	100	500
90	10	100	100
95	5	100	50
100	0	100	Leerwert

**Tab. 5: Verdünnungen des RiboGreen-Farbstoffes.**

#### **2.2.4 RNA-Agarosegelelektrophorese (denaturierend)**

Die Qualität der RNA wurde mit einer vereinfachten Agarosegelelektrophorese durchgeführt, wobei für den Probenauftrag ein denaturierender Formamid-Puffer verwendet wurde. Vor der Gelelektrophorese wurde das Zubehör in 3% (v/v) H<sub>2</sub>O<sub>2</sub> für 1 h inkubiert. Es wurden je 100 ng Probe mit 5x RNA-Ladepuffer (Sigma) versetzt, bei 65°C für 10 min denaturiert und für 5 min auf Eis gekühlt. Die Probe wurde dann zusammen mit einem RNA-Standard (RNA-Ladder, MBI Fermentas) auf ein 1,2%iges Agarosegel aufgetragen. Anschließend erfolgte eine Gelelektrophorese für 45 min bei 10 V/cm.

#### **2.2.5 Bestimmung der DNA-Konzentration**

Die DNA-Konzentration wurde ebenfalls fluorometrisch bestimmt. Die Quantifizierung der DNA erfolgte mittels des DNA-interkalierenden Farbstoffes Hoechst 33258. Der Farbstoff (1 mg/ml) wurde dazu in 1x TNE-Puffer 1/2000 verdünnt. Um den oberen dynamischen Meßbereich (high range) von DNA-Konzentrationen zu erfassen die zwischen 100 ng/ml und 5000 ng/ml liegen, wurden 4 DNA-Standardkonzentrationen (Kalbsthymus DNA, Sigma) definiert, die in 1x TNE-Puffer gelöst und mit Hoechst 33258 markiert wurden. Als Leerwert wurde 1x TNE-Puffer mit Hoechst 33258 verwendet. Es wurden 2x (Doppelprobe) 200 µl

jeder Verdünnung in ELISA 8-well F-Strips (Greiner) pipettiert und zusammen mit dem Standard in einem FLUOstar Galaxy Fluorometer (BMG) gemessen. Die Anregungswellenlänge betrug dabei 352 nm und die Emissionswellenlänge 520 nm.

### **2.2.6 Herstellung komplexer genomischer Kontrollen für die Microarray-Analyse**

Als Positivkontrolle für die Hybridisierungen auf dem Microarray wurden Proben mit genomischer *Drosophila*-DNA verwendet. Die Restriktion der DNA erfolgte in einem Volumen von 30 µl. Dazu wurden je 6 µg genomischer DNA mit 3 µl 10x Buffer R+ (MBI Fermentas) und 2 µl *Mbo* I (10 u/µl) bei 37°C für 2h inkubiert. Es folgte eine Inaktivierung des Restriktionsenzym bei 65°C für 20 min. Die DNA wurde mittels DNA-Präzipitation in 2x vol. Ethanol (100%) und 2 M Ammoniumacetat präzipitiert und nach Zentrifugation (40 min bei 14.000 rpm, 4°C) durch waschen mit 70% Ethanol ausgefällt und getrocknet. Nach einer Konzentrationsmessung im Photometer bei 260 nm in einer UV-Küvette (Uvette, Eppendorf) wurde das DNA-Pellet in 1x Spotting-solution aufgenommen und Lösungen zu je 50 ng/µl in die 384er-Platten (V-bottom, Genetix) pipettiert.

### **2.2.7 Herstellung von Positionsmarkern für die Microarray-Analyse**

Um Positionsmarker („Landmark-Spots“) für je einen Block auf dem Microarray zu generieren, wurde Fischsperma-DNA (Roche) bei 70°C für 30 min erhitzt und mittels einer Kanüle (0,9x40mm, Becton Dickinson) geschert. Von dieser Lösung wurden je 2 µg DNA für die Fluoreszenzmarkierung verwendet (siehe 2.2.9). Nach der Aufreinigung der markierten DNA wurde diese eingetrocknet (3 Tage bei RT) und anschließend in 1 ml 1x Spotting-solution resuspendiert und entsprechend dem Array-Layout zu je 10 µl in die 384er-Platten (V-bottom, Genetix) pipettiert. Durch den Auftrag der markierten Proben auf den Microarray konnten so die Blockpositionen („Grids“) genau zugeordnet werden.

### **2.2.8 Herstellung der cDNA-Microarrays („Heidelberg FlyArray“)**

#### **Vorbereitung der PCR-Produkte**

Die ausgewählten PCR-Produkte in 96-well PCR-Platten, die mit universellen Primern erfolgreich amplifiziert werden konnten, wurden für die weitere Verwendung aufgeteilt. Dazu wurden je 50 µl aus den Platten entnommen und in 384er-Platten (V-bottom, Genetix) überführt. Zu diesem Zweck wurden je 4 96-Platten reihenweise versetzt (12 Kanalpipette

Biohit) in einer 384er Platte kombiniert. Die Platten wurden abgedeckt, bei RT für ca. 2 Wochen eingetrocknet und anschließend in je 10 µl 1x Spotting-solution aufgenommen. Die Platten wurden um 2 384er-Kontrollplatten ergänzt und für 1 min bei 1000 rpm (Megafuge 1.0R, Heraeus) zentrifugiert.

### **Auftragung der DNA („Spotting“) auf Glasobjektträger**

Die DNA wurde auf aminosilanisierte QMT Amino slides (Quantifoil) mit Hilfe des Microarray-Roboters *MicroGrid II*<sup>®</sup> (BioRobotics) aufgetragen. Es wurden 2 x 12 spezielle Nadeln („Pins“) (SMP3 pins, TeleChem) mit einem Flüssigkeitsreservoir („Splitpins“) für den Auftrag verwendet. Diese Nadeln fassen ein Volumen von ca. 100 nl und durch den DNA-Auftrag in 1x Spotting-solution wurde ein Spot-Durchmesser von ca. 80 µm generiert. Der Abstand zwischen dem Mittelpunkt der Spots betrug 140 µm und der Übertrag fand bei RT und 30 – 40 % Luftfeuchtigkeit statt.



**Abb.2.1:** Der Microarray-Roboter MicroGrid II<sup>®</sup> zur Herstellung der DNA-Chips.

### **Nachbehandlung („Processing“) der cDNA-Microarrays**

Nach dem Auftrag der DNA wurden die Microarrays in Objektträgerkästen bei RT in einem Excikator mit Trockenmittel für mindestens 12 h gelagert. Um die DNA-Fläche später eindeutig zu identifizieren, wurde der Array mit einem Glasschreiber markiert. Für die Hybridisierungsexperimente erfolgte die Entnahme der Microarrays, die dann in einem UV-Crosslinker mit einer Energie von 250 x 100 µJ/cm<sup>2</sup> behandelt wurden. Anschließend wurden die Microarrays bei 80°C für 4 h in einem Hybridisierungssofen erhitzt.

Dann wurden die Objektträger bei RT für 10 s in „Rinsing solution 1“ (Quantifoil) unter starkem Schütteln gewaschen. Es folgte ein Waschschrift bei RT für 10 s in H<sub>2</sub>O (bidest) und danach ein Erhitzen der Arrays bei 95°C für 3 min. Die Prähybridisierung zur Blockierung der aktivierten Glassoberfläche der Microarrays erfolgte bei 55°C für 45 min mit 20 ml

Prähybridisierungs-Puffer in Atlas™ Hybridisierungskammern (BD Biosciences). Anschließend wurden die Microarrays bei RT für 10 s in H<sub>2</sub>O (bidest) gewaschen und sofort mit Druckluft (Stickstoffgas) getrocknet. Die so behandelten DNA-Chips waren dann sofort für die Hybridisierungsexperimente einsetzbar.

## 2.2.9 Fluoreszenzmarkierung von Nukleinsäuren

### **Fluoreszenzmarkierung von DNA**

Doppelsträngige DNA wurde unter Einbau von dCTP-Cy5 bzw. -Cy3 mit Hilfe des „Random Priming“ fluoreszenzmarkiert. In einem 50 µl Reaktionsansatz wurden 2 µg dsDNA, 6 µl Hexamer-Primer (pd(N)<sub>6</sub>, 1,0 µg/µl), 5 µl 10x Klenow-Puffer gemischt und mit H<sub>2</sub>O auf 43 µl aufgefüllt. Der Ansatz wurde daraufhin für 5 min bei 95°C erhitzt und sofort für 5 min auf Eis abgekühlt. Es folgte die Zugabe von 2,5 µl dNTPs (dATP, dGTP, dTTP jeweils 2,5 mM, dCTP 0,4 mM), 2 µl Cy5- bzw. Cy3-dCTP (1mM) und 2,5 µl Klenow-Fragment (3'-5' Exo<sup>-</sup>, 4 U/µl) wodurch die Reaktion gestartet wurde. Es folgte eine Inkubation der Lösung für 12 h bei 37°C. Die Reaktion wurde dann durch Zugabe von 2 µl 0,5 M EDTA abgestoppt.

### **Fluoreszenzmarkierung von RNA durch cDNA-Erststrangsynthese (Reverse Transkription)**

Die Fluoreszenzmarkierung der RNA-Proben erfolgte in zwei Schritten, wobei zuerst aus der Gesamt-RNA durch reverse Transkription mit Aminoallyl-markiertem dUTP (aa-dUTP) die cDNA synthetisiert wurde und anschließend erfolgte die Kupplung eines monoreaktiven Fluoreszenzfarbstoff-NHS-Esters an die Amino-Gruppe des eingebauten Aminoallyl-dUTP.

Es wurden für die Reaktion 20 µg Gesamt-RNA (isoliert aus *Drosophila*) mit je 8 ng der acht Arabidopsis mRNA-Kontrollen gemischt („spiking“) und in einem Vakuum-Konzentrator eingetrocknet. Das Pellet wurde in je 9,5 µl DEPC- H<sub>2</sub>O und 9,0 µl Hexamer-Primern (pd(N)<sub>6</sub>, 1,0 µg/µl) aufgenommen, für 10 min bei 70°C inkubiert und für 2 min auf Eis abgekühlt. Es folgte die Zugabe von 6 µl 5x Erststrang-Puffer, 0,6 µl 50x dNTP-Mix (je 25 mM dATP, dCTP, dGTP, 10 mM aa-dUTP, 15 mM dTTP), 3 µl DTT (0,1 M), 2,0 µl SuperScript II RT (200 U/µl) und 3 µl DEPC- H<sub>2</sub>O. Die reverse Transkription wurde über Nacht bei 42°C durchgeführt. Die Reaktion wurde durch Zugabe von 10 µl 0,5 M EDTA gestoppt und die RNA durch Zugabe von 10 µl 1 N NaOH und nachfolgender Inkubation für 15 min bei 65°C hydrolytisch gespalten. Es folgte eine Neutralisierung mit 10 µl 1 M HCl. Um nicht eingebaute Nukleotide und Puffer zu entfernen wurde ein PCR Reinigungskit (QIAquick® PCR Purification Kit, Qiagen) verwendet. Zu dem Reaktionsansatz wurde 300 µl Puffer PB gegeben und der gesamte Ansatz in eine QIAquick-Säule pipettiert. Es folgte eine Zentrifugation für 1 min, das Filtrat wurde verworfen. Anschließend wurde die Säule 2x mit je

750 µl Phosphatwasch-Puffer gewaschen und der Durchfluss verworfen. Es folgte eine Zentrifugation ohne Lösung und die gebundene cDNA wurde daraufhin mit 30 µl Phosphateluions-Puffer eluiert. Das Eluat mit der aa-dUTP markierten cDNA wurde in einem Vakuum-Konzentrator für ca. 30 min getrocknet.

Für die Kupplung der Farbstoffester an die aa-dUTP markierte cDNA wurde ein Aliquot eines Farbstoffes in 72 µl DMSO gelöst und auf 16 Aliquots zu je 4,5 µl verteilt. Nach der Trocknung der Aliquots im Vakuum-Konzentrator wurden diese bei -80°C unter Lichtausschluß gelagert.

Für die Kupplungsreaktion wurde je ein Aliquot der gelagerten monoreaktiven NHS-Farbstoffester (Cy5 bzw. Cy3) in 4,5 µl DMSO gelöst und zu der cDNA gegeben, die zuvor in 4,5 µl NaHCO<sub>3</sub> (pH 9,0) aufgenommen wurde. Der Ansatz wurde gemischt und für 2 h bei RT im Dunkeln inkubiert. Die Farbstoff-markierte cDNA wurde aufgereinigt (siehe 2.2.10) um nicht gebundenen Farbstoffester zu entfernen. Nach der Reinigung wurde die Konzentration und Einbaurrate des Farbstoffes der markierten cDNA photometrisch (siehe 2.2.11) gemessen und die cDNA erneut im Vakuum-Konzentrator getrocknet. Die getrocknete cDNA konnte direkt für die Hybridisierung verwendet werden, oder die Proben wurden bei -20°C lichtgeschützt für max. 3 Wochen aufbewahrt.

### **2.2.10 Reinigung von markierten Nukleinsäuren**

Markierte cDNA- oder DNA wurde mithilfe des QIAquick® PCR Purification Kit gereinigt. Der jeweilige Ansatz wurde mit 35 µl 100 mM Natriumacetat (pH 5,2) und 250 µl Puffer PB versetzt. Die Lösung wurde auf ein QIAquick-Säulchen pipettiert und für 1 min zentrifugiert. Dann wurde die Säule mit 750 µl Puffer PE gewaschen und erneut für 1 min zentrifugiert. Das Filtrat wurde verworfen und erneut mit 400 µl Puffer PE gewaschen. Nach erneuter Zentrifugation wurde das Filtrat verworfen und die Säule noch einmal für 1 min zentrifugiert. Die cDNA bzw. DNA wurde durch Zugabe von zweimal 30 µl Puffer EB eluiert, wobei der Puffer jeweils 5 min auf der Säule belassen wurde und anschließend für 1 min zentrifugiert wurde.

### **2.2.11 Photometrische Bestimmung der Einbaurrate von Cy5/Cy3**

Die Effizienz des Einbaus von modifizierten Nukleotiden in die cDNA erfolgte durch die photometrische Bestimmung des Anteils der eingebauten Fluoreszenzfarbstoffe Cy5 und Cy3 an der Gesamtmenge der cDNA. Zu diesem Zweck wurden die eluierten cDNA-Proben im Verhältnis 1:1 mit H<sub>2</sub>O verdünnt und in eine UVette® (Eppendorf) pipettiert. Die Messung



erfolgte im Photometer bei den Wellenlängen 260 nm, 550 nm und 650 nm.

Nach dem Lambert-Beer'schen Gesetz und den folgenden Extinktionskoeffizienten  $\epsilon$  für:

$$\text{cDNA: } \epsilon_{260} = 10000 \text{ l/mol}\cdot\text{cm}$$

$$\text{Cy3: } \epsilon_{550} = 150000 \text{ l/mol}\cdot\text{cm}$$

$$\text{Cy5: } \epsilon_{650} = 250000 \text{ l/mol}\cdot\text{cm}$$

wurden die jeweiligen Konzentrationen der cDNA und der beiden Farbstoffe bestimmt. Die Absorption für die cDNA bei 260 nm wurde mit einem Korrekturfaktor berechnet, da bei dieser Wellenlänge eine gleichzeitige Absorption der Farbstoffmoleküle erfolgt:

$$\text{Für Cy3: } A_{260 \text{ kor.}} = A_{260} - (0,08 \times A_{550})$$

$$\text{Für Cy5: } A_{260 \text{ kor.}} = A_{260} - (0,05 \times A_{650})$$

Durch die nachfolgenden Formeln konnte über das Verhältnis der Konzentrationen die Einbaurrate (in %) für den jeweiligen Farbstoff Cy3 bzw. Cy5 ermittelt werden:

$$(c[\text{Cy3}] / c[\text{cDNA}]) \times 100 = \% \text{ Einbau Cy3}$$

$$(c[\text{Cy5}] / c[\text{cDNA}]) \times 100 = \% \text{ Einbau Cy5}$$

### 2.2.12 Hybridisierung auf cDNA-Microarrays

Die Hybridisierungen wurden jeweils mit zwei unterschiedlich markierten Proben (Cy5 und Cy3) auf einem Microarray durchgeführt. Zu diesem Zweck wurden die beiden Proben in 5  $\mu\text{l}$  10 mM EDTA vereinigt und für 5 min bei 95°C denaturiert. Dann wurden die cDNA-Proben für 5 min bei RT abgekühlt und zentrifugiert. Die mit ddH<sub>2</sub>O gereinigten Deckgläschen („LifterSlips<sup>®</sup>“, Menzel Gläser) wurden mit dem Abstandshalter nach unten auf die DNA-Fläche des Microarrays gelegt. Der Microarray befand sich bereits in der Hybridisierungskammer (TeleChem), in deren Aussparungen zuvor je 30  $\mu\text{l}$  3 X SSC zwecks Befeuchtung der Kammer pipettiert worden waren. Zur Hybridisierung wurde „Slide Hyb-buffer#1“ (Ambion) als Hybridisierungs-Puffer verwendet. Dieser wurde 15 min vor Verwendung bei 68°C vorgewärmt. Die Proben wurden in 35  $\mu\text{l}$  Slide Hyb-buffer#1 aufgenommen und direkt auf den Microarray neben die Kante des Deckglases pipettiert, so daß die Hybridisierungslösung durch die Kapillarwirkung gleichmäßig unter das Deckglas gezogen wurde. Danach wurde die Hybridisierungskammer fest mit dem Deckel verschlossen und die Hybridisierung erfolgte in einem Wasserbad für 16 h bei 55°C. Anschließend wurde der Microarray entnommen und zur Ablösung des Deckglases in Waschlösung A (Quantifoil) hin- und herbewegt. Danach wurde der Objektträger für 10 min in Waschlösung A, für 10 min in Waschlösung B (Quantifoil) und anschließend für 10 min in

Waschlösung C (Quantifoil) gewaschen. Der Microarray wurde dann sofort mit Druckluft (Stickstoffgas) getrocknet. Bis zur Signaldetektion wurden die Microarrays bei RT in Boxen unter Lichtausschluß aufbewahrt.

### **2.2.13 Auswertung der Hybridisierungsdaten**

#### ***Signaldetektion***

Die Messung der Cy3- und Cy5-markierten Proben auf den DNA-Chips erfolgte mit dem konfokalen Laserscanner ScanArray 5000 (Perkin Elmer). Der Scanner verwendet einen roten und einen grünen Helium-Neon Laser der die durch Cy5 erzeugten Signale mit einer Anregungswelle von 633 nm und einer Emissionswellenlänge von 670 nm detektiert. Die durch Cy3 erzeugten Signale wurden mit einer Anregungswelle von 543 nm und einer Emissionswellenlänge von 570 nm detektiert. Die Laserstärke und Photomultiplerverstärkung wurden jeweils für die Farbstoffsignale auf jedem Microarray angepaßt. Zu diesem Zweck wurden die externen Arabidopsis-mRNA-Kontrollen verwendet, die definierte Konzentrationen jeweils nur des Cy3- und des Cy5-Signals, sowie ein 1:1-Verhältnis beider Farbstoffe repräsentierten. Als Abgleich der Wellenlängen für die einzelnen Farbstoffe wurde die entsprechende Signalamplitude als 100% Sättigung (Wert max.: 65535, min.: 0) definiert und der Ratiowert beider Farbstoffe als 50%.

Als erster Qualitätsfilter wurde die Standardabweichung (SD) des Farbstoffverhältnisses (Cy5/Cy3) der beiden Replikate für jedes Gen auf dem Microarray genutzt. Es wurden nur Hybridisierungen weiterverwendet, bei denen mindestens 30% aller Gene auf dem Array die Standardabweichung des  $\log(632/532)$  zwischen den Replikaten weniger als ein Drittel betrug. Das photoinstabilere Cy5 wurde jeweils vor dem Cy3 Farbstoff gemessen. Für jeden der beiden Farbstoffe wurde ein separates 16-bit TIFF-Format Bild erzeugt.

#### ***Quantifizierung der gemessenen Signalintensitäten***

Die Quantifizierung erfolgte mit der Software GenePix™ 4.0 (Axon). Die Bilder wurden dazu in Falschfarben konvertiert, wobei Cy3 die Farbe Grün und Cy5 die Farbe Rot zugeordnet wurde. Wird ein Gen in beiden Proben in gleicher Menge transkribiert, so ergibt sich die Farbe Gelb als eine Mischung aus Grün und Rot, da das entsprechende Signal in beiden Laserkanälen die gleiche Intensität besitzt. Differentiell exprimierte Gene werden somit grün oder rot oder in allen möglichen Mischfarben zwischen Grün und Rot dargestellt, entsprechend ihrer Transkriptionshäufigkeit in den beiden Proben. Die Quantifizierung der Signale wurde ermöglicht, indem die Software beide Bilder überlagert und anschließend ein Raster über das Falschfarbenbild gelegt wurde. Die Anpassung dieses Rasters erfolgte manuell und semi-automatisch. Die Intensitätswerte der Farbstoffe wurden dann für jedes

Signal berechnet und die erzeugten Werte wurden in einem Tabellenformat (GPR-Format) gespeichert. Neben den Signalintensitätswerten (Mittelwert und Median der Pixel) wurden zusätzlich Hintergrundintensitäten, Pixelzahl, Standardabweichung und weitere Daten für jedes Gen („Spot“) in der Tabelle gespeichert.

### ***Normalisierung und Datenanalyse mit M-CHiPS***

Die Softwareumgebung M-CHiPS („Multi-Conditional Hybridization Intensity Processing Software“) implementiert in Matlab® (MathWorks), wurde am DKFZ für die Analyse von Expressionsdaten auf DNA-Microarrays entwickelt. mit Hilfe der Software wurden die Daten normalisiert und verschiedene Filtermethoden auf die Datensätze angewendet. Die gefilterten Daten wurden dann in der Korrespondenzanalyse [22] eingesetzt. Bei der Korrespondenzanalyse handelt es sich um eine statistische Methode um Beziehungen zwischen Variablen zu untersuchen, z.B. ob in einer Tabelle (oder Matrix) ein Verhältnis zwischen Spalten (Hybridisierungsbedingungen) und Reihen (Genen) besteht. Ähnlich anderen Projektionsmethoden repräsentiert die Korrespondenzanalyse Variablen wie Genexpressionsdaten als Vektoren in einem mehrdimensionalem Raum. Dabei läßt sich jedes Gen in Anbetracht der Bedingungen in einem Hybridisierungs-dimensionalem – und jede Hybridisierung in einem Gen-dimensionalem Raum gleichzeitig darstellen. Daraufhin wird dieser Raum unter Reduzierung der effektiven Dimensionalität ohne signifikanten Informationsverlust auf eine zweidimensionale Ebene reduziert. Der Grad des Informationsverlustes für jede Korrespondenzanalyse wird über die Darstellung der Varianz der Prinzipiellen Komponenten („principal components“) errechnet. Durch die Darstellung der Varianzen läßt sich der beste „Blickpunkt“ auf die Daten und damit die geeignetste Trennung der Daten bezüglich der Bedingungen ermitteln. Die Korrespondenzanalyse benötigt keine vorherigen Parameterangaben und läßt damit eine unverfälschte Analyse der Datenstrukturen zu. Aus diesem Grund lassen sich Experimente mit mehreren Bedingungen gleichzeitig visualisieren und die Qualität der Hybridisierungen überprüfen. Zuerst erfolgte eine Normalisierung der Signalintensitäten. Aufgrund unterschiedlicher Faktoren wie z.B. generierte cDNA-Menge, unterschiedliche Einbauraten bei der Markierung der RNA, unterschiedliche Photolabilität der Fluoreszenzfarbstoffe und Qualität der Hybridisierungen kommt es zu Unterschieden in den Experimenten die biologisch nicht relevant sind, aber trotzdem das Ergebnis beeinflussen. Um die verschiedenen Hybridisierungsbedingungen untereinander vergleichbar zu machen, wurden die Signalintensitätswerte durch eine lineare Regression normalisiert. Dazu wurden die Intensitätswerte der zu untersuchenden Bedingung gegen die der Kontrollbedingung in einem zweidimensionalen Koordinatensystem („Bi-plot“) aufgetragen. Wurden die Experimente für eine Bedingung mehrfach wiederholt, so wurde der jeweilige Median aus den betreffenden Bedingungen

gebildet und gegen den Median aller Datensätze der Kontrollbedingung aufgetragen. Die Regressionsgerade wurde mit  $y = mx + b$  bestimmt und die Geradensteigung durch eine multiplikative Korrektur auf den Wert 1 errechnet. Durch eine additive Korrektur des Y-Achsenabschnittes (Hintergrundintensität) verlief die Regressionsgerade durch den Ursprung des Koordinatensystems. Zur Berechnung der Regressionsgeraden wurden die Signalintensitäten der Mehrheit der Gene hinzugezogen anstatt heterologer Kontrollen oder konstitutiv exprimierter Gene. Ein für jeden Kanal ermittelter Hintergrundwert (Hintergrund-Mittelwert  $\times$  Spot-Pixelanzahl) wurde von jedem Signalintensitätswert abgezogen. Nach der Normalisierung erfolgte eine Filterung der Signalintensitäten nach 3 verschiedenen Kriterien:

#### Signalintensität:

Um zu verhindern daß Signale resultierend aus unspezifischer Kreuzhybridisierung, oder Signale mit geringer Intensität die gleich der Hintergrundintensität waren, mit in die Analyse aufgenommen wurden, wurde ein Schwellenwert definiert. Der Wert lag oberhalb der normalisierten Intensitäten, um unspezifische und schwache Signale herauszufiltern.

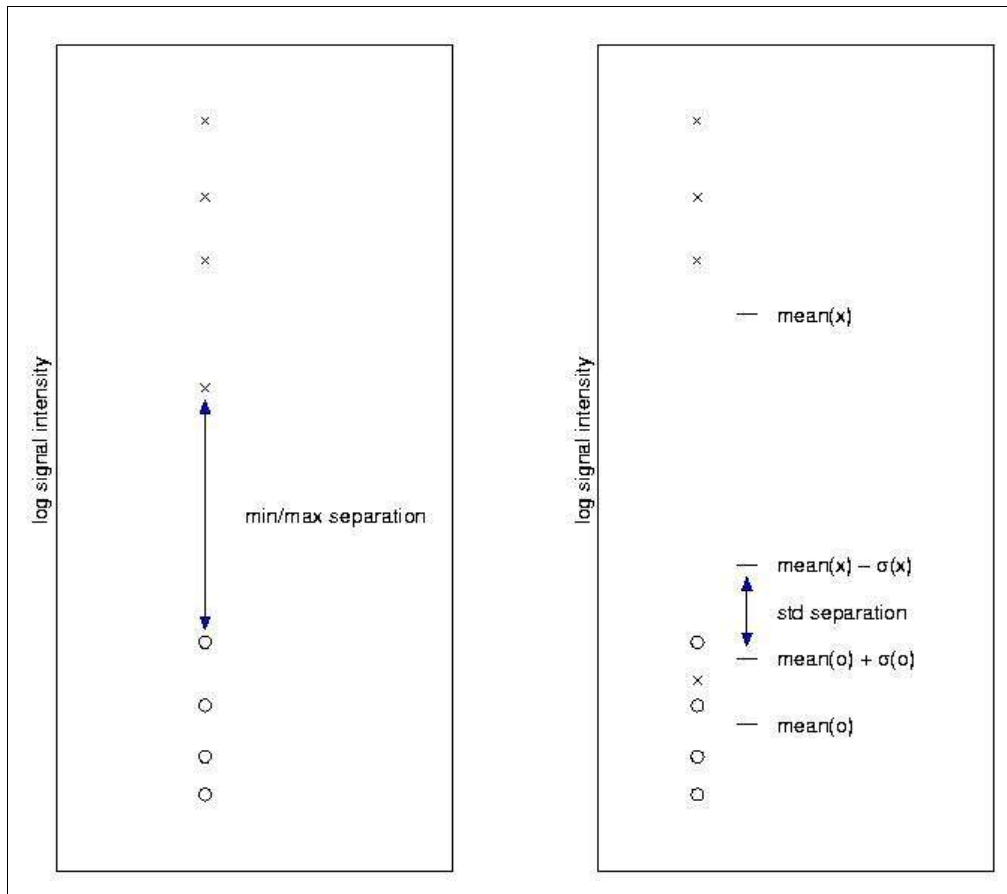
#### Verhältnis der Intensitäten von Cy5 zu Cy3:

Das Verhältnis (Ratio) von Cy5-markierter und Cy3-markierter Probe gibt an, in welchem Maß ein Gen in einer Bedingung exprimiert wird oder nicht. Es wurden die Signale herausgefiltert, deren Expression sich um weniger als einen definierten Mittelwert der bestimmten Bedingung gegenüber der Kontrolle änderte. Damit wurden Gene, die nicht differentiell reguliert worden waren ausgeschlossen.

#### Reproduzierbarkeit:

mit Hilfe zweier Methoden unterschiedlicher Stringenz (Min/Max-Trennung und Standardabweichung-Trennung) wurde die reproduzierbare signifikante Expressionsänderung für ein Gen in einer bestimmten Bedingung ermittelt. Dabei werden alle Hybridisierungs-Wiederholungen einer Bedingung für ein Gen auf reproduzierbare Expression -oder Repression hinzugezogen. Wenn das Gen in einer Bedingung stärker exprimiert wird, so müssen alle Datenpunkte (siehe Abb. 2.2) aus den Wiederholungen in der gleichen Bedingung stärker exprimiert sein als z.B. in der Kontrollbedingung. Es wird ein Differenzwert angegeben, der den Minimalwert  $\min(x)$  der Bedingung mit höheren Intensitäten (x) von dem Maximalwert  $\max(o)$  der Kontrollbedingung mit niedrigeren Intensitäten (o) trennt. Diese Methode ist stringenter als die Verwendung der Standardabweichung-Trennung, bei der die Standardabweichungen  $\sigma(x,o)$  von den jeweiligen Mittelwerten  $\bar{M}(x,o)$  addiert oder subtrahiert werden.

Bei der Datenanalyse wurde ausschließlich die Min/Max-Trennung verwendet.



**Abb.2.2:** Min/Max- und Standardabweichung-Trennung. Filtermethoden für die Beurteilung der Reproduzierbarkeit bei signifikanter Änderung der Genexpression.

### **Statistische Analyse der differentiellen Genexpression**

In Microarray-Experimenten repräsentieren die markierten RNA-Proben, auch wenn sie gemeinsam hybridisiert wurden, zwei unterschiedliche Bedingungen. Aus den beiden Signalintensitätswerten für jede Probe wird die Genexpressionsdifferenz ermittelt. Dieses Verhältnis gibt Aufschluss ob ein Gen in einer Bedingung reprimiert oder exprimiert wird.

Ab welchem Differenzwert aber ist ein Gen differentiell reguliert [41] ? Die meist verwendete Methode der 2-fachen Differenz („two-fold difference“) besitzt keine statistische Grundlage [62]. Um signifikante Änderungen der Genexpressionsdifferenz zu detektieren, wurde für den Vergleich mehrerer Bedingungen die Varianzanalyse („Analysis of variance“, ANOVA) [42] verwendet. Dazu wurde die Hypothese formuliert: „Welche Gene sind im Verlauf der Entwicklung bei *Drosophila* signifikant differentiell reguliert?“.

Wird die Nullhypothese auf die Experimente angewendet, so bedeutet dies, daß keine Differenz in der Expression zwischen den Bedingungen (Entwicklungsstadien bei *Drosophila*) besteht und das Verhältnis zwischen jedem Gen in den Bedingungen ist gleich eins. Ist die Expression unterschiedlich, so ist das Verhältnis ungleich eins. In der

Varianzanalyse wird die Expression eines Gens in einer Probe relativ zu einem Mittelwert des Gens in allen Proben des Experimentes berechnet.

Für die Varianzanalyse wurde ein „gemischtes“ Model („mixed-effect ANOVA“)[43] eingesetzt. Dafür wurde die Software R/maanova® 0.91 in der R-Statistiksprache verwendet. Aus den gefilterten Genen der MCHiPS-Analyse wurde eine Liste mit den normalisierten Signalintensitäten (Mittelwert der Replikate) für jedes Microarray-Experiment erstellt und für die statistische Auswertung verwendet. Dabei wurden einige Experimentfaktoren als zufällig („random factors“) definiert, d.h. würde das Experiment wiederholt, wären nicht exakt die gleichen aber ähnliche Effekte zu erwarten, beeinflusst durch Zufallsfaktoren („Array-Effekt“ (AG), „Farbstoff-Effekt“ (DG)) als Varianzquelle. Neben den Zufallsfaktoren wurde das Entwicklungsstadium („stage“) als unabhängiger Faktor definiert. Nach dem genspezifischen ANOVA-Model:

$$r_{ijgr} = G + VG_{ij} + DG_j + AG_i + \epsilon_{ijr}$$

Wobei gilt:

$r_{ijgr}$	Genspezifischer Effekt mit Array (i), Farbstoff (j), Gen (g) und Messung (r)
$G$	Mittelwert der Intensität des Gens (G)
$VG_{ij}$	Variation der Expression des Gens (VG) durch Array (i) und Farbstoff (j)
$DG_j$	Farbstoff-Effekt (DG) auf das Gen
$AG_i$	Array-Effekt (AG) auf das Gen
$\epsilon_{ijr}$	Differenzwert ( $\epsilon_{ijr}$ ) aus ANOVA-Model und Messung

wurden drei  $F$ -Test-Variationen (Vergleich der Variationen zwischen wiederholten Experimenten und Bedingungen) durchgeführt. Dabei wurde für jedes Gen ein  $p$ -Wert errechnet, der anzeigt ob eine signifikante Änderung der Expression vorliegt oder nicht.

Die Ergebnisse der  $F$ -Tests wurden visualisiert in einem Vulkan-Plot („Volcano plot“) zusammengefasst.

### **Clusteranalysen der Transkriptionsprofile exprimierter Gene**

Für die weitere Auswertung der Genexpressionsprofile wurden die aus M-CHiPS-gefilterten Signalwerte (Mediane der Intensitäten für jedes Stadium) logarithmiert. Dazu wurden die gemittelten Intensitätswerte jedes Stadiums jeweils durch das Kontrollstadium (Embryo 0-4 h) geteilt und der  $\log(2)$  aus den Verhältnissen („ratios“) gebildet.

Die Clusteranalyse erfolgte mit der J-Express Pro™ (Molmine AS) Software [44]. Um Muster in der Genexpression von Microarraydaten [45] zu finden werden die Genexpressionswerte

allgemein in Gruppen (Cluster) nach ähnlichem Profil geordnet. Die Einteilung der Genprofile in die Cluster wird von verschiedenen Clusteralgorithmen [46,47] übernommen. Für diese Arbeit wurde zunächst ein initialer „Self-organizing maps“ (SOM)-Algorithmus [25] angewendet. Dieser auf einem neuronalen Netzwerk basierende Algorithmus weist Gene [48,49] anhand ihrem ähnlichen Expressionsprofil einer Reihe von Partitionen zu, die durch einen Referenzvektor repräsentiert werden. Zuerst wird eine geometrische Konfiguration, meist ein zweidimensionales rechteckiges – oder in unserem Fall ein hexagonales Gitter aus 11 x 11 Knotenpunkten, vorgegeben. Zufallsvektoren werden für jede Partition generiert und „trainiert“ während eines iterativen Prozesses, bei dem eine effektive Trennung der Daten erfolgen soll. Dabei werden zufällig ausgewählte Gene aufgrund ihrer Distanz, wie verwendet die euklidische Distanz, zum Referenzvektor identifiziert. Der Referenzvektor wird daraufhin dem zugewiesenen Genvektor angepasst. Gleichzeitig werden Referenzvektoren aus der direkten Nachbarschaft des hexagonalen Gitters „herangezogen“ und ebenfalls dem Genvektor angepasst. Diese Schritte werden iterativ mehrfach wiederholt (wir benutzten 10.000 Wiederholungen) und mit zunehmender Stringenz werden die Gene den Partitionen zugeordnet, deren Referenzvektoren sie am Ähnlichsten sind. In einem zweiten Schritt wurde ein k-means-Algorithmus auf diese Partitionen angewendet. Bei diesem Algorithmus werden die Gene in eine vorgegebene Anzahl (k) Cluster eingeteilt, die sich intern ähnlich, aber nach außen unähnlich sind. Die iterative Methodik der Zuweisung der Gene ist dem SOM-Algorithmus sehr ähnlich, bis auf die Verwendung eines Referenzvektors. Durch den k-means-Algorithmus erfolgte die eindeutige und endgültige Klassifizierung der Genprofile in Cluster.

### ***Datenbankanwendung und Bioinformatik***

Den Genen innerhalb der Cluster wurde entsprechend ihren Gen-IDs (Flybase Accession Number) Funktionen mit Hilfe von Begriffen des „Gene Ontology™ Consortium“ (GO) aus der Flybase-Datenbank zugeordnet. Mit der Flybase Accession number konnten die annotierten Gene in der öffentlichen Flybase-Datenbank mit den Heidelberg FlyArray-IDs kombiniert werden. Die FlyArray-ID wird eindeutig jedem einzelnen Spot auf dem Microarray zugeordnet und kann über die Webseite „Heidelberg FlyArray“ variabel mit Genannotationen verbunden werden. Das bedeutet, jedem Spot auf dem DNA-Chip kann direkt über Web-Datenbanken ein Gen oder auch re-annotierte Genfunktionen zugeordnet werden. Für die weitere Analyse wurden die gefilterten Gene mit den Gene Ontology (GO) Annotationen [50] des Gene Ontology™ Konsortiums verknüpft. Das GO - Konsortium stellt eine offene Plattform von Begriffen für Genfunktionen für eine Vielzahl von Organismen dar [51]. Durch diese standardisierten Beschreibungen der Genfunktionen werden Organismen auf genetischer Ebene untereinander vergleichbar. Der Begriffskatalog ist in einer Baumstruktur

hierarchisch angeordnet und beschreibt mit sog. GO-IDs eindeutig die Ebene und Funktion eines Gens. Der Katalog ist in drei Unterkategorien eingeteilt: Biologischer Prozess, Molekulare Funktion und Zelluläre Komponente. Dabei werden die Begriffe entlang eines hierarchischen Pfades in der Form: Oberbegriff (z.B. Entwicklung) immer weiter detailliert (z.B. Entwicklung/Organogenese/Neurogenese...) fortgesetzt.

Die Analyse der Microarray-Daten wurde hauptsächlich auf einem GNU/Linux Debian Betriebssystem durchgeführt. Dieses diente als Grundlage für die R-Statistikumgebung und das Datenbankmanagementsystem PostgreSQL mit denen die Auswertung und Formattierungen der Daten erfolgte. Zur Manipulation von Text- und Tabellendateien wurde die Scriptsprache Python verwendet. Als Schnittstelle zur Darstellung der Expressionsdaten in der Datenbank wurde der Web-Applikationsserver Tomcat mit Java Server Pages™ (Sun Microsystems) verwendet.



## 3 Ergebnisse

### 3.1 Optimierung der Eigenschaften für Microarrays mit hoher Dichte

#### 3.1.1 Ziel der Optimierungen

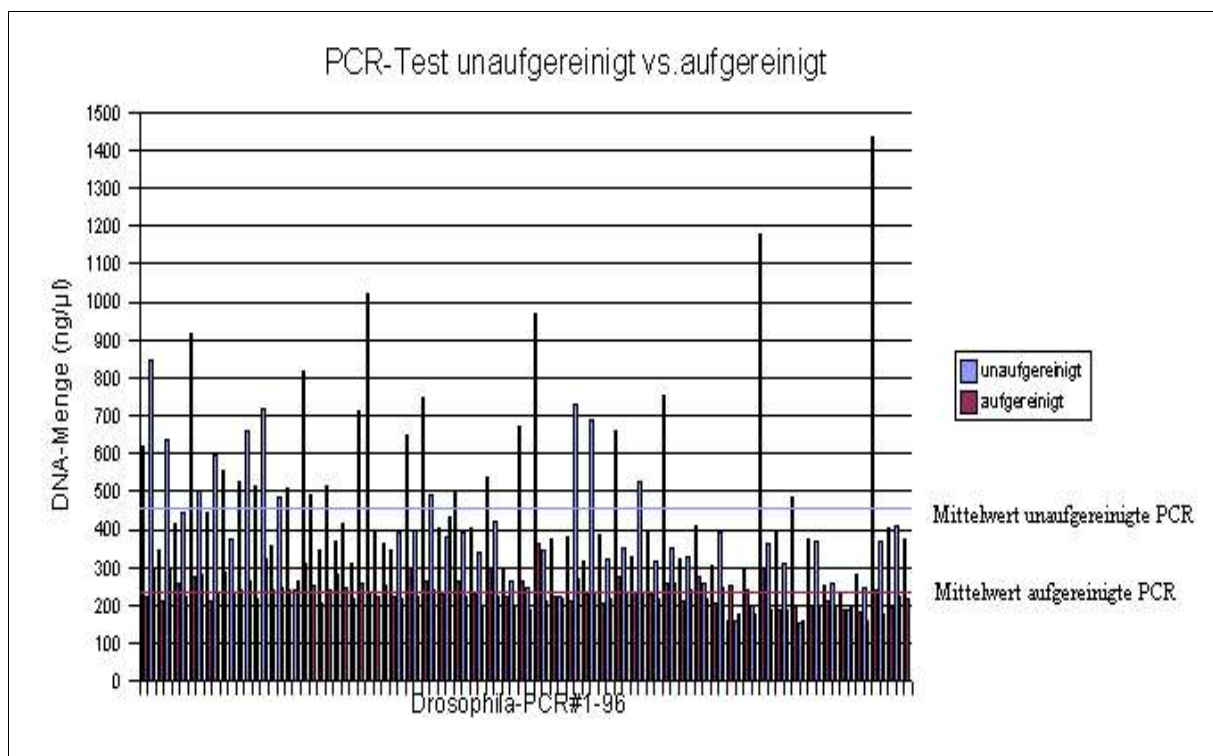
Um die hohe Anzahl der vorhergesagten PCR-Produkte aus dem Primerdesign-Programm GenomePride auf einen Glasobjektträger aufbringen zu können, mußten einige Modifizierungen bestehender Protokolle und Methoden zur Herstellung von DNA-Microarrays vorgenommen werden. Das Ziel war, einen Microarray als Standard für die Untersuchung des *Drosophila* Transkriptoms zu etablieren. Zu diesem Zweck wurden einige Optimierungsexperimente durchgeführt, da zu diesem Zeitpunkt noch keine Erfahrung mit der Herstellung von Microarrays mit hoher Dichte (ab 30.000 Features auf einem Objektträger) vorlag. Zusätzlich sollte aus den angefertigten PCR-Reaktionen reproduzierbar eine große Anzahl qualitativ hochwertiger DNA-Microarrays schnell und kostensparend hergestellt werden.

Als Grundlage wurde das von Frank Diehl [52] entwickelte PCR-Protokoll mit Betain-Zusatz für die Verwendung als Spotting-Lösung übernommen und für Microarrays mit hoher Auftragsdichte weiterentwickelt. Durch den Zusatz von 1.5 M Betain, das einen wasseranziehenden Effekt aufweist, konnte frühzeitiger Verdunstung während des Spotting-Vorgangs (eine 384er Platte entspricht ~ 1 h) entgegengewirkt werden. Der Nachteil dabei jedoch ist, daß sich der Spot-Durchmesser aufgrund der hygroskopischen Wirkung des Betains vergrößert. Um die gesamten PCR-Produkte zweifach (Replikate) auf einem Glasobjektträger unterzubringen (über 40.000 Spots) mußte der Spot-Durchmesser für einen Standard-Objektträger (22 x 75 mm) von bisher üblichen 200 µm auf 140 µm reduziert werden. Es wurde deshalb nach unterschiedlichen Möglichkeiten gesucht, die Protokolle zu optimieren.

#### 3.1.2 Verwendung unaufgereinigter PCR-Produkte

Um Puffer, Enzyme oder Nukleotide aus PCR-Reaktionen zu entfernen, werden diese z.B. mit Hilfe von PCR-Filterplatten aufgereinigt. Da die in dieser Arbeit generierten PCR-Produkte nicht für weitere sensitive enzymatische Reaktionen, sondern für den Auftrag auf die Objektträger dienten, war eine Aufreinigung nicht nötig, solange die aufgetragenen Proben spezifisch hybridisieren. Die Verwendung unaufgereinigter PCR-Produkte für die

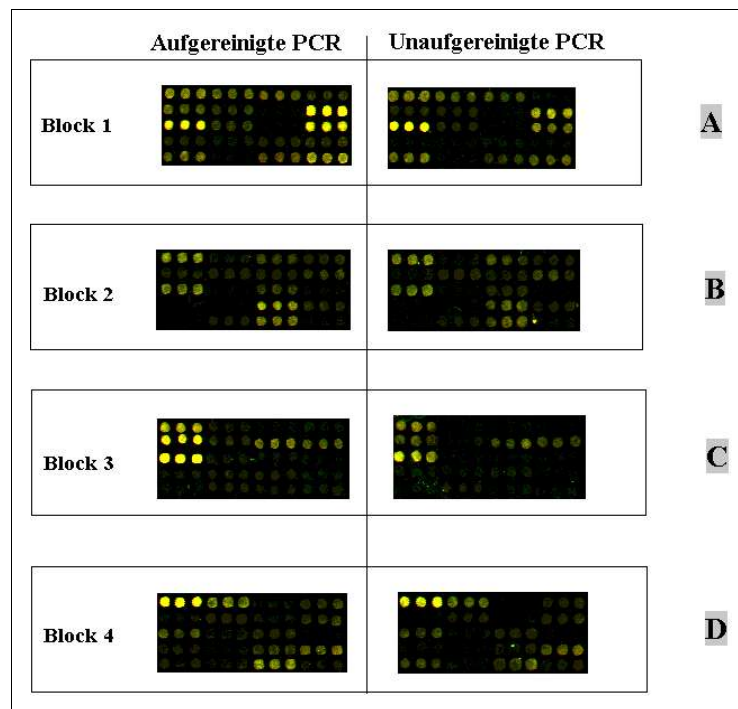
Herstellung von Microarrays hat den Vorteil, daß die Zeit- und kostenintensive Aufreinigung entfällt. Um das Hybridisierungsverhalten beurteilen zu können, wurden 96 *Drosophila* PCR-Produkte zufällig ausgewählt, wovon die Hälfte jedes PCR-Ansatzes (je 100 µl) aufgereinigt wurde und die andere Hälfte nicht. Die DNA-Konzentrationen aller PCR-Produkte wurden nach für beide Verfahren gemessen und in einer Tabelle aufgetragen (siehe Abb.3.1). Anschließend wurden die Proben als Triplikate (siehe Abb.3.2) auf Glasobjektträger aufgebracht. Um Nadelartefakte zu vermeiden, wurde das Array-Layout so gewählt, daß ein „Pin“ jeweils unaufgereinigte- und aufgereinigte PCR-Proben in einem Block auftrug.



**Abb.3.1:** Gemessene DNA-Konzentrationen der 96 unaufgereinigten- und aufgereinigten PCR-Produkte und der jeweilige Mittelwert.

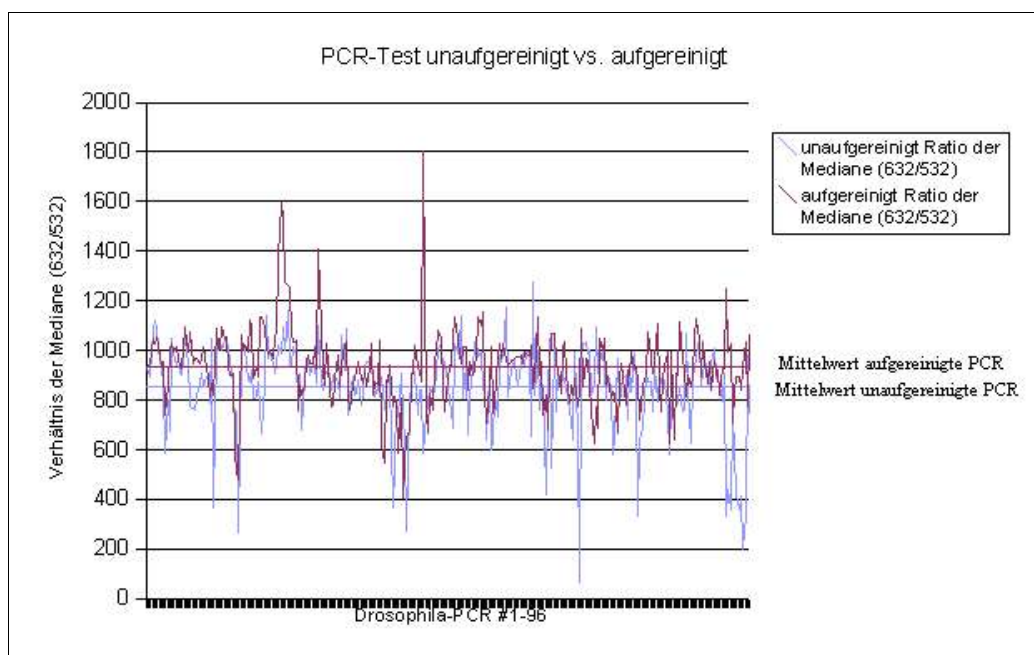
Es wurden insgesamt 4 Blöcke (3x 96 PCR-Proben mit Kontrollen) mit einem 6 x 12 Muster pro Block auf den Array aufgebracht. Als RNA wurde Gesamt-RNA von 0-4 Stunden alten Embryos verwendet. Diese wurde wie unter 2.2.9. beschrieben mit Cy5 und Cy3 markiert und aufgereinigt. Die DNA-Chips wurden mit markierter Gesamt-RNA (s. 2.2.12.) hybridisiert und anschließend im Scanner eingelesen.

Die ermittelten Signalintensitäten wurden anschließend ausgewertet (siehe Abb. 3.3) und verglichen.



**Abb.3.2:** Vergleich der Signalintensitäten von aufgereinigtem -und unaufgereinigtem PCR-Produkt auf dem Microarray. Die Darstellung zeigt die Selektion der vergleichbaren PCR-Produkte. Die Intensitäten spiegeln das Verhältnis des Cy3-und Cy5-Farbstoffe wider.

In Abb.3.2 zeigten sich bei den unaufgereinigten PCR-Produkten teilweise geringere Intensitäten (siehe Abb.3.2 A und C). Das Hybridisierungsmuster ist für jeden Block identisch, so daß eine unspezifische Hybridisierung und ein Verunreinigungseffekt durch nicht entfernte Oligonukleotide mit universeller Sequenz ausgeschlossen werden konnte.



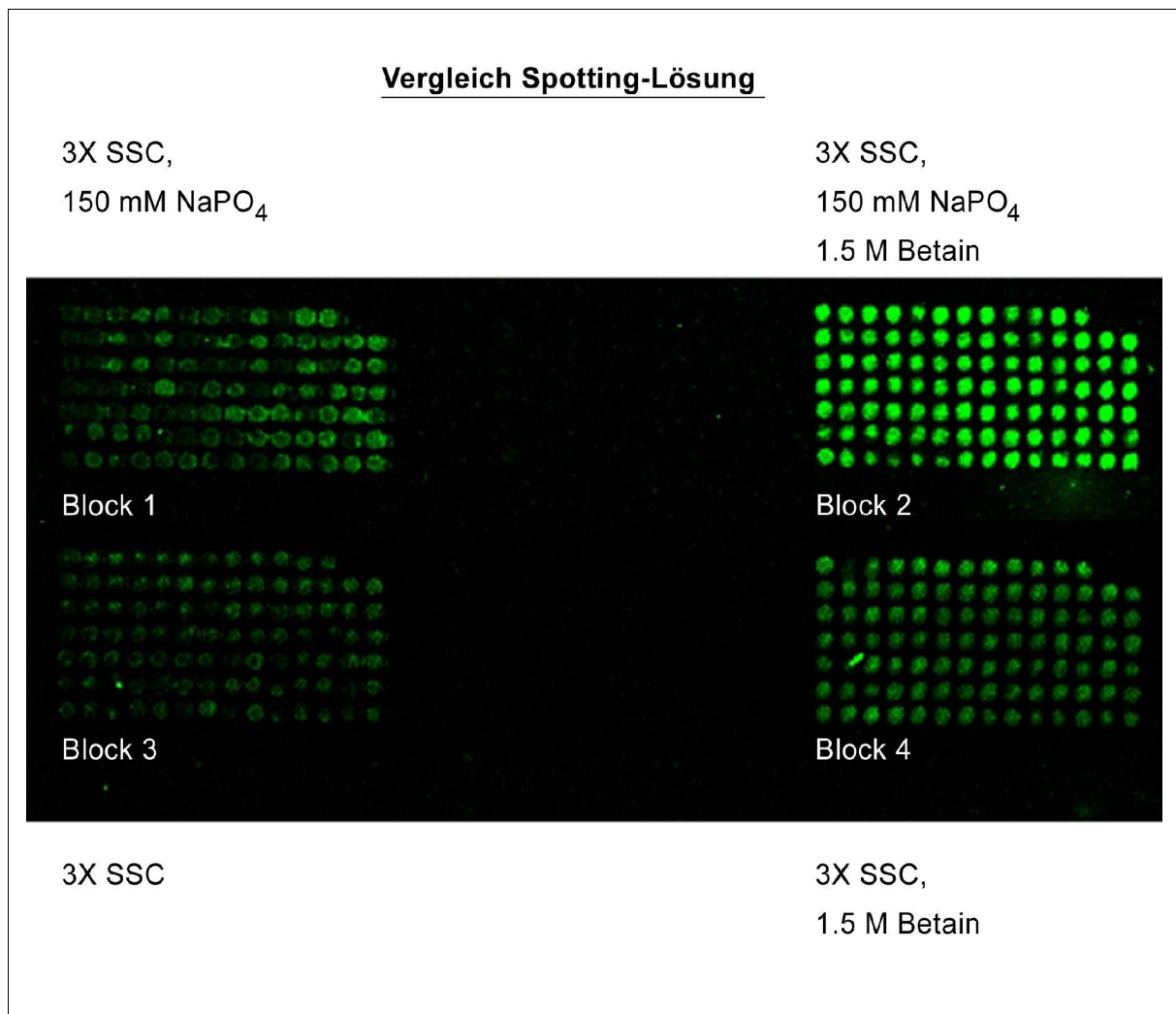
**Abb.3.3:** Signalintensitäten der 96 PCR-Produkte. Aufgetragen wurde das Verhältnis der Mediane der Kanäle (632/532) und der jeweilige Mittelwert für die 96 PCR-Produkte.

Bei der Messung der DNA-Konzentration (siehe Abb.3.1) ergaben sich die Mittelwerte 454 ng/ $\mu$ l für die unaufgereinigte PCR und 234 ng/ $\mu$ l für die aufgereinigte PCR. Das bedeutet einen Verlust von durchschnittlich ca. 50% der PCR-Produkte während des Aufreinigungsprozesses. Die Auswertung der Signalintensitäten (siehe Abb.3.3) ergab für die unaufgereinigten PCR-Produkte einen Mittelwert von 852 und für die aufgereinigten PCR-Produkte einen Mittelwert von 931. Das bedeutet eine ca. 10% durchschnittlich höhere Signalintensität bei der aufgereinigten PCR. Der Effekt der Aufreinigung der DNA mit Hilfe der PCR-Filterplatten ist also durchaus sichtbar, da sie bei wesentlich geringerer aufgetragener DNA-Konzentration eine leicht erhöhte mittlere Signalintensität aufweist. Ein Verlust von ca. 50% der PCR-Produkte ist aber gemessen an der Signaldifferenz von ca. 10% zwischen unaufgereinigter- und aufgereinigter PCR zu hoch und rechtfertigt damit nicht die wesentlich erhöhten Kosten und den höheren Zeitaufwand für die Aufreinigung. Ähnliche Resultate ergaben sich bei anderen Mitarbeitern in der Arbeitsgruppe. Die Ergebnisse sind in einer Publikation [53] zusammengefaßt.

### 3.1.3 Erhöhung der DNA-Bindungskapazität auf Glasobjektträgern

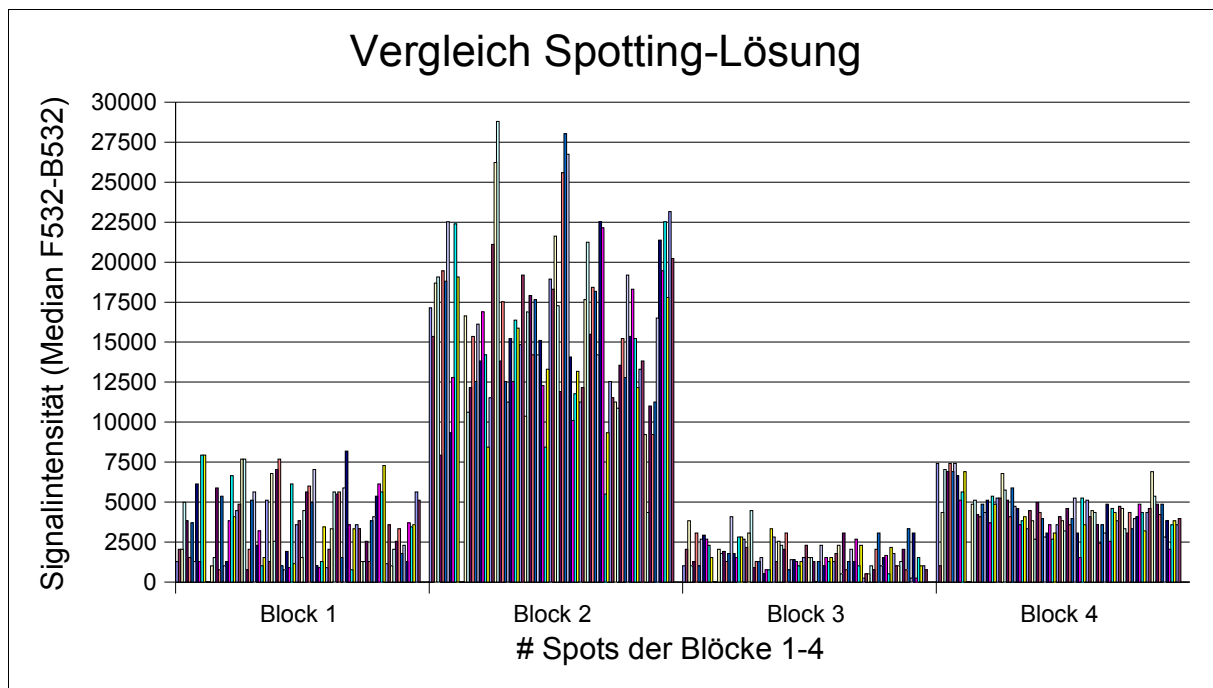
Die Qualität der Hybridisierungen und damit verbunden der Erfolg der Genexpressionsanalysen hängt maßgeblich von der Qualität der verwendeten Microarrays ab. Dazu müssen die PCR-Produkte in ausreichender Menge effektiv an die Glasoberfläche der Microarrays gebunden werden. Die Menge an amplifizierter DNA in den PCR-Platten sollte für die Herstellung mehrerer hundert Microarrays ausreichen.

Die Bindung der PCR-Produkte an aminosilanisierte Glasobjektträger erfolgt über ionische Wechselwirkungen, wobei das negativ-geladene Phosphatrückgrat der DNA an die positiv-geladenen Aminogruppen der Glasoberfläche bindet. Um die Bindungskapazität der DNA an die Glasoberfläche zu erhöhen und gleichzeitig einen Puffer gegen pH-Wert- Schwankungen für die Spotting-Lösung (3 X SSC, 1.5 M Betain) verfügbar zu machen, wurde der Einfluß von  $\text{NaPO}_4$ -Lösung auf die Bindungseffizienz der DNA untersucht. Zu diesem Zweck wurde der Spotting-Lösung 150 mM  $\text{NaPO}_4$  zugesetzt und zusammen mit anderen herkömmlichen Varianten der Spotting solution auf Glasobjektträger aufgetragen. Es wurde jeweils eine konstante Konzentration Herings-DNA (200 ng/ $\mu$ l) und Kresolrot (0,1 mM) zu vier Spotting-Lösungen gegeben. Durch den Kresolrot-Zusatz können die Microarrays nach dem Auftrag direkt mit dem Cy3-Kanalfilter im Laserscanner gescannt und ausgewertet werden.



**Abb.3.4:** Signalintensitäten der vier verwendeten Spotting-Lösungen. Gescannt wurde bei 633 nm Wellenlänge für den Cy3-Kanal. Das Bild wurde in eine Falschfarbe (grün) übersetzt.

Die in Abb. 3.4 dargestellten Signale, wurden mit der GenePix-Software ausgewertet und in eine Tabelle eingetragen. Die stärksten Intensitäten, wie bereits visuell zu erkennen ist, befinden sich im zweiten Block mit 3X SSC, 150 mM NaPO<sub>4</sub> und 1.5 M Betain als Spotting-Lösung. Zusätzlich erkennt man eine homogenere Morphologie der Spots, die auch in Block 4 (3X SSC, 1.5 M Betain) deutlich wird. Die Ursache dafür liegt in der Zugabe von Betain, welches eine gleichmäßigere Verteilung der DNA über die Auftragsfläche bewirkt [52]. Die Auswertung der Signalintensitäten (siehe Abb. 3.5) zeigte ebenfalls deutlich höhere Werte für Block 2. Wie aus der Tabelle 3.1 hervorgeht, ist der Mittelwert der Intensitäten von Block 2 gegenüber Block 4 mehr als dreimal so hoch. Diese Erhöhung der Signalintensität konnte nur durch die Zugabe des NaPO<sub>4</sub> zu der sonst verwendeten Spotting-Lösung bewirkt werden. Ein experimenteller Fehler, wie unterschiedliche Auftragsvolumina der Nadeln konnte ausgeschlossen werden, da die Positionen der Nadeln in wiederholten Experimenten getauscht wurden. Die anderen Spotting-Lösungen (Block 1 und Block 3) zeigten sehr geringe Intensitätswerte mit heterologer Spotmorphologie, die daher für eine spätere Datenanalyse schlecht verwendbar sind.



**Abb.3.5:** Signalintensitäten für die Spots der 4 Blöcke. Verwendet wurden die Medianwerte der Pixel für den Cy3-Kanal bei 633 nm. Von den jeweiligen Medianwerten wurden die Hintergrundwerte (B532) abgezogen.

Für die Herstellung der Microarrays wurde aus diesem Grund ausschließlich das modifizierte Protokoll mit Phosphat-Zusatz verwendet.

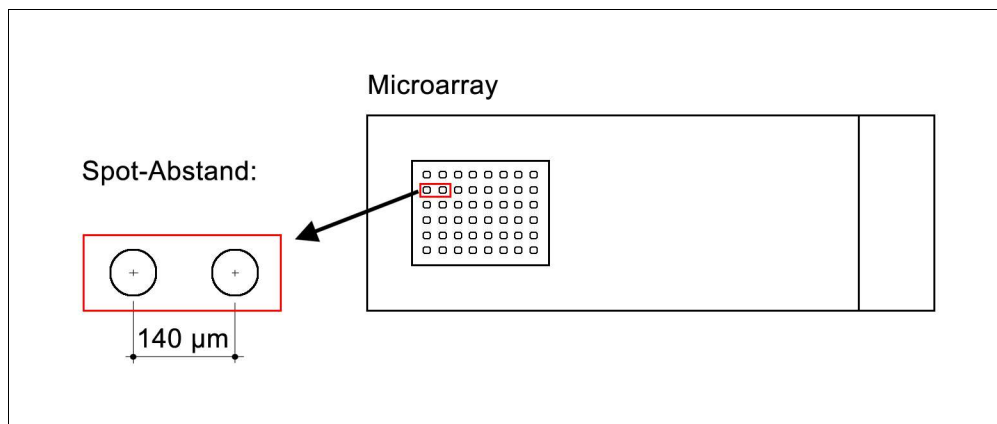
<b>Block #</b>	<b>Mittelwert (Block)</b>
1	3475,59
2	15345,63
3	1665,31
4	4291,92

**Tab. 3.1** Die 4 Blöcke mit unterschiedlichen Spotting-Lösungen und der Mittelwert der Intensitäten für jeden Block.

### 3.1.4 Auswahl der geeigneten Glasobjektträger

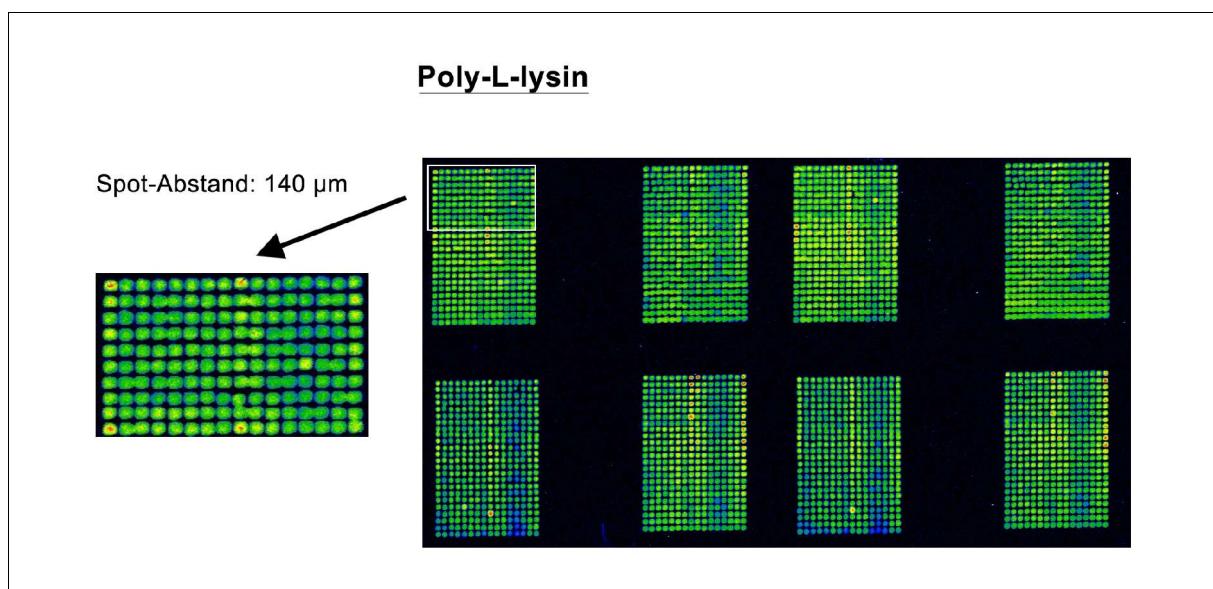
Um die PCR-Produkte mit der optimierten Spotting-Lösung (3X SSC, 150 mM NaPO<sub>4</sub>, 1.5M Betain) in ausreichendem Abstand zweier Spot-Mittelpunkte („Spot distance“) auf die Glasobjektträger aufzubringen, wurden verschiedene Microarrays mit aminosilanisierter Oberfläche getestet. Dabei durfte der Durchmesser der Spots einerseits nicht zu gering sein, da die Größe der Spots die Menge an gebundener DNA limitiert. Andererseits müssen die

Spotaußengrenzen genügend Abstand zueinander haben, um den Kontakt und damit ein Ineinanderfließen zu verhindern. Der errechnete Durchmesser der jeweiligen Spot-Abstände um die PCR-Proben als Replikate mit Kontrollen auf einem Areal von 22 x 75 mm (Standardobjektträger) mit 24 Nadeln aufzubringen, lag bei maximal 140 µm (siehe Abb. 3.6).

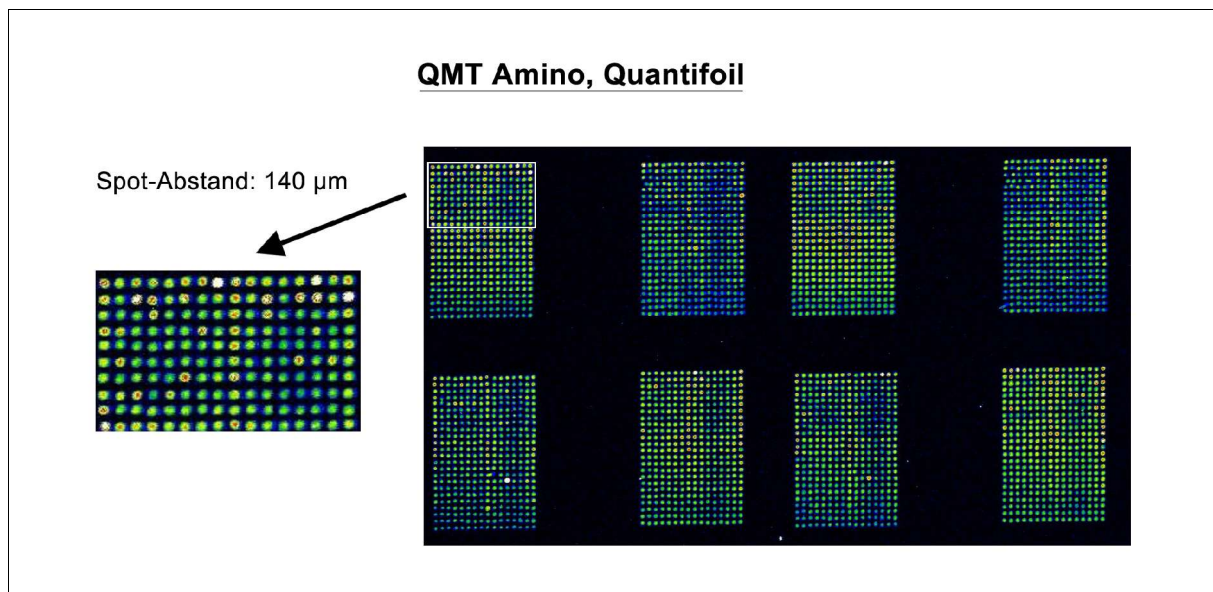


**Abb.3.6:** Schema des Durchmessers zwischen dem Mittelpunkt zweier Spots.

Die Software des Array-Roboters MicroGrid II™ gibt abhängig vom gewählten Programm mit bestehender Nadelkonfiguration den Mindestabstand (140 µm) vor. Für die Versuche wurde die 1x Test Spotting-Lösung verwendet, um zu gewährleisten, dass alle Glasobjektträger mit der gleichen Lösung behandelt wurden. Es wurden Poly-L-Lysin-Slides und QMT Amino-Slides miteinander verglichen. Die Poly-L-Lysin-Slides werden in unserer Arbeitsgruppe selbst hergestellt, die QMT Amino-Slides stammen von der Firma Quantifoil.



**Abb.3.7:** Ausschnitt aus einem Poly-L-Lysin-Slide mit Vergrößerung. Der Auftrag der Testlösung erfolgte mit einem Abstand von 140 µm. Gemessen wurden die Signalintensitäten des Cy3-Kanals.



**Abb.3.8:** Ausschnitt aus einem QMT Amino-Slide mit Vergrößerung. Der Auftrag der Testlösung erfolgte mit einem Abstand von 140 µm. Gemessen wurden die Signalintensitäten des Cy3-Kanals.

Bei den Poly-L-Lysin-Slides (Abb.3.7) erkennt man in der Vergrößerung den erhöhten Durchmesser der Spots (ca. 120 µm) und resultierend daraus teilweise einen Kontakt zwischen den Spots. Da jeder Spot ein individuelles PCR-Produkt für die Hybridisierung darstellt, muß eine Kontamination der DNA-Proben vermieden werden. Bei Verwendung der QMT Amino-Slides (Quantifoil) (siehe Abb.3.8) zeigen sich wesentlich geringere Spotdurchmesser (ca. 80 µm) und ein ausreichender Abstand zwischen den Mittelpunkten der Spots. Für die weiteren Hybridisierungsexperimente wurden daher ausschließlich die QMT Amino-Slides verwendet.

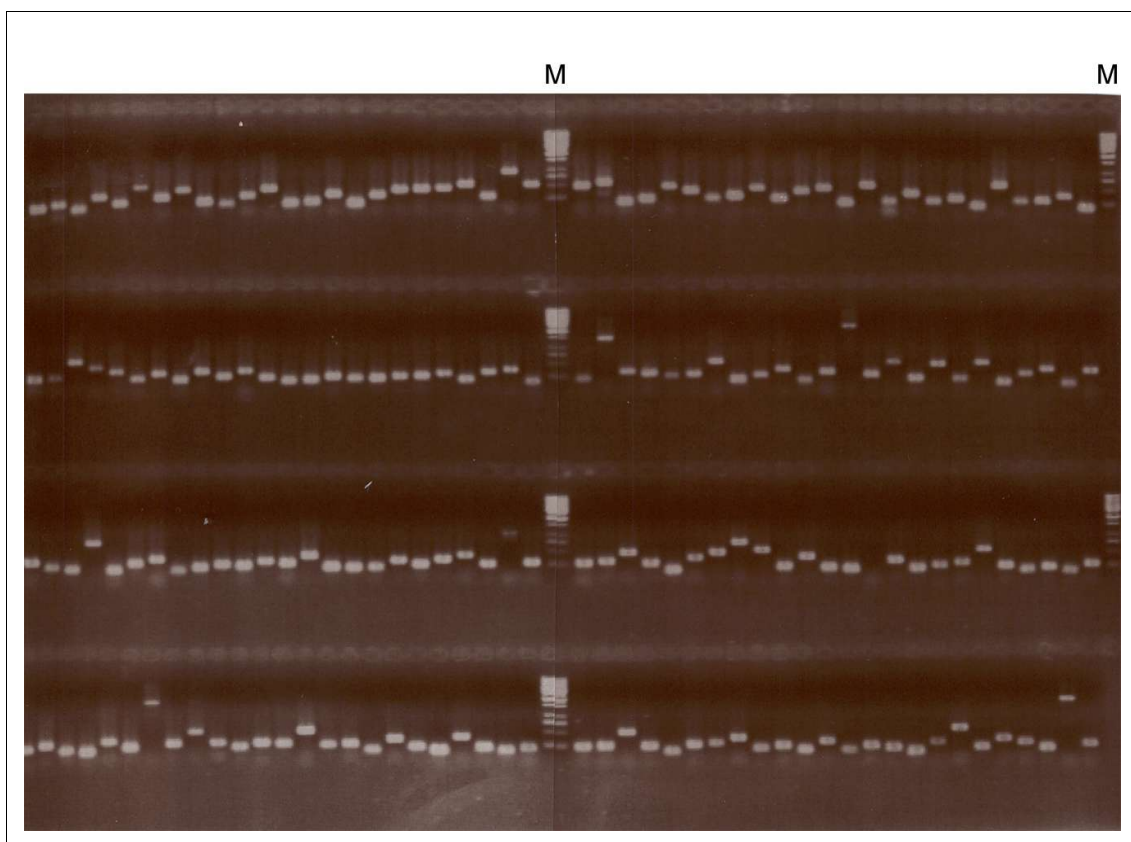
## 3.2 Aufbau des Microarray „Heidelberg FlyArray“

### 3.2.1 PCR-Amplifikation der Heidelberg Kollektion R1

mit Hilfe der spezifischen Primer wurden, wie unter 2.2.1. beschrieben, die 21.306 PCR-Produkte mit genomischer DNA als Vorlage amplifiziert. Die Bewertung der amplifizierten Fragmente erfolgte anhand der Gelelektrophorese. Die Größe für jedes Amplifikat war durch die Vorhersage mit GenomePride bekannt (siehe 1.4), und so erfolgte ein Größenvergleich auf dem Gel (siehe Abb.3.9). Entsprach ein PCR-Produkt der errechneten Größe wurde es mit '+' bezeichnet, fehlende Banden oder PCR-Produkte mit falscher



Größe erhielten ein '-' und Mehrfachbanden ein 'db'. Konnten in einer 96er PCR-Platte nur weniger als 90 % der Fragmente amplifiziert werden, so wurde die gesamte 96er PCR-Platte wiederholt. Die amplifizierten Produkte dienten als Vorlage für die zweite PCR mit universellen Primern. Die entstandenen PCR-Produkte wurden ebenfalls, wie für die erste PCR, auf ihre Größe kontrolliert und gleichermaßen bewertet. Von den jeweils 100 µl PCR-Reaktionen der zweiten PCR wurde jeweils die Hälfte (50 µl) entnommen und in die 384er Spotting-Platten zum Eintrocknen übertragen. Von den insgesamt voraussberechneten 21.396 PCR-Produkten konnten bei der zweiten PCR und den entsprechenden Wiederholungen 20.948 (97,9%) erfolgreich amplifiziert werden. Diese standen damit als Grundlage für die Herstellung des Microarrays zur Verfügung.



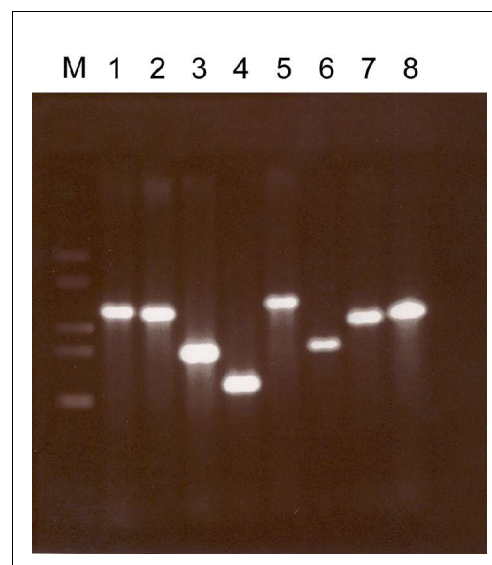
**Abb.3.9:** Ergebnis einer zufällig ausgewählten 1. PCR-Amplifikation mit universellen Primern. Je 2 µl aus 192 PCR-Proben ( 2x 96er PCR-Platten) wurden auf ein 2%iges Agarosegel aufgetragen und elektrophoretisch bei 100 V für 25 min aufgetrennt. In den Spuren M wurde der Größenmarker GeneRuler aufgetragen.

### 3.2.2 Herstellung der heterologen Kontrollen

Wie unter 2.2.2 beschrieben wurden 8 heterologe cDNA-Kontrollen von *Arabidopsis thaliana* hergestellt, die einerseits als PCR-Produkte auf den Microarray aufgetragen und

andererseits mittels *in-vitro* Transkription in mRNA umgeschrieben wurden. Durch die synthetisierte mRNA konnten die Kontrollen in definierter Konzentration zu der komplexen *Drosophila* Gesamt-RNA hinzugefügt und hybridisiert werden. Anhand der Signalintensitäten der Kontrollen konnten die DNA-Chips während des Scanvorganges kalibriert werden (siehe 2.2.13.). Damit wurde der qualitative Vergleich der Datensätze ermöglicht. Außerdem bestand die Möglichkeit, anhand der standardisierten Kontrollen eine Normalisierung durchzuführen, da die zugegebene Konzentration der Kontrollen in den Hybridisierungsexperimenten konstant gehalten wurde.

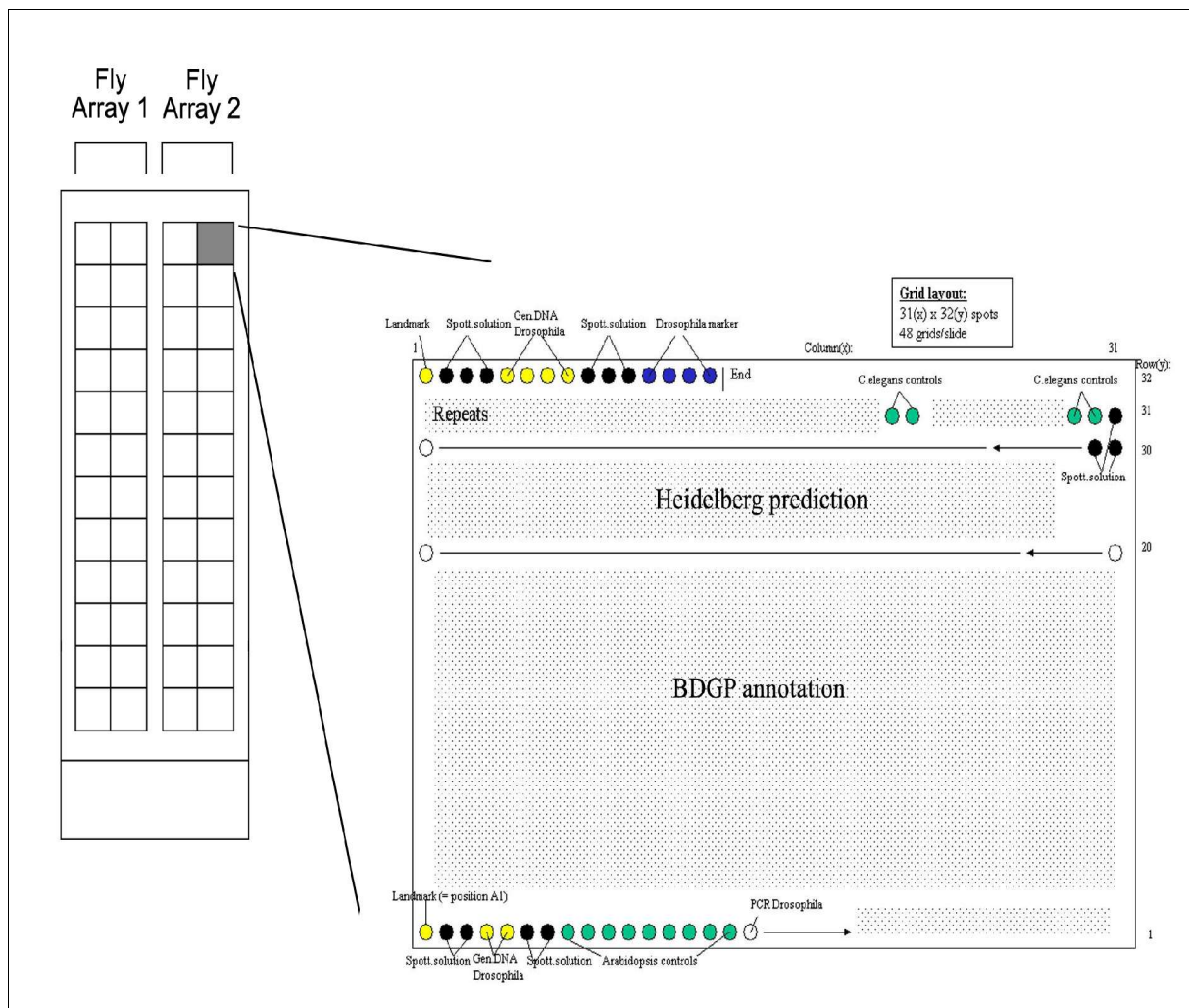
Es konnten 1 bis 2  $\mu\text{g}$  an mRNA der 8 cDNAs durch die *in-vitro* Transkription hergestellt werden. Die Messung der RNA-Konzentration erfolgte nach 2.2.3. und eine Kontrolle der mRNA erfolgte anhand von 2  $\mu\text{l}$  Aliquots auf einem denaturierenden Agarosegel, wie unter 2.2.4. beschrieben (siehe Abb. 3.10). Diese mRNA wurde im Verhältnis 1:1000 verdünnt und mit einer Konzentration von je 8 ng/ $\mu\text{l}$  für jede Kontroll-cDNA zu den einzelnen Hybridisierungen gegeben („spiking“). Kreuzhybridisierungen zwischen der Kontroll-RNA und aufgetragenen *Drosophila*-Proben auf dem Microarray konnten nicht beobachtet werden (Daten nicht gezeigt).



**Abb.3.10:** mRNA der 8 *A.thaliana* Kontrollen gelelektrophoretisch aufgetrennt. Es wurden je 2  $\mu\text{l}$  der Kontrollen auf ein denaturierendes 1,2%iges Agarosegel aufgetragen und für 45 min bei 100 V getrennt. In Spur M ist der Größenmarker RNA-Ladder und in den Spuren 1-8 die mRNA-Proben „Arab“ 1-8 aufgetragen.

### 3.2.3 Herstellung der cDNA-Microarrays

Die Herstellung der cDNA-Microarrays („Heidelberg FlyArray“) erfolgte wie unter 2.2.8. beschrieben. Es wurden 61 (59 *Drosophila* PCR-Platten + 2 Kontrollplatten) 384er Spotting-Platten mit je 10 µl 1x Spotting-Lösung eingesetzt. Zusammen mit den 20.948 PCR-Produkten einschließlich redundanter Wiederholungen aus den 96er PCR-Platten wurden ebenso die *Arabidopsis* Kontroll-cDNAs, *C.elegans* Kontroll-cDNAs (aus dem ZMBH), die Positionsmarker, Spotting-Lösung als Negativkontrolle und genomische *Drosophila*-DNA als Positivkontrolle auf die QMT Amino™-slides als Replikate aufgebracht. Insgesamt bestand der Microarray aus 47.616 Spots. Der Spot-Mittelpunktabstand betrug 140 µm. Der Auftrag erfolgte mit 24 (2 x 12) SMP3-Pins (TeleChem) auf dem MicroGrid II™ (BioRobotics) Array-Roboter. Die Replikate wurden an verschiedenen Stellen auf dem Objektträger aufgetragen, um lokale Signalintensitätsdifferenzen zu vermeiden. Auf einem Glasobjektträger wurden auf einer Fläche von 22 x 60 mm daher 2 identische Arrays (48 Blocks, bzw. „Grids“) untergebracht (siehe Abb. 3.11). Um eine einheitliche Spotmorphologie zu gewährleisten, wurden sog. „Prespot-Slides“ verwendet, auf denen die ersten 10 Nadelaufträge erfolgten. Es konnten während eines gesamten Roboterlaufes (72 h) 108 DNA-Chips hergestellt werden, von denen in der Regel (abzüglich „Prespot“-Slides) 75 DNA-Chips (69.5%) für weitere Hybridisierungsexperimente verwendet werden konnten.



**Abb.3.11:** Schematischer Aufbau des Heidelberg FlyArray. Auf dem Objektträger wurden 2 gleiche Arrays (je 2 x 12 Blöcke) aufgetragen. Die Vergrößerung stellt schematisch einen der 48 Blöcke dar. Jeder Block besteht aus 31 x 32 Spots. Die unterste und oberste Reihe besteht im wesentlichen aus heterologen Kontrollen. Den oberen Bereich bilden die wiederholten PCR-Produkte. Die PCR-Produkte der Heidelberg-Genvorhersage folgen darunter. Den größten Bereich nehmen die PCR-Produkte aus der BDGP Annotation im unteren Bereich ein.

### 3.3 Durchführung der Experimente

#### 3.3.1 Fluoreszenzmarkierung der RNA

Die jeweils 20 µg Gesamt-RNA entnommen aus den jeweiligen stadienspezifischen RNA-Pools wurden durch die reverse Transkription in cDNA überführt. Dabei wurde Aminoallyl-

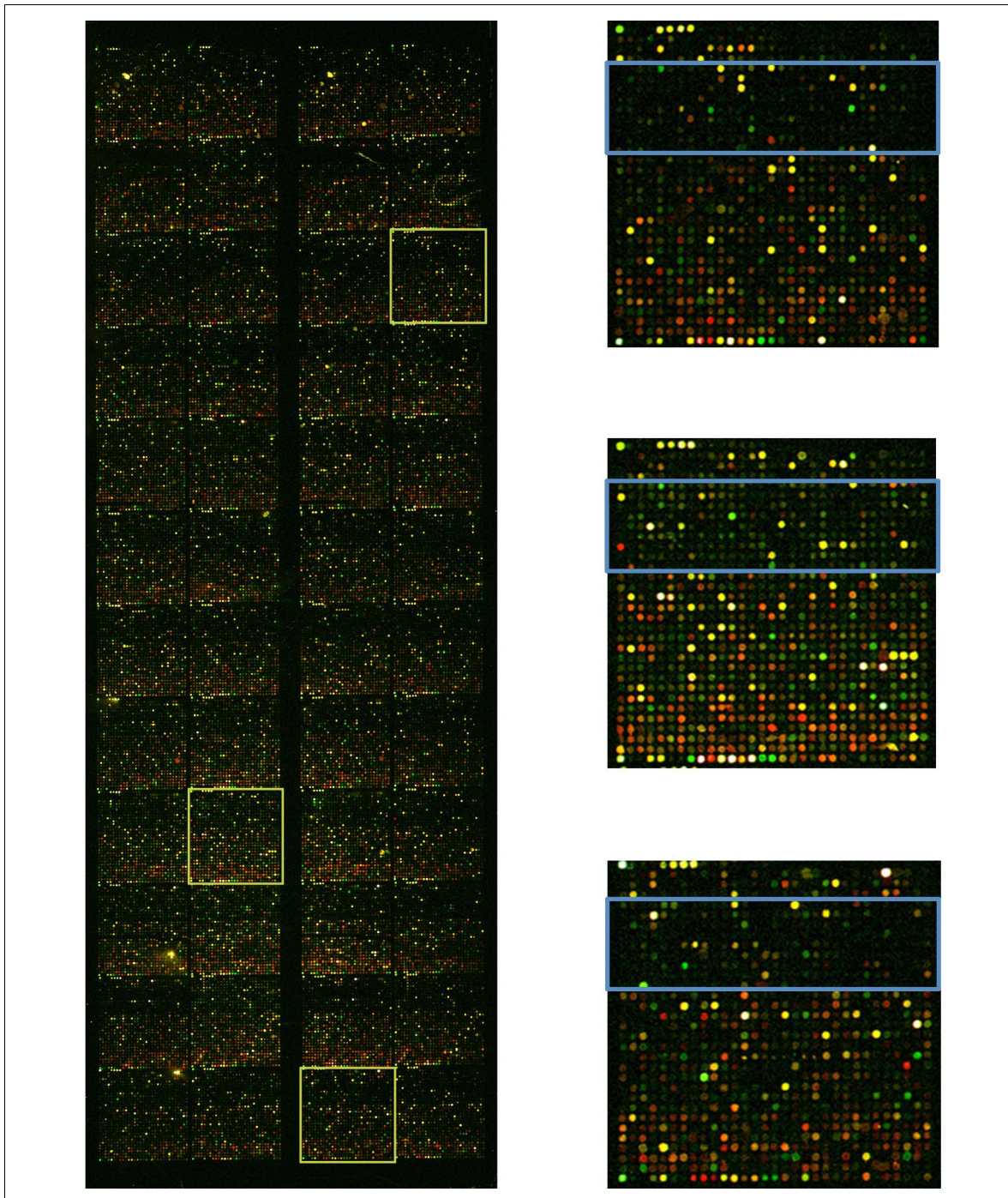
dUTP in die synthetisierte einzelsträngige cDNA eingebaut und anschließend in einem zweiten Schritt mit einem Cy3- bzw. Cy5-Farbstoff markiert (siehe 2.2.9.). Um den Erfolg der indirekten reversen Transkription zu messen, wurde die Einbaurate der gekopplten Farbstoffe photometrisch bestimmt. Wie unter 2.2.11. beschrieben, wurde für die markierten cDNAs eine Absorptionmessung durchgeführt, wobei jeweils Einbauraten zwischen 1 – 3% gemessen wurden, d.h. durchschnittlich alle 30 bis 150 Nukleotide wurde ein Farbstoffmolekül inkorporiert. Die Menge an gewonnener markierter cDNA lag im Durchschnitt bei 1,0 – 2,5 µg.

### 3.3.2 Hybridisierungen auf den Microarrays

Die Hybridisierungen der markierten cDNAs auf die Microarrays erfolgte als kompetitive Co-Hybridisierung (siehe 2.2.12). Dabei wurde jeweils die Kontrollbedingung (0-4 h Embryo) mit einem anderen *Drosophila*-Entwicklungsstadium kombiniert. Das erste Embryonalstadium wurde als Referenzkontrolle in allen Hybridisierungsexperimenten verwendet, da über die Genexpression zu diesem frühen Zeitpunkt die meisten Literaturinformationen vorlagen. Jede Hybridisierung wurde dreimal wiederholt. Mit den auf dem DNA-Chip vorhandenen Replikaten standen somit insgesamt 6 Datenpunkte für jedes Gen in einer Kondition zur Verfügung. Dadurch sollte eine ausreichende statistische Verteilung der Ergebnisse erreicht werden. Innerhalb der drei Wiederholungsexperimente wurden mindestens jeweils einmal die Farbstoffe (Cy3 bzw. Cy5) zwischen den Proben gewechselt („dye swap“), um zu gewährleisten, daß experimentelle Abweichungen durch unterschiedliches Verhalten der Farbstoffe bei der Signaldetektion ausgeglichen werden konnten [54]. Es wurden insgesamt 8 verschiedene Entwicklungsstadien (siehe Tab. 3.2) gegen die Kontrolle von *Drosophila melanogaster* hybridisiert.

Stadium	Embryo				Larve	Puppe (Stadien)			Adult
Zeitpunkt	0-4 h	4-8 h	8-12 h	12-16 h	1-3	1	2	3	1
	Kontrolle								

Tab. 3.2 Die verwendeten Entwicklungsstadien und Zeitpunkte der RNA-Proben von *Drosophila*. Die Zeitpunkte für die Embryonalstadien sind in Stunden, für die Larve die 3 Larvenstadien, für die Puppe die 3 Verpuppungsphasen und die junge ausgewachsene Fliege angegeben. Das erste Embryonalstadium (0-4h) wurde als Kontrolle verwendet.



**Abb.3.12:** Der komplette Heidelberg FlyArray. Co-Hybridisiert wurde hier markierte cDNA von 0-4h alten Embryos gegen die adulte Fliege. Das gescannte Bild zeigt Signale des gesamten Objektträgers. Die vergrößerten Ausschnitte entsprechen jeweils einem Block (Grid). Die blau-umrandeten Bereiche kennzeichnen die Heidelberg-spezifischen Gene. Bereits visuell erkennt man das Vorhandensein von Signalen in diesen Arealen.

### 3.3.3 Signaldetektion und Quantifizierung der Hybridisierungsdaten

Die gescannten Bilder wurden mit Hilfe von GenePix™ quantifiziert. Durch die Berechnung der Standardabweichung (SD) zwischen den Replikaten eines Arrays (siehe 2.2.13.) wurde die Qualität der Hybridisierungen beurteilt. Von insgesamt 73 angefertigten Hybridisierungen wurden 24 (32,8%) für die weiteren Genexpressionsanalysen verwendet (siehe 4.1.4). Die Ergebnisse der Signaldetektion und die Quantifizierung der Intensitäten mit GenePix™ wurden in Tabellenform abgespeichert. Dabei wurden jedem Spot, welcher ein Gen repräsentiert, neben den Signalintensitätswerten für den Cy3- und Cy5-Kanal noch Werte für den lokalen Hintergrund, Pixelzahl, Sättigungsgrad, Standardabweichung und weitere Werte zugeordnet. Diese Datentabellen dienten als Grundlage für die nachfolgenden Genexpressionsanalysen.

Die Microarraydaten und Signalintensitäten für alle 24 Hybridisierungen sind im Rahmen der Publikation [37] bei der „Gene Expression Omnibus“ (GEO) Datenbibliothek über eine Webdatenbank (siehe 2.1.11) unter den Zugriffsnummern („Accession no.“) GPL517, GSM10917-GSM10940 verfügbar.

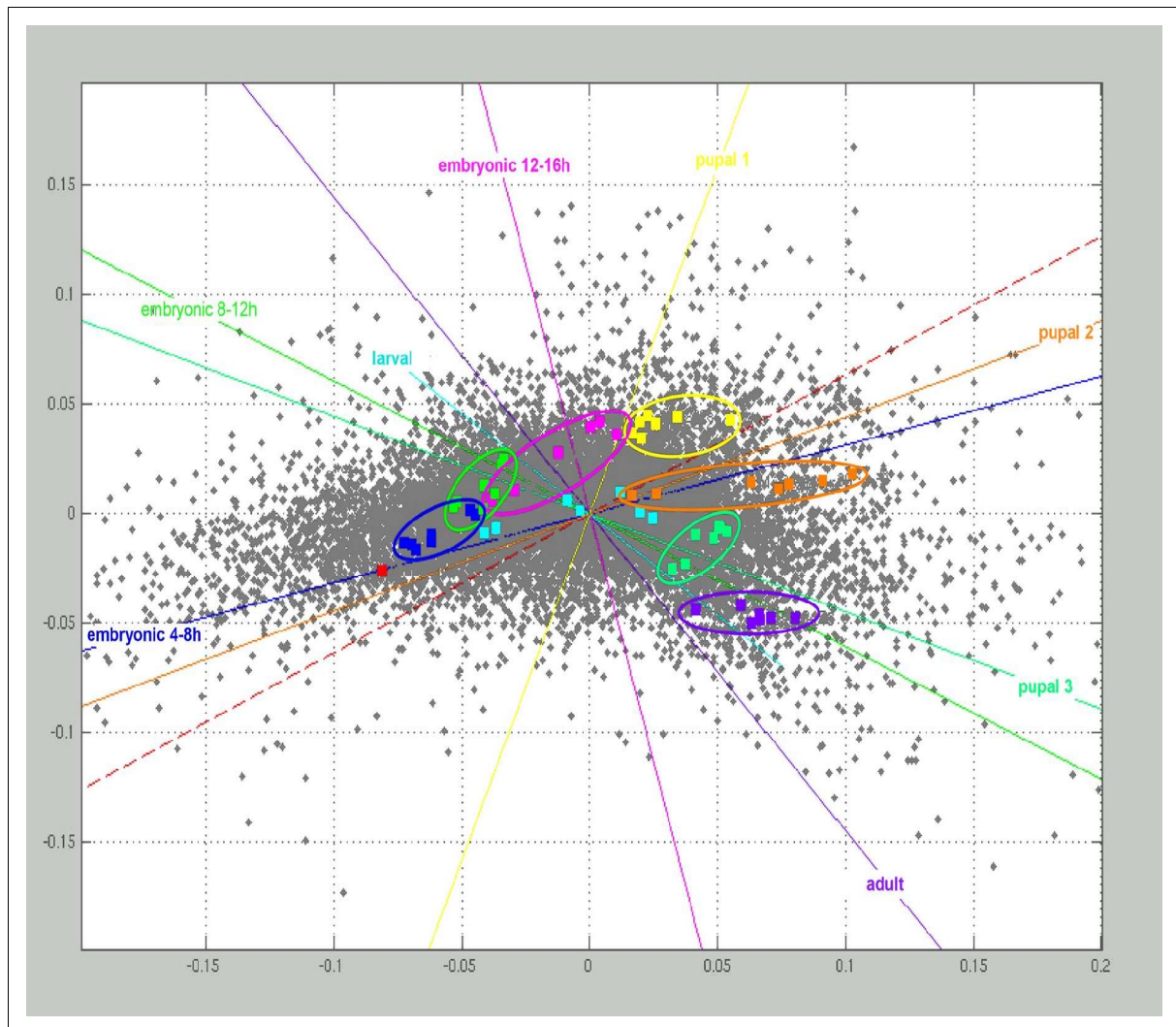
## 3.4 Analyse der differentiellen Genexpression in der Entwicklung von *Drosophila melanogaster*

### 3.4.1 Filterung und Auswertung der Daten mit M-CHiPS

Mit dem Softwarepaket M-CHiPS [21] wurden die Signalintensitäten in Form von Tabellen (.GPR-Format) aus der GenePix™-Software weiter ausgewertet. Die Hybridisierungen und ihre Wiederholungen unter der gleichen Bedingung konnten dabei zu Experimenten zusammengefaßt werden. Die Experimente konnten beliebig miteinander kombiniert werden, wobei immer eine Bedingung (= Cy3- oder Cy5-markierte cDNA) als gemeinsame Referenzkontrolle definiert werden mußte. Diese Kontrolle mußte in allen Zweifarbenexperimenten gleich sein, um die kompetitiven Hybridisierungen untereinander vergleichbar zu machen. Die 24 Hybridisierungen wurden nach den Entwicklungsstadien (siehe Tab. 3.2) der Fliege in 8 Bedingungen mit je 3 Hybridisierungen aufgeteilt. Die gemeinsamen Kontrollbedingungen wurden zusammengefaßt, gleichgültig ob sie mit Cy3- oder Cy5-Farbstoff markiert wurden. Zuerst erfolgte die Normalisierung der Daten (siehe 2.2.13.). Zu diesem Zweck wurden die selektierten Hybridisierungen entsprechend ihren Bedingungen in einem Plot gegeneinander aufgetragen und eine lineare Regressionsgerade für die Daten errechnet. Dabei wurden die Werte für die Mehrheit aller Gene der beteiligten

Hybridisierungen in die Berechnung einbezogen und durch eine multiplikative und additive Korrektur wurde die Regressionsgerade mit der Steigung = 1 so angepaßt, das sie durch den Koordinatenursprung verläuft. Für die Normalisierung wurde der Median aller beteiligten Kontrollbedingungen verwendet. Aufgrund dessen können mehrere Bedingungen gleichzeitig in ein Experiment einbezogen und analysiert werden. Nach der Normalisierung erfolgte die Filterung der Daten auf Gene, die eine Änderung in der Expression aufwiesen. Dabei kamen die erwähnten drei Filterkriterien zum Einsatz (siehe 2.2.13). Durch jeweils drei wiederholte Hybridisierungen einer Bedingung und den Replikaten auf jedem Microarray wurden insgesamt 6 Datenpunkte erhalten, die zur statistischen Stabilisierung der Signalintensitätswerte beitragen. Dies war wichtig für die Bestimmung eines reproduzierbaren Signalintensitätsschwellenwert, der Signale von dem Hintergrundwert trennt. Weiterhin wurde nach Genen gefiltert, bei denen das gemittelte Verhältnis der Signalintensitäten der Farbstoffe auf eine Änderung der Genexpression hinwies. Außerdem wurde auf die Reproduzierbarkeit der Genexpressionsänderung [55] gefiltert. Nach Anwendung der Filter verblieben noch 13.812 (64,5%) von 21.396 Genen. Davon entsprachen 3.473 (46,5%) (siehe Abb. 3.12) den unbekanntem Genen der Heidelberg Kollektion R1 und 10.339 (74,2%) der BDGP Genomannotation. Damit wurde der Beweis erbracht, daß bisher unbekannte Gene zusätzlich zu bekannten Genen aus der BDGP Annotation transkribiert werden. Dieser Gensatz repräsentiert die Anzahl der Gene, deren Expression im Transkriptom von *Drosophila* für alle untersuchten Entwicklungsstadien nachgewiesen werden konnte. Mit diesen gefilterten Genen wurde anschließend eine Korrespondenzanalyse [56] durchgeführt. mit Hilfe der Korrespondenzanalyse (siehe 2.2.13.) wurde die Qualität der Experimente überprüft, da sich durch diese Analyse die Beziehungen zwischen Genen und Hybridisierungen in einem Plotdiagramm darstellen lassen (siehe Abb. 3.13). In der Abb. 3.13 erkennt man anhand der umkreisten Cluster die Auftrennung der Entwicklungsstadien im Uhrzeigersinn. Jede Farbe steht dabei für ein Stadium. Die Korrespondenzanalyse projiziert die Ähnlichkeit der Bedingungen relativ zueinander in einen mehrdimensionalen Raum. Dadurch wird deutlich, daß sich die Embryonalstadien relativ ähnlich sind, wohingegen die Puppenstadien und das Adultstadium größere Abstände zueinander und zu den Embryonalstadien aufweisen, jeweils repräsentiert durch distinkte Cluster. Das Larvenstadium (türkis) steht dazu im Gegensatz, da sich hier die Hybridisierungen zwischen allen Stadien bewegen. Die Reproduzierbarkeit der Ergebnisse äußert sich durch die geringe Distanz der Quadrate einer Farbe zueinander, die jeweils Wiederholungen der Hybridisierungen auf einem Microarray darstellen. Bis auf kleinere Abweichungen (oranger Kreis) war die Streuung der Experimente gering, was für die Güte der Hybridisierungen sprach.



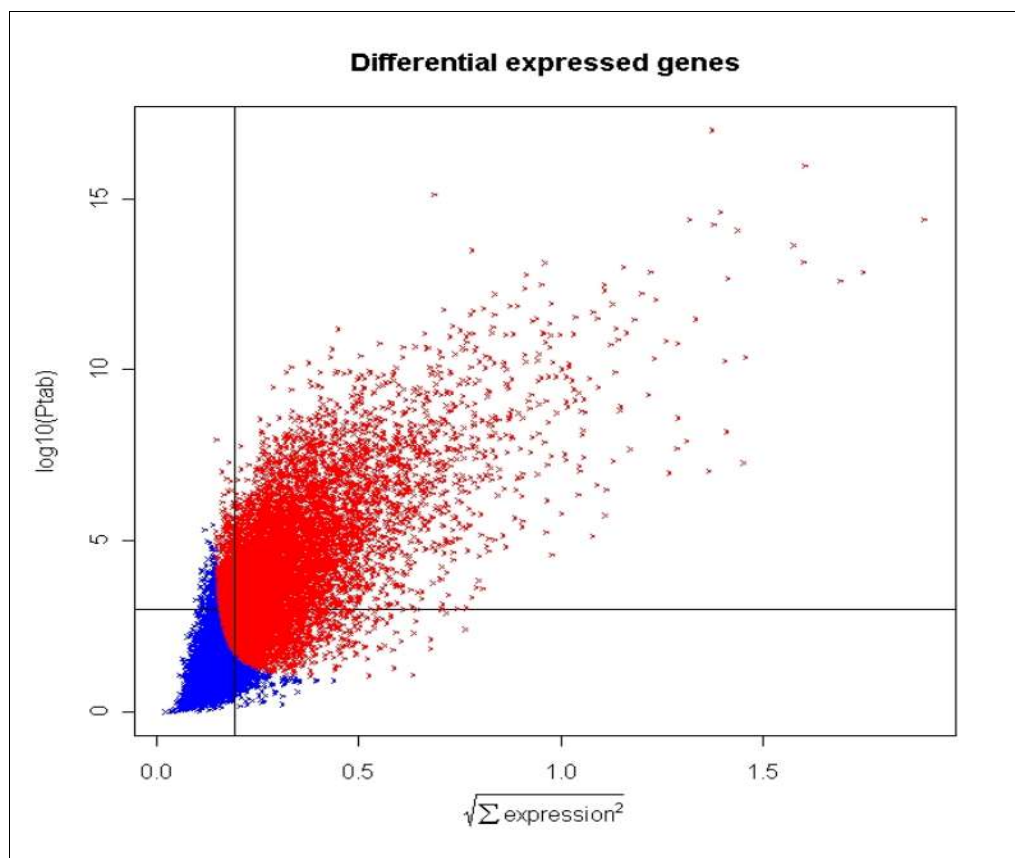


**Abb. 3.13:** Korrespondenzanalyse der Entwicklungsstadien bei *Drosophila*. Die Farben stellen die Stadien dar. Jedes Replikat eines Arrays entspricht einem farbigen Quadrat. Die Cluster sind entsprechend umkreist. Die grauen Punkte repräsentieren die Gene. Das rote Quadrat stellt den Median der Kontrolle (0-4h Embryo) dar. Die farbigen Hilfslinien sind das Transkriptionsprofil „virtueller“ Gene die jeweils der Standardkoordinate einer Bedingung entsprechen.

Die Linien sind entsprechen Verbindungen zu hypothetischen Genen mit einem idealisiertem Transkriptionsprofil, deren gesamte Signalintensität einer Bedingung entspricht. Sie führen vom Ursprung des Diagramms zu der Standardkoordinate, die das ideale Genprofil für die jeweilige Kondition darstellt. Aufgrund dessen entsprechen Gene, die näher an den Hilfslinien und vom Ursprung entfernt liegen, „ihrer“ Bedingung am meisten. Da ein differentielles Verhalten der Gene beschrieben durch ihr Transkriptionsprofil in den Entwicklungsstadien beobachtet und validiert werden konnte, erfolgte mit den 13.812 gefilterten Genen die weitere Analyse. Die Resultate der Korrespondenzanalyse wurden in einer farbkodierten Liste gespeichert, in die der Median der normalisierten Signalintensitäten jedes Gens für die 9 Entwicklungsstadien eingetragen wurde.

### 3.4.2 Statistische Analyse der differentiellen Genexpression

Um den Microarray auf die statistische Signifikanz differentiell exprimierter Gene zu untersuchen, wurde die Software R/maanova<sup>®</sup> implementiert in das R-Statistikpaket verwendet. Wie unter 2.2.13. beschrieben, wurde ein gemischtes ANOVA-Modell („Analysis of variance“) [43] auf die gefilterten Gene angewendet. Dazu wurden die Mittelwerte der Signalintensitäten der Replikate für jedes Entwicklungsstadium mit seinen 3 Wiederholungen als Tabelle in die Software geladen. Neben diesen Werten wurde zusätzlich eine Experimentdesign-Datei eingesetzt, die den experimentellen Aufbau und die Reihenfolge der Versuche beschreibt. Die Analyse hatte den Zweck, die Anzahl signifikant differentiell regulierter Gene im Verlauf der gesamten untersuchten Entwicklungsstadien zu ermitteln, da die sonst übliche Methode der 2-fachen Signaldifferenz keine ausreichend statistische Grundlage besitzt (siehe 2.2.13). Nach der Modellberechnung zeigte sich, daß 6.286 Gene (ANOVA) (siehe Abb. 3.14) bei einem Signifikanzniveau von  $p < 0,001$  differentielle Expression aufwiesen. Dementsprechend sind 45,5% der durch M-CHiPS gefilterten Gene (13.812) während der Entwicklung von *Drosophila* differentiell reguliert.












**Abb.3.14:** Vulkandiagramm differentiell regulierter Gene bei *Drosophila*. Dargestellt sind die negativen  $\log_{10}$ -transformierten  $p$ -Werte des  $F_1$ -Tests (ANOVA) gegen die Standardabweichung (SD) des Quotienten der Signalintensitäten. Die horizontale Linie repräsentiert die nominale Signifikanzschwelle  $p < 0,001$ . Die vertikale Linie ist der gruppenweise-korrigierte Signifikanzlevel  $p < 0,01$ . Rote Punkte sind signifikante Gene die den  $F_2$ -Test bestanden, blaue Punkte verfehlten den Test. Die 6.286 signifikant differentiell regulierten Gene sind im oberen rechten Quadranten zusammengefaßt.

### 3.4.3 Analyse der Genregulation mit Self-organizing maps und k-means - Clusteralgorithmen.

Die weitere Auswertung der Genexpression während der Entwicklung von *Drosophila* erfolgte mit den gesamten 13.812 gefilterten Genen und nicht ausschließlich mit den 6.286 differentiell regulierten Genen. Der Grund dafür lag darin, daß das Transkriptom so umfassend wie möglich untersucht werden sollte, also sowohl Gene, die differentiell reguliert werden, als auch konstitutiv oder permanent exprimierte Gene, wie schon für andere Organismen berichtet wurde [57]. Zu diesem Zweck wurde der Median eines Gens aus den Signalintensitäten der drei Hybridisierungen für das jeweilige Stadium durch den Median der Kontrolle (0 - 4 h Embryonalstadium) geteilt und logarithmiert (Basis 2). Die Tabelle mit den log(2)-transformierten Werten wurde mit den zugeordneten Gennamen (Flybase), der Flybase Accession number (Fbgn) und der Heidelberg FlyArray-ID mit Hilfe der PostgreSQL-Datenbank verknüpft (siehe 2.2.13.).

Die Gennamen aus der Tabelle wurden mit dem für *Drosophila melanogaster* verfügbaren Begriffssatz über die Flybase Accession number verknüpft (Christian Busold, pers. Mitteilung) und es konnten 7.452 GO-IDs den gefilterten Genen zugeordnet werden (siehe 2.2.13). Mit Hilfe dieser GO-IDs wurden Ober- und Unterbegriffe definiert, die Gene funktionell charakterisierten (siehe Tab. 3.3) und über das Farbschema in die weiteren Clusteranalysen einbezogen wurden.

GO term	Colour
Development	
Cell organization/cytoskeleton	
Energy pathway/mitochondrion	
Structural insecta	
Muscle	
Protein regulation/metabolism	
RNA regulation/transcription	
DNA replication/Chromatin	
Cell communication/Signal transduction	

**Tab. 3.3 Gene Ontology-Terminologie. Den Gene Ontology-Begriffen (GO) wurden die in der Tabelle dargestellten Farben zugewiesen. Die Farbkodierung wurde in der weiteren Clusteranalyse für die Markierung der Genprofile beibehalten.**

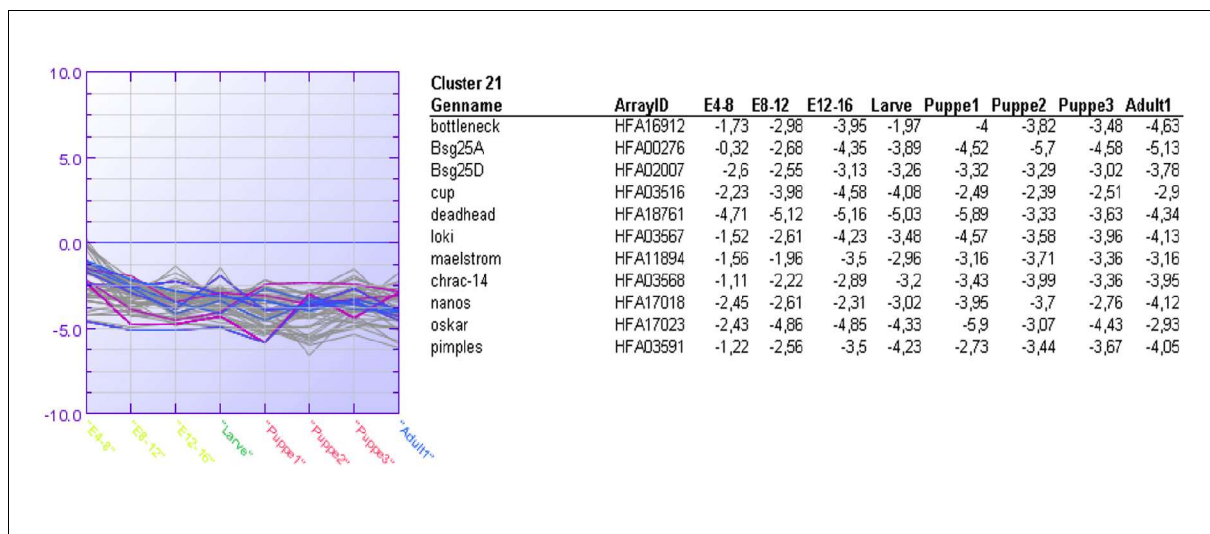
Als nächster Schritt wurde die Datentabelle in die J-Express™-Software geladen und mit 2 Cluster-Methoden analysiert. Das Clustering von Genexpressionsdaten ermöglicht es, ähnliche Gene in Gruppen einzuteilen und bestimmte Muster in der Expression zu erkennen. Durch die Zuordnung der Gene in Gruppen lassen sich Homologien und gemeinsame Regulation bestimmter Gene ermitteln [58]. Ebenso können regulatorische Unterschiede anhand der Transkriptionsprofile beobachtet werden. Die gefilterten Gene wurden daher auf ihr Expressionsmuster im Verlauf der Entwicklung von *Drosophila*, vom Embryo zur adulten Fliege, untersucht. Von besonderem Interesse waren Gene- oder Gengruppen, die spezifisch für ein bestimmtes Stadium exprimiert oder reprimiert wurden.

Zum Einsatz kam zuerst der Self-organizing maps (SOM) – Algorithmus, eine Methode die auf einem neuronalen Netzwerk beruht und sich für explorative Datenanalyse gut eignet [59]. Dabei wird ein hexagonales Gitter bestehend aus „Knoten“ über die Datenpunkte in einem zweidimensionalen Diagramm gelegt, und die Gene diesen Knoten aufgrund ihrer Ähnlichkeit zugeordnet. Daraufhin wurden die initiierten Partitionen mit einem k - means - Algorithmus in Cluster eingeteilt, um ähnliche Expressionsmuster zu identifizieren und die Transkriptionsprofile zu klassifizieren (siehe 2.2.13.). Insgesamt wurden 121 Cluster mit unterschiedlicher Genzahl ermittelt. In diesen Clustern wurde dann nach Vorkommen der verwendeten GO-Begriffe (siehe Tab. 3.3) mit übereinstimmenden Genfunktionen gesucht, um festzustellen, welche Gene in einem Stadium signifikant reguliert sind und ob funktionell ähnliche Gruppen von Genen homolog exprimiert werden [60,61]. Aufgrund der großen Anzahl der Gene in den Clustern wurden jeweils die Cluster mit der höchsten Prozentzahl übereinstimmender GO-Begriffe und die funktionell interessantesten Gene stellvertretend selektiert und dargestellt. Die kompletten Cluster und die Liste der 13.812 Gene mit den logarithmierten Verhältniswerten für jedes Gen sind auf CD vorhanden.

## Embryonalstadium

Der Verlauf der Transkriptionsprofile in Abb.3.16 zeigt, eine starke Reprimierung der Genexpression in den 8 Entwicklungsstadien von *Drosophila*. Das bedeutet jedoch gleichzeitig, aufgrund der logarithmischen Darstellung der Werte, eine starke Expression der Gene im frühesten Stadium, nämlich dem Kontrollstadium des 0 - 4 h alten Embryos.

Da das Signalverhältnis jeweils aus dem Wert des Entwicklungsstadiums geteilt durch den Wert der Kontrolle (0-4 h Embryo) errechnet wurde, ergaben sich für die ersten frühen Gene die gezeigten Transkriptionsprofile.



**Abb.3.16:** Transkriptionsprofile früher embryonaler Gene. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien geteilt durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Die Gennamen der farbig markierten Gene und die zugehörige HDFlyArray-ID sowie die Clusternummer sind in der Tabelle rechts eingetragen. Beispielhaft sind die interessantesten Gene dargestellt.

Die beiden Gene *oskar*, *nanos*, sowie *loki* gehören zu der Gruppe der maternalen Gene und besitzen eine hohe Transkriptionsrate in diesem Stadium. Über den ganzen Entwicklungsverlauf bleiben die Gene in Abb. 3.16 nur im frühesten Stadium hochreguliert. Im weiteren Verlauf (Abb. 3.17) zeigen sich Gene, die für das Embryonalstadium spezifisch hohe Genexpressionen aufweisen. In Abb. 3.17.A zeigt sich der Verlauf für Gene im 4 -8 h Embryonalstadium. Frühe Entwicklungsgene, wie *Notch*, *patched* und *armadillo* zeigen in dieser Phase ihre höchste Expression, ebenso wie *Tollo*, *Toll-6* und *failed axon connection* und im Embryonalstadium 8 -12 h (Abb. 3.17.B) *Toll-7* und *unc-5*. Im 12 – 16 h Stadium (Abb. 3.17.C) werden einige insektenspezifische Gene exprimiert, wie *aristaless*, *a methyl dopa resistant* und *Larval visceral protein L*. Die Profile mancher Gene überschneiden sich teilweise zwischen den Embryonalstadien, bzw. eine exakte Auflösung ist aufgrund redundanter Funktionen der Gene nicht möglich.

In Abb.3.18 sind die Transkriptionsprofile von Genen entsprechend ihrer

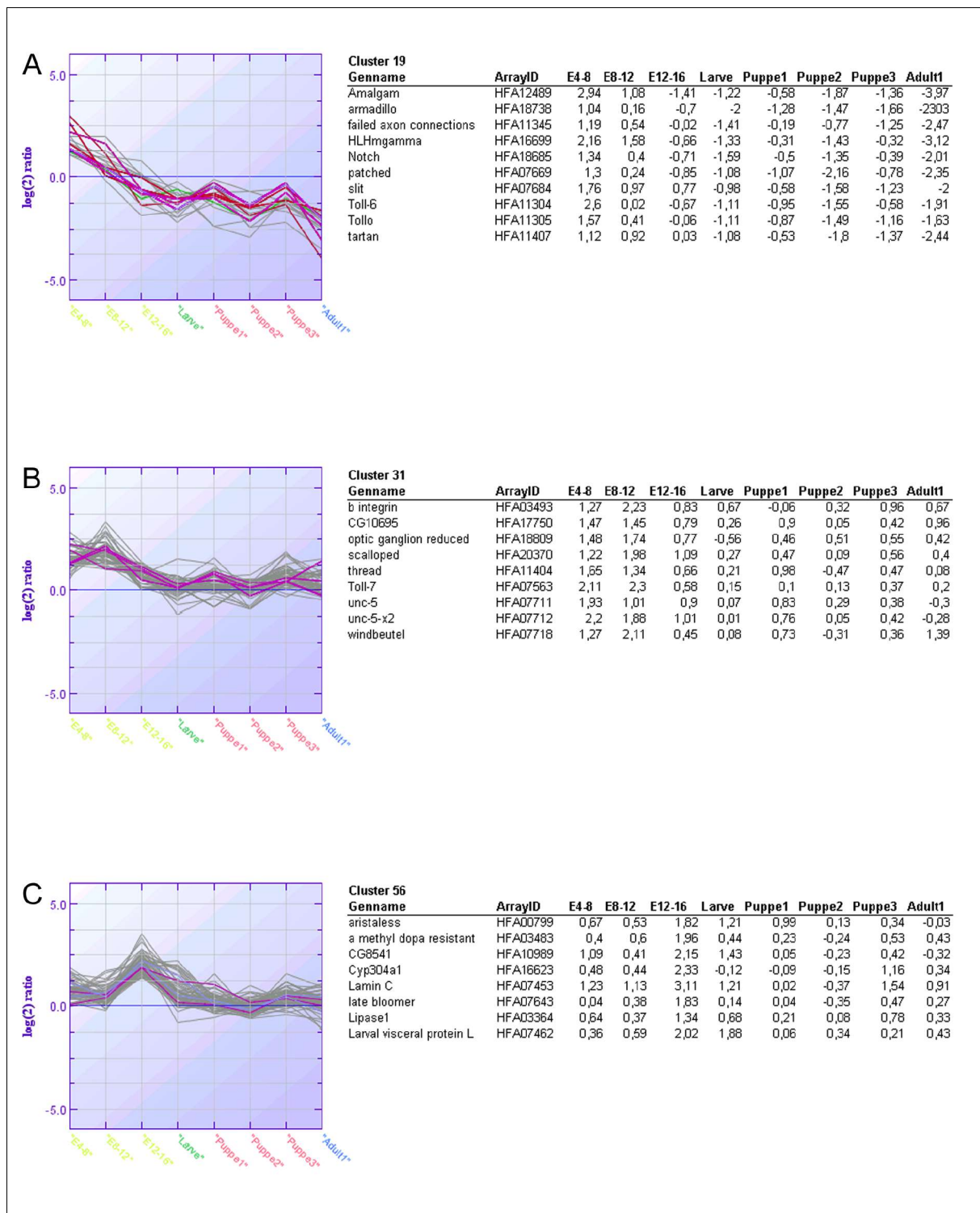
Funktionszugehörigkeit dargestellt. Es zeigt sich an der Reihenfolge der Genexpressionsprofile, das in der sehr frühen Phase der Embryonalentwicklung neben entwicklungsspezifischen Genen (Abb.3.18.B) viele Gene mit Zellfunktionen wie Chromatinregulation (Abb. 3.18.A), Organisation der Zellorganellen (Abb.3.18.C) oder Regulation der Transkription (Abb.3.18.D) beteiligt sind. Diese Gengruppen zeigen ihre höchste Transkription im 0 - 4 h und 4 - 8 h Embryonalstadium und die Cluster weisen untereinander starke Ähnlichkeiten im Profilverlauf auf. Ebenso werden zu dieser Zeit Gene aktiviert, die am Aufbau des Proteasom-26 Komplex beteiligt sind (Abb. 3.18.E).

Neben den Genen, die an der Proteinproteolyse beteiligt sind, findet man zu einem späteren Zeitpunkt (4 – 16 h) eine erhöhte Expression von ribosomalen Genen und Genen des Proteinmetabolismus (Abb.3.18.F). Zum Larvenstadium hin sinkt die Transkriptionsrate dieser Gene abrupt.

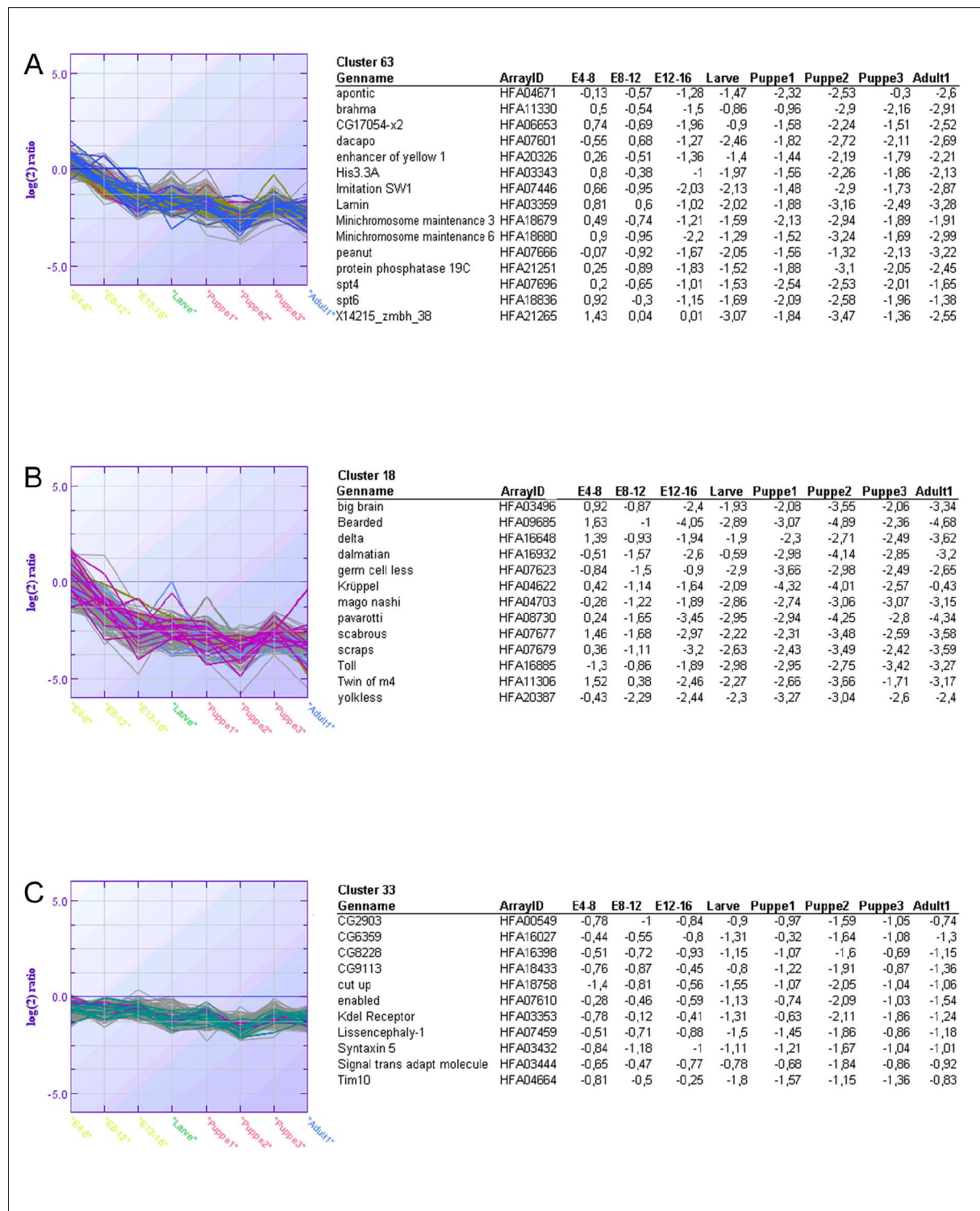
### **Larvenstadium**

Das Larvenstadium nimmt im Leben der Insekten eine wichtige Position ein, denn es ist die Phase des Übergangs zwischen Embryo und Puppe an deren Ende die Metamorphose eingeleitet wird. Dies wird deutlich bei Abb. 3.19.A und Abb. 3.19.C, bei denen larvenspezifische Gene extrem hochreguliert werden, wie *Ecdysone-induced gene 71*, *Ecdysone-induced proteine 28* sowie *Larval cuticle proteine* (Hautbildung) und *Salivary gland secretion* (Speicheldrüse) Gene. Neben proteolytisch wirksamen Genen wie *Lysozym E*, *Lysozym P* und *z Trypsin* (Abb. 3.19.A) finden sich in Abb. 3.19.B sehr hohe Mengen ( $\log(2) > 5.0$ ) an Serin Proteasen und weiteren Trypsinen ( $\alpha$ -,  $\gamma$ -, und  $\delta$ -Trypsin).

Wie in Abb. 3.19.D werden Gene aus dem *Larval serum proteine complex 1 - 2* (Larvenserum Proteinkomplex) zu Beginn der Larvenphase sehr schnell und in großer Menge ( $\log(2) > 8.0$ ) transkribiert. Die Transkriptmenge sinkt jedoch langsam ab und verbleibt auch in den nachfolgenden Stadien auf hohem Niveau.

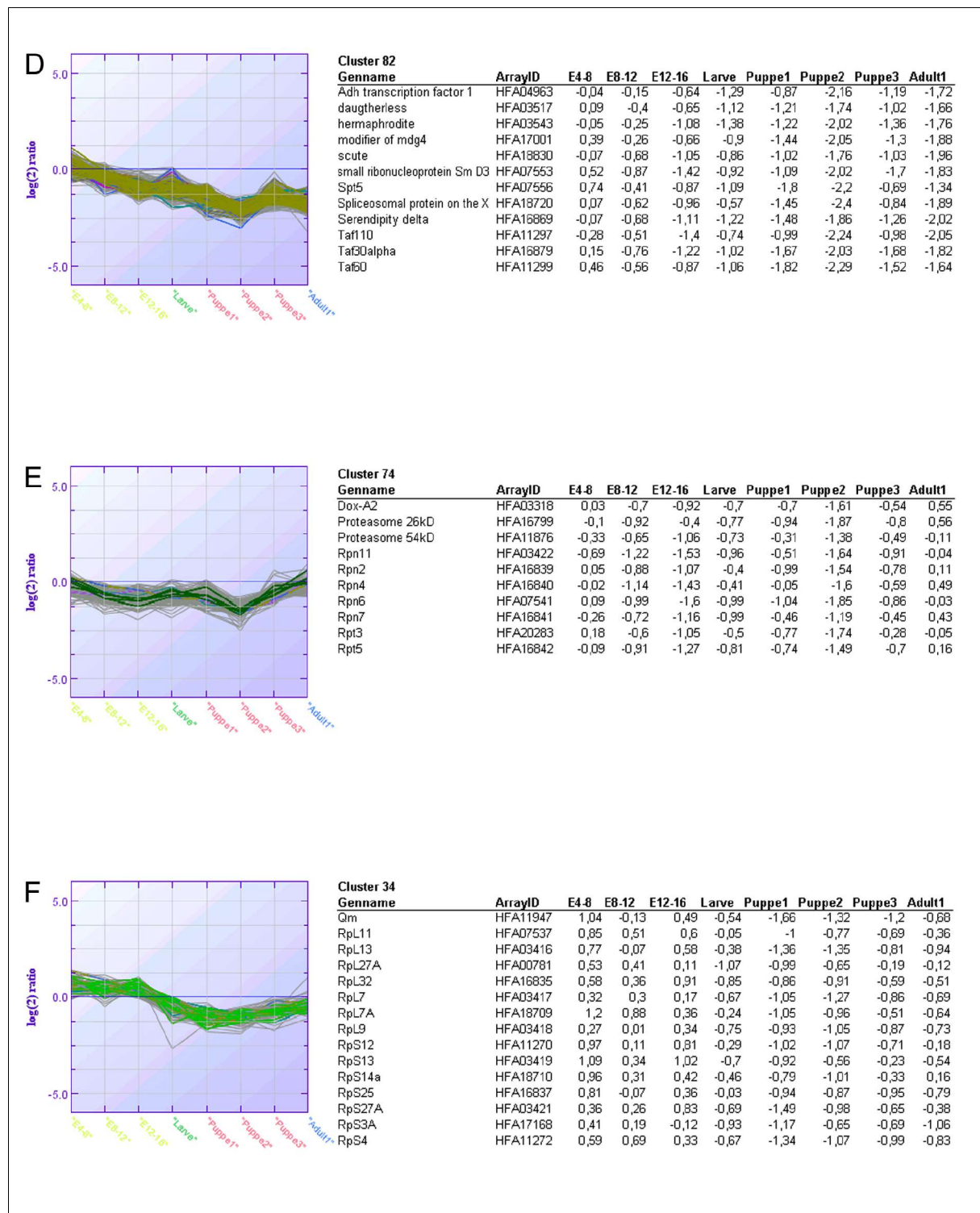


**Abb.3.17:** Transkriptionsprofile der Embryonalstadien. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. A) 4-8 h Embryo. B) 8-12 h Embryo C) 12-16 h Embryo. Die Gennamen der farblich markierten Gene und die zugehörige HDFlyArray-ID sowie die Clusternummer sind in der Tabelle rechts eingetragen.

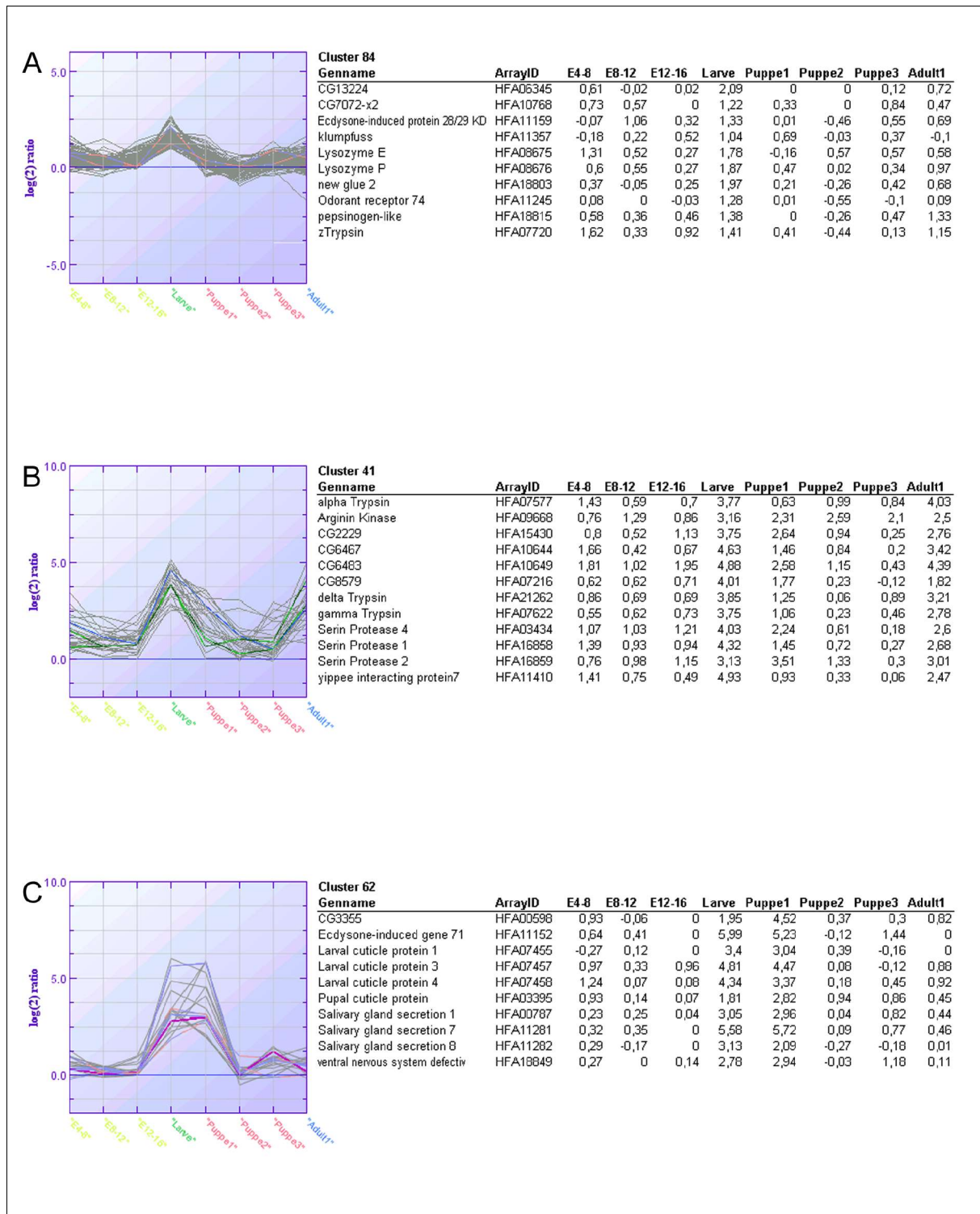


**Abb.3.18 A-C:** Funktionelle Transkriptionsprofile der Embryonalstadien. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Die Zuordnung erfolgte gemäß der GO-Terminologie. A) Chromatin regulierte Gene. B) Entwicklungsgene. C) Proteintransport.

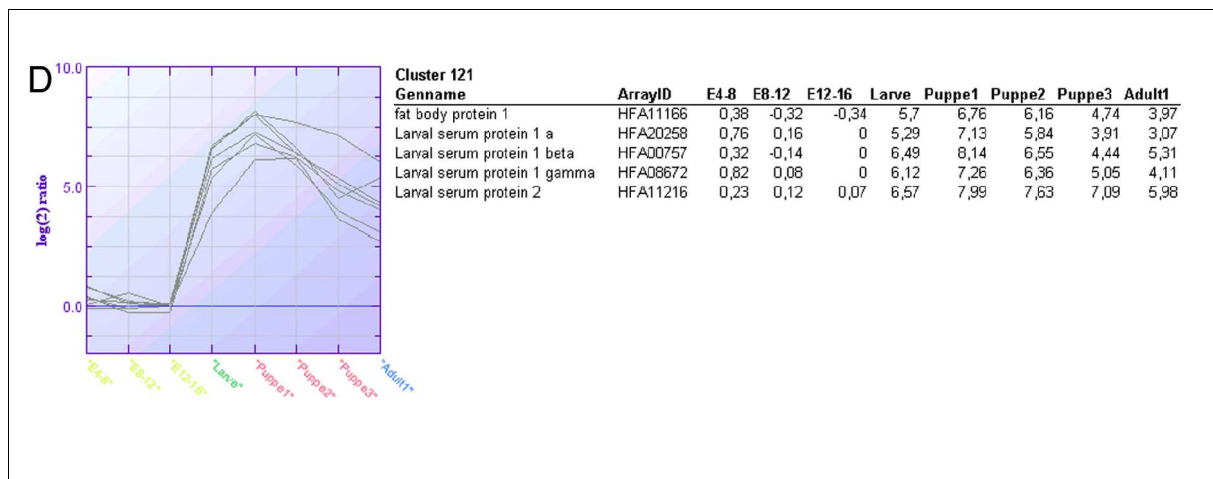




**Abb.3.18 D-F:** Funktionelle Transkriptionsprofile der Embryonalstadien. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Die Zuordnung erfolgte gemäß der GO-Terminologie. D) Transkriptionsregulation. E) Proteolyse. F) Proteinmetabolismus.



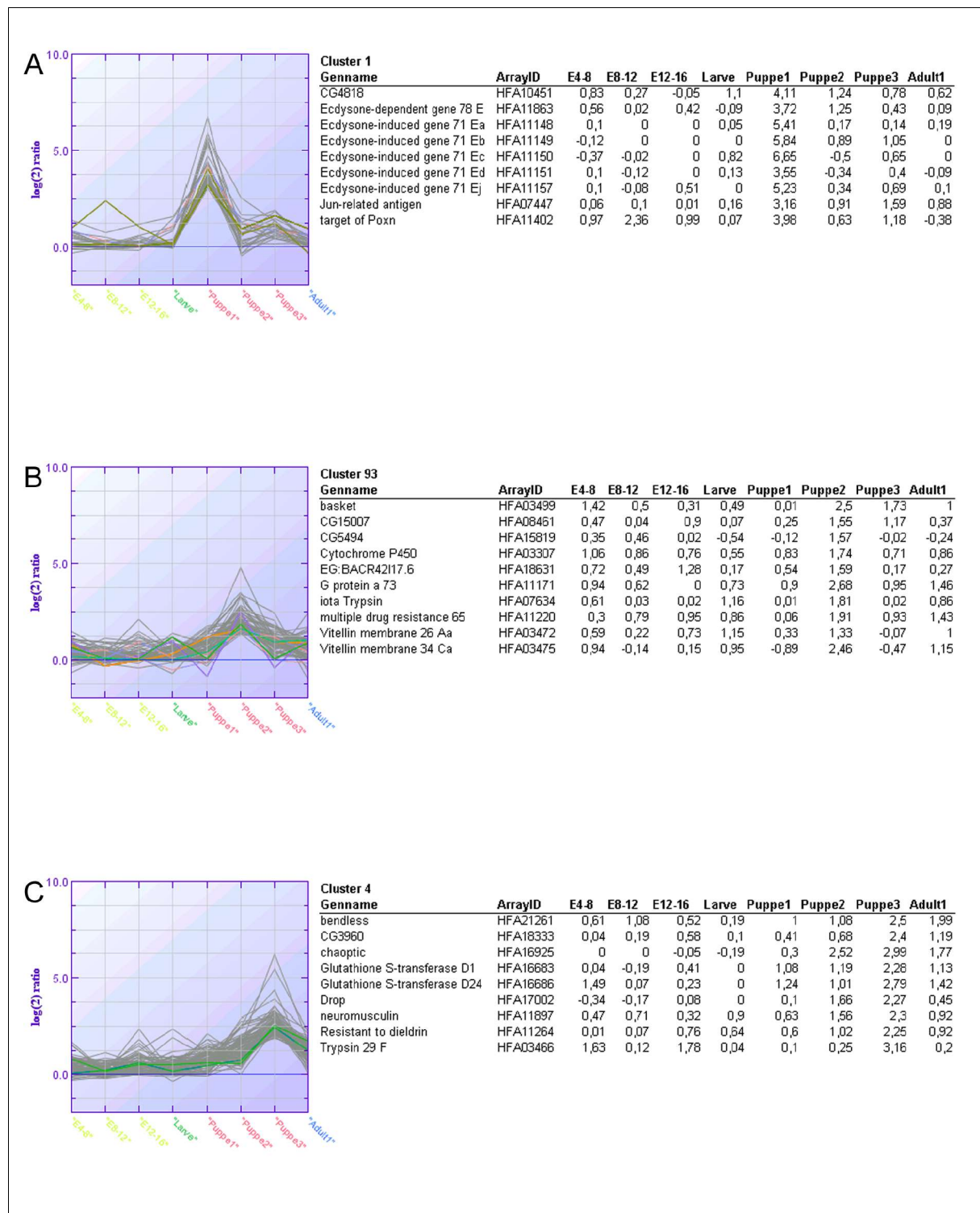
**Abb.3.19 A–C:** Transkriptionsprofile des Larvenstadiums. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0–4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Die 3 Larvenstadien sind experimentell zusammengefasst. Die Gennamen der farblich markierten Gene und die zugehörige HDFlyArray-ID sowie die Clusternummer sind in der Tabelle rechts eingetragen.



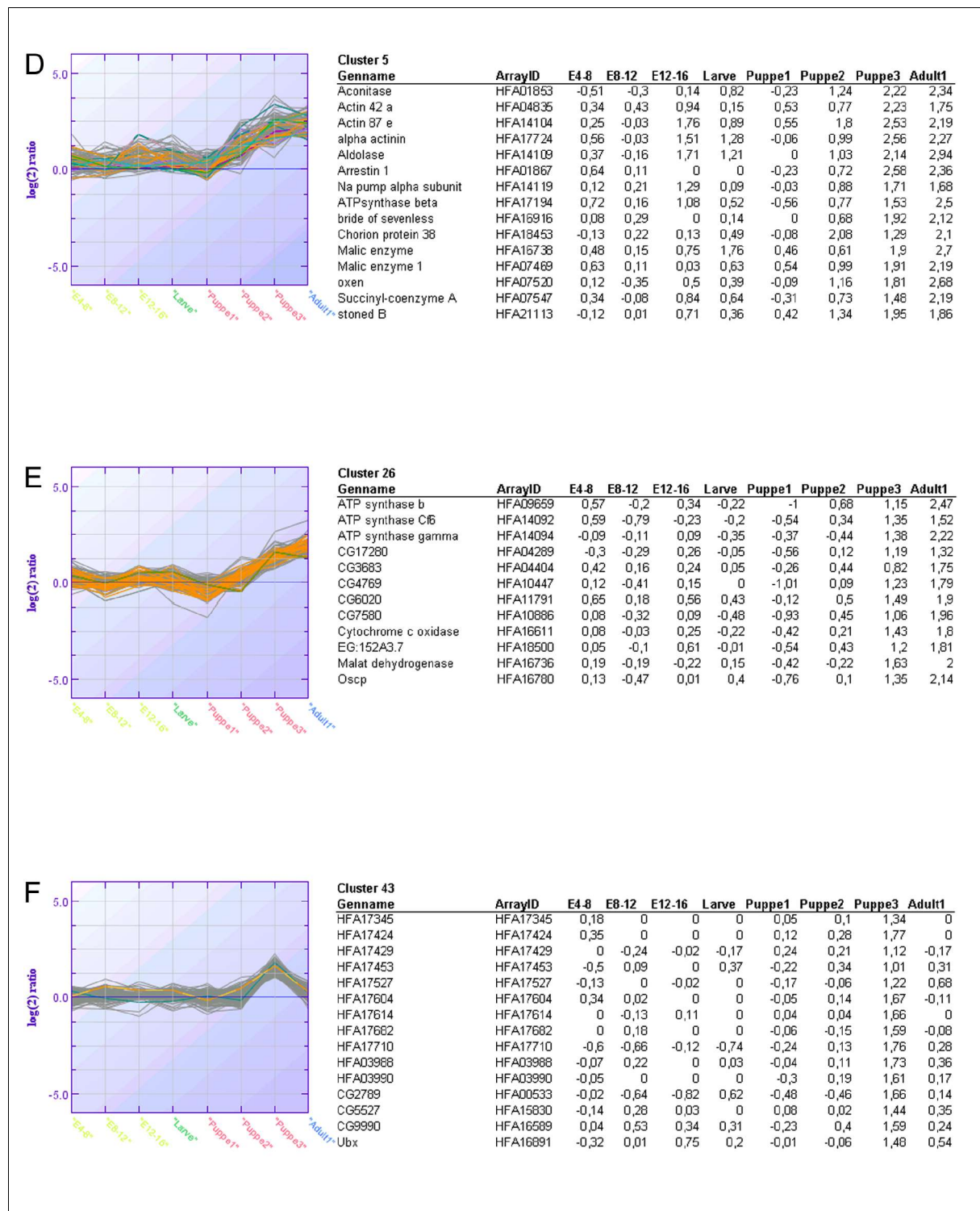
**Abb.3.19 D:** Transkriptionsprofile des Larvenstadiums. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Die 3 Larvenstadien sind experimentell zusammengefasst. Die Gennamen der farbig markierten Gene und die zugehörige HDFlyArray-ID sowie die Clusternummer sind in der Tabelle rechts eingetragen.

### **Puppenstadium**

Im Verlauf der Verpuppung setzt sich der Transformationsprozess zur erwachsenen Fliege fort. Wie in Abb. 3.20.A zu erkennen ist, werden am Übergang vom Larvenstadium zum Vorpuppenstadium hohe Transkriptionsraten von *Ecdysone-induced gene 71* gebildet. Neben Genen für die Entwicklung der Fotorezeptoren in den Augen, *Jun-related antigen* (Abb. 3.20.A) und *bride of sevenless* (Abb. 3.20.D) werden mit Beginn der Verpuppung und während der 3 Puppenstadien verstärkt Gene exprimiert, die für die Ausbildung und Modellierung des Cytoskelettes verantwortlich sind. Vor allem gegen Ende der dritten Puppenphase zeigt sich eine starke Expression von Genen wie *neuromusculin* (Abb. 3.20.C) und Aktin-Genen (Abb. 3.20.D). Verstärkt werden ebenfalls Gene aus dem Transport funktioneller Gruppen exprimiert, wie *Glutathione S-transferase D1,D24* (Abb. 3.20.C) sowie Gene aus dem Kohlehydratmetabolismus (z.B. der Glykolyse) *Aldolase*, Malat-spaltenden Enzymen und *Malat-Dehydrogenase* (Abb. 3.20.E), außerdem Gene des Citratzyklus wie *Succinyl-coenzyme A* (Abb. 3.20.D). Eine weitere auffällige Gruppe mit hoher Transkription in diesen Entwicklungsstadien ist mit dem Energiemetabolismus in der Zelle verbunden. Dazu gehören mitochondriale Gene wie *Na pump alpha subunit* und *oxen* (Abb. 3.20.D) und die an der Atmungskette beteiligten Gene wie *NADH-Dehydrogenasen*, *Cytochrom-c Oxidase* und *ATP-Synthase*-Untereinheiten (Abb. 3.20.E). Wie erwähnt befanden sich in dem Gensatz der Heidelberg Kollektion R1 auch bisher unbekannte Gene, die in der BDGP Annotation (Version 3.1) nicht vorhanden waren. In Abb. 3.20.F ist ein Cluster dargestellt, in dem 87% dieser unbekanntenen Gene spezifisch im letzten Puppenstadium eine signifikant hohe Transkriptionsrate aufweisen.



**Abb.3.20 A–C:** Transkriptionsprofile der Puppenstadien. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Das Puppenstadien wurde in 3 Phasen unterteilt. A) Puppe 1. B) Puppe 2. C) Puppe 3. Die Gennamen der farbig markierten Gene und die zugehörige HDFlyArray-ID sowie die Clusternummer sind in der Tabelle rechts eingetragen.



**Abb.3.20 D–F:** Funktionelle Transkriptionsprofile der Puppenstadien. Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. Die Zuordnung erfolgte gemäß der GO-Terminologie. D) Glykolyse- und Kohlehydratmetabolismus. E) Energiemetabolismus und Mitochondrium. F) Unbekannte Gene der Heidelberg Prediction R1.

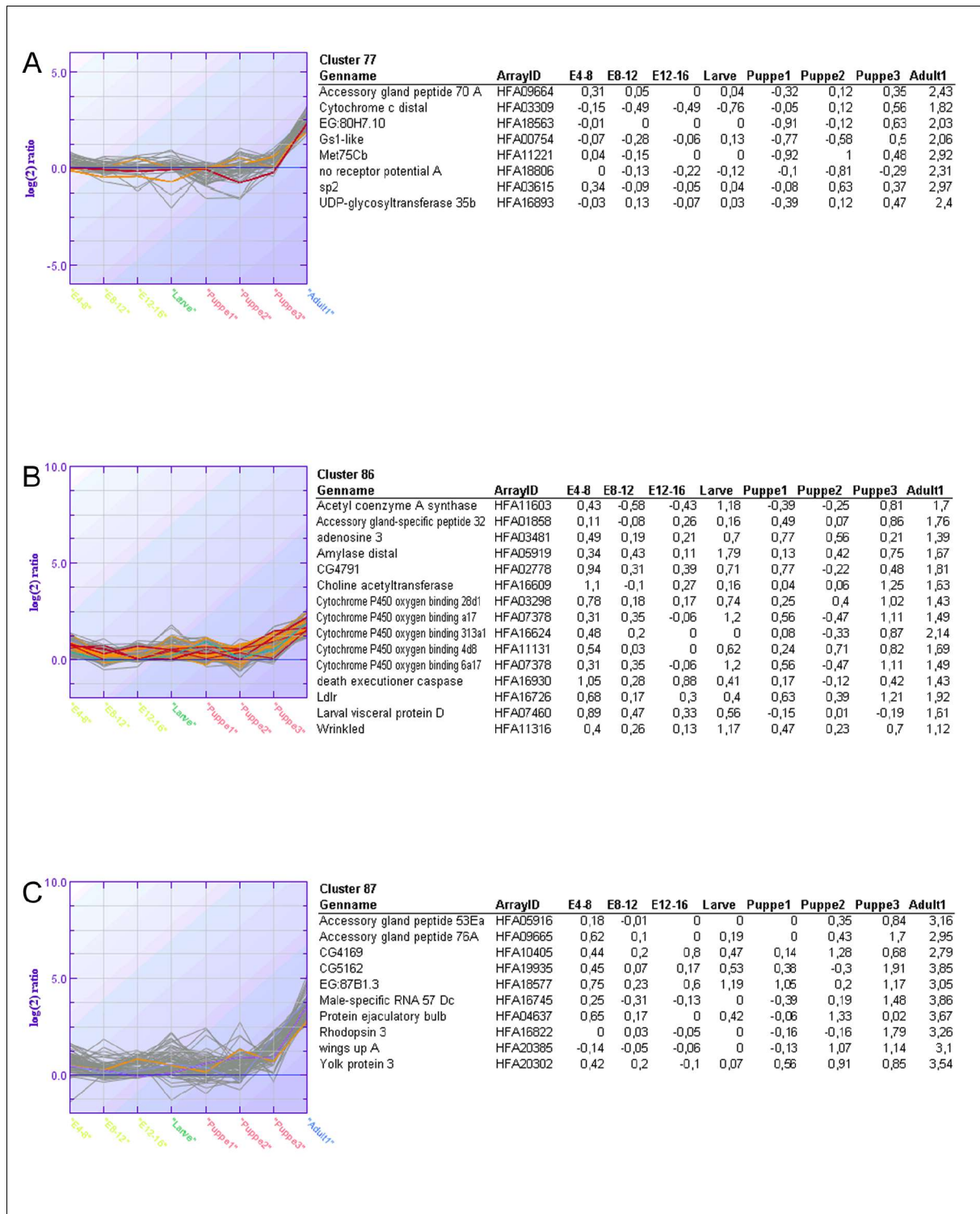
**Adultstadium**

In der jungen erwachsenen Fliege werden, wie in den Puppenstadien, Gene aus dem Cytochrom-Komplex (Abb. 3.21.A) und der Atmungskette z.B. *Cytochrome P450 binding genes* (Abb. 3.21.B) hochreguliert. Außerdem finden sich die gehäuft Gene, die mit der Sexualdifferenzierung verbunden sind, wie die *Accessory gland peptides* (Abb. 3.21.A, 3.21.B und 3.21.C), *Male-specific RNA 57 Dc* und *yolk protein 3* (Abb. 3.21.C).

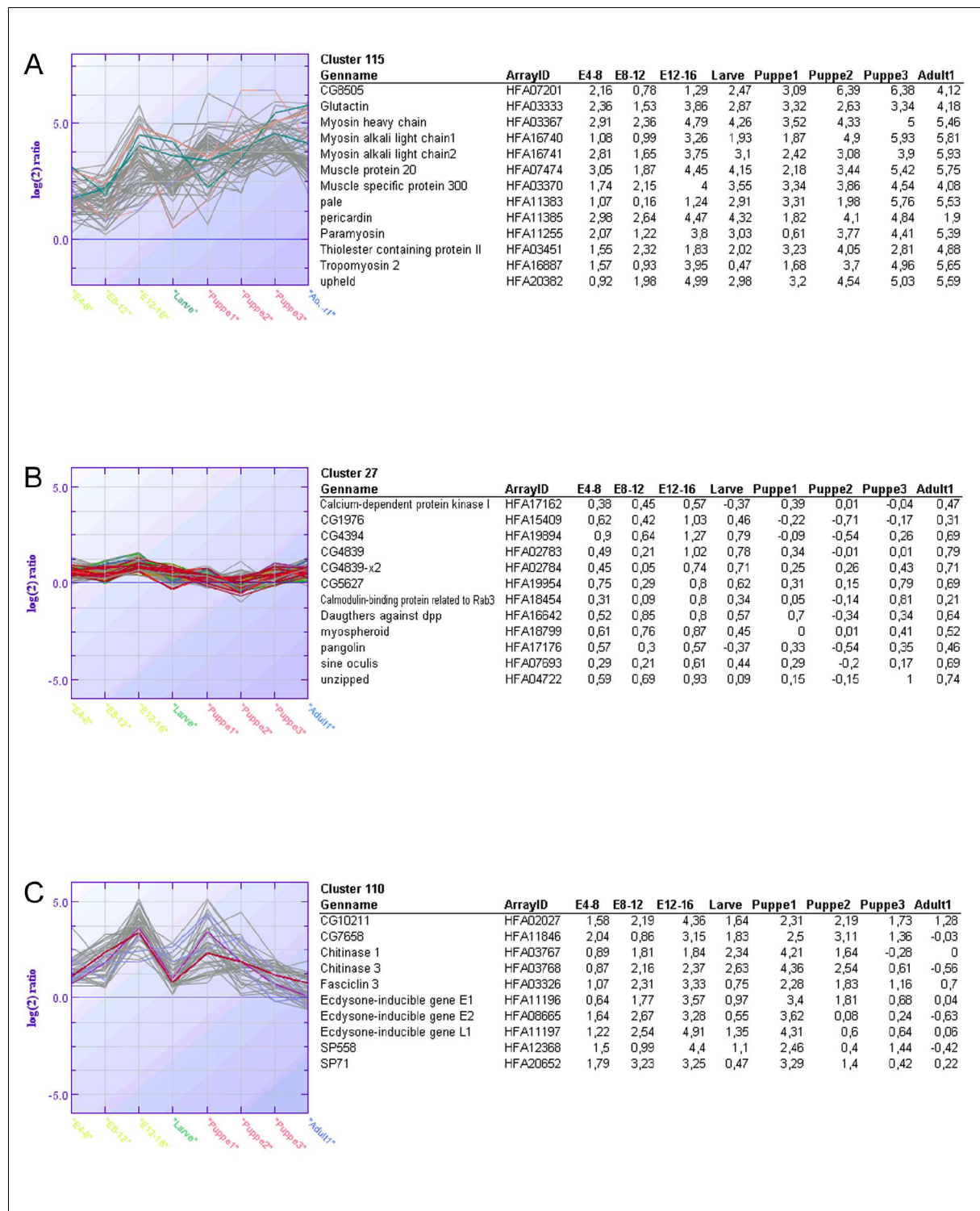
In diesem Stadium waren allgemein die wenigsten Cluster mit hohen Transkriptionsprofilen zu finden.

**Sonstige Cluster**

Neben den für jede Entwicklungsphase spezifisch-transkribierten Genen zeigten sich auch signifikant hohe Expressionsraten bei über alle Stadien der Entwicklung konstitutiv regulierten Genen. Dabei konnte auch Gewebs-spezifische Genregulation identifiziert werden, wie in Abb. 3.22.A die Muskel-spezifischen Gene mit einer hohen Transkriptionsrate. Auf wesentlich niedrigerem Niveau, aber ebenfalls konstitutiv exprimiert werden manche Signaltransduktionsgene mit Kalzium-bindender Funktion (Abb. 3.22.B). Auch Cluster mit mehr als einem Optimum, wie in Abb. 3.22.C wo die höchste Genexpression direkt vor- und -nach dem Larvenstadium vorhanden ist.



**Abb.3.21:** Transkriptionsprofile der adulten Fliege. Aufgetragen sind die  $\log_2$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. A – C) Gene der Atmungskette (Energiemetabolismus) und der Sexualdifferenzierung. Die Gennamen der farbig markierten Gene und die zugehörige HDFlyArray-ID sowie die Clusternummer sind in der Tabelle rechts eingetragen.



**Abb.3.22: Stadien-unspezifische Transkriptionsprofile.** Aufgetragen sind die  $\log(2)$ -Werte der jeweiligen Entwicklungsstadien durch die Kontrolle (0-4h Embryo) gegen den Entwicklungsverlauf bei *Drosophila*. A) Muskel-spezifische Gene. B) Signaltransduktionsgene. C) Hochregulierte Gene in zwei Stadien.



---

## 4 Diskussion

### 4.1 Methoden zur Untersuchung der Genexpression

#### 4.1.1 Optimierung der DNA-Chip-Technologie

Die parallele Analyse von mehreren tausend Genexpressionsdaten gleichzeitig ist eine der großen Stärken der DNA-Chip-Technologie [62]. Seit der Etablierung der Microarrays [63] zur Genomanalyse entwickelten sich unterschiedliche Methoden der Herstellung [52, 64] und Optimierungen der Technologie [65,66] mit hoher Geschwindigkeit. Im wesentlichen werden heutzutage zwei Herstellungsmethoden von cDNA-Microarrays verwendet: zum einen die Bindung über ionische Wechselwirkungen, wobei die negativ geladenen Phosphodiester, die als Rückgrat der DNA fungieren, an die positiv geladenen Amino- oder Lysingruppen der modifizierten Glasobjekträger binden. Zum anderen werden Objekträger mit Aldehydgruppen funktionalisiert [67], die Bindung der aminomodifizierten DNA erfolgt kovalent unter Bildung einer Schiff'schen Base. Für die Anwendung dieser Methode muß die DNA (z.B. PCR-Produkte) zuvor mit Aminogruppen als Adaptoren versehen werden, z.B. durch Verwendung von amino-modifizierten Primern während der PCR. Diese Herstellungsweise könnte für den Heidelberg FlyArray ebenfalls leicht verwendet werden, da für die zweite PCR zu Amplifikationszwecken universelle Primer (siehe 2.2.1.) benutzt wurden, die kostengünstiger als spezifische Primer synthetisiert, und im Anschluß mit Aminogruppen gekoppelt werden können. Aufgrund der Erfahrung mit aminosilanisierten Microarrays (Poly-L-Lysin, QMT Amino slides) [68] und den darauf entwickelten Protokollen wurde dieser Variante des DNA-Chips der Vorzug gegeben. Für den Heidelberg FlyArray wurde eine hohe Anzahl von DNA-Fragmenten (47.616) auf einen Glasobjekträger aufgetragen. Dazu mußte die verwendete Flüssigkeit, in der die DNA gelöst ist („Spotting solution“) Eigenschaften aufweisen, wie hohe Rasterdichte der Spots und geringe Evaporation.

Das Prinzip der Genexpressionsanalyse durch die Hybridisierung der markierten Gesamt-RNA-Probe auf einem Microarray beruht auf der Annahme, daß die detektierte Signalintensität eines Gens proportional zu dessen Transkriptionsrate ist [69]. Der dynamische Bereich der detektierbaren Signalintensität verläuft über drei Größenordnungen unterschiedlicher DNA-Konzentrationen linear und die kleinste detektierbare mRNA-Menge betrug laut [70] 2 pg. Die maximal detektierbare RNA-Menge entspricht dem Konzentrationsverhältnis, an dem die Signalintensität trotz steigender Konzentration der markierten cDNA-Proben nicht weiter zunimmt da die Menge immobilisierter DNA auf dem

Microarray zu gering sein. An diesen Punkten ist eine Sättigung der immobilisierten Probe auf dem DNA-Chip erreicht und eine Quantifizierung der DNA-Konzentration ist nicht mehr möglich. Aus diesem Grund muß die Konzentration der immobilisierten DNA deutlich im Überschuß vorliegen [71]. Um dies zu erreichen, mußten vorliegende Protokolle dahingehend optimiert werden, möglichst viel DNA auf einer kleinen Fläche zu binden. Es wurden DNA-Konzentrationen zwischen 200 und 500 ng/ $\mu$ l auf einem durchschnittlichen Spot-Durchmesser von 80  $\mu$ m aufgetragen. Das von Frank Diehl entwickelte Betain/SSC-Puffersystem [52] diente dazu als Grundlage. Das Betain besitzt einen stabilisierenden Effekt auf DNA und Proteine [72] und führt zu einer Angleichung der Stabilität von A-T- und G-C-Basenpaarungen [73]. Deswegen wird es ebenfalls zur Optimierung von PCR-Reaktionen [74,75] als Additiv [76] eingesetzt. Durch den Zusatz von Betain wird die Viskosität des Spots auf der Glasoberfläche des Microarray erhöht, wodurch ein Austrocknen des Tropfens verzögert wird. Da der Spot länger feucht bleibt, kann sich die darin enthaltene DNA gleichmäßiger über die Fläche des Spots verteilen, was zu einer höheren Homogenität der Spotmorphologie führt. Ähnliche Effekte wurden bei der Verwendung eines DMSO/SSC-Puffers beschrieben [77]. Zusätzlich wird durch die verringerte Evaporation die Reaktionsdauer verlängert, mit der die DNA durch ionische Wechselwirkung an der Glasoberfläche binden kann. Als Konsequenz daraus wird insgesamt mehr DNA an das Trägermaterial gebunden. Der Betain/SSC-Puffer wurde um den Zusatz von  $\text{NaPO}_4$ -Puffer (pH 8,5) erweitert. Durch die Zugabe des Phosphatsalzes zu der Lösung konnte die Bindungsfähigkeit der DNA an die Objektträgeroberfläche um den Faktor drei gegenüber dem Puffersystem Betain/SSC erhöht werden. Die Wirkung des Phosphates läßt sich zum einen dadurch erklären, daß sich die Konformation der DNA dahingehend verändert, daß sie als relaxierte Duplex vorliegt [78,79] und dadurch die negativ-geladenen Phosphatgruppen besser an die Glasoberfläche binden können [80]. Zum anderen bewirkt der leicht basische pH-Wert von 8,5 des Phosphat-Puffers in der Spotting-Lösung vermutlich eine Deprotonierung der Phosphatreste und damit eine erhöhte negative Ladungsdichte in den Phosphatgruppen der DNA. Somit verstärkt sich die Wechselwirkung zwischen dem Phosphatrückgrat der DNA-Stränge und der positiven Aminogruppen der Glasoberfläche [81], was zu Folge hat, daß mehr DNA pro Fläche gebunden werden kann. Weiterhin wurden die verwendeten PCR-Produkte mit Hilfe der modifizierten Spotting-Lösung ohne weitere Aufreinigung auf die Objektträger aufgebracht, so daß dem erhebliche Verlust von bis zu 50% der DNA-Ausgangsmenge während des Aufreinigungsprozesses vorgebeugt werden konnte. Obwohl die bei der PCR eingesetzten Primer ebenfalls an den Glas-Chip binden könnten, da sie nicht durch die Aufreinigung entfernt wurden und in bis zu 5-fachem molarem Überschuß gegenüber dem PCR-Produkt vorliegen, kommt es nicht zu einer Konkurrenz um Bindungsstellen auf der aktivierten Glasoberfläche, da aufgrund der wesentlich höheren Fragmentgröße hauptsächlich PCR-Produkte gebunden werden. Sollte

doch ein geringer Anteil der Oligonukleotide an die Oberfläche gebunden werden, so ist der Anteil an der gesamten Signalintensität nicht signifikant genug und die Bindung der Oligonukleotide zu instabil [53]. Ein wesentlicher Vorteil des Betain/SSC/Phosphat-Puffersystems ist die geringe Verdunstungsrate bedingt durch das Betain. Während des Spotting-Vorganges überträgt eine Nadel des Microarray-Roboters ein Volumen von je 0,5 – 1 nl Lösung auf die Glasoberfläche, welches bei einer Spottingdauer von ca. 1 Stunde pro 384er Platte ohne Betainzusatz zu starken Verdunstungen führt. Ebenso verdunsten bereits aufgetragene Spots wesentlich langsamer während der DNA-Chip-Herstellung. Der Nachteil der reduzierten Verdunstung durch den hygroskopischen Effekt ist jedoch ein höherer Spotdurchmesser (100 – 120  $\mu\text{m}$ ). Damit erhöht sich ebenfalls der Spot-zu-Spot-Durchmesser. Die Ausbreitung der Tropfen wird durch die hydrophoben Eigenschaften der Glasoberfläche der Microarrays bestimmt [82]. Je hydrophober die Oberfläche, desto dichter lässt sich die hydrophile Spotting-Lösung aufbringen. Zu diesem Zweck wurden die zuvor benutzten Poly-L-Lysin-Slides [83] mit aminosilanisierten QMT Amino-Slides verglichen. Da die QMT Amino-Slides wesentlich hydrophobere Oberflächeneigenschaften besitzen (pers. Mitteilung Quantifoil) als die Poly-L-Lysin-Slides, konnten Spots mit einem Durchmesser von ca. 80  $\mu\text{m}$  generiert werden. Ein Zusammenfließen der Tropfen wurde so verhindert und damit die spätere Auswertbarkeit der Arrays gewährleistet.

Durch die Kombination der genannten optimierten Methoden konnte daher der Heidelberg FlyArray mit 47.616 Spots auf einem Glasträger in hoher Qualität reproduzierbar hergestellt werden.

#### **4.1.2 PCR-Amplifikation der Heidelberg Kollektion R1 und Konstruktion des Heidelberg FlyArray**

Ausgehend von der Heidelberg Kollektion R1, die aus Annotationen der BDGP Genom Annotation Version 2.0, der BDGP cDNA Kollektion Version 1.0 und der eigenen Heidelberg Genannotation (HDC) kombiniert wurde, lagen 21.306 zu amplifizierende ORFs („open reading frames“) vor. Insgesamt konnten 20.948 (97,9%) der PCR-Produkte mit einer durchschnittlichen Länge von 500 bp erfolgreich amplifiziert werden. Die hohe Prozentzahl von 97,9% PCR-Produkte amplifiziert von genomischer DNA auf der Grundlage von reinen *in silico*-Vorhersagen der spezifischen Primersequenzen durch die GenomePride-Software [38] war ein Hinweis auf die hohe Genauigkeit des eingesetzten Algorithmus. Die daraus resultierende Genliste, die Heidelberg Prediction, bildete die Grundlage für die kombinierte Heidelberg Kollektion R1 und die nachfolgenden PCR-Amplifikationen. Lag die Amplifikationsrate einer 96er PCR-Platte bei der ersten PCR unter 90%, wurde die gesamte

PCR wiederholt. Bei erneut geringer Rate unter 90% erfolgte nocheinmal eine Wiederholung dieser 96er PCR-Platte. In den meisten Fällen reichte die Anzahl von 2 Wiederholungen aus, um jeweils über 90% der PCR-Produkte zu generieren. Schwer zu amplifizierende Fragmente wurden in neuen Platten zusammengefaßt und mit veränderten Annealingbedingungen zu amplifizieren versucht. Die PCR-Platten aus der ersten Runde mit der höchsten Amplifikationsrate wurden für die zweite PCR verwendet, die zur Reamplifikation des ersten PCR-Produktes diente, um eine ausreichende Menge an Ausgangsmaterial für den Spottingprozess zu erhalten. Im Durchschnitt wurden in der zweiten PCR DNA-Konzentrationen von 150 – 350 ng/µl erreicht und auf die Microarrays gespottet. In diesem Konzentrationsbereich ist die exakte DNA-Menge in einem Spot nicht von Relevanz, wenn Betain in der Spotting-Lösung vorhanden ist [53]. Die Beurteilung der Qualität der zweiten PCR und das weitere Verfahren erfolgte analog zur ersten PCR (siehe 2.2.1). Die wiederholten und neu kombinierten PCR-Platten wurden so zusammengefaßt, daß sie im Microarray-Design einen eigenen Bereich belegten (siehe Abb. 3.11). Neben der Aufteilung jedes Blockes in die drei verschiedenen Bereiche der gespotteten *Drosophila* PCR-Produkte wurden die Kontrollelemente wie *Arabidopsis* cDNAs, genomische *Drosophila*-DNA und Spotting-Lösung jeweils in der obersten und untersten Reihe eines „Grids“ (Blockes) aufgetragen. Durch die verwendeten *Arabidopsis*-mRNAs als externe („spiking“) Kontrollen konnte die Qualität der Markierungsreaktion und der Hybridisierung [84] schnell beurteilt werden. Die markierte und gespottete Herings-DNA diente als Positionsmarker in der ersten Spalte, wodurch die spätere Anpassung der Maske der Auswertungssoftware GenePix™ wesentlich erleichtert wurde, da die einzelnen Blöcke leichter identifiziert werden konnten. Zur Herstellung des Transkriptom-Arrays wurde der Microarray-Roboter MicroGrid II™ (BioRobotics) eingesetzt, der nach dem Kontaktverfahren [85] arbeitet. Der Aufbau des Roboters erlaubt die Beladung mit 24 Platten, die automatisiert auf die Objektträger gespottet werden. Die dabei verwendeten Pins, sogenannte „Split-Pins“ (SMP 3 Pins, TeleChem) besitzen ein Flüssigkeitsreservoir mit einem Volumen von ungefähr 100 nl. Da bei jedem Kontakt der Nadeln mit der Glasoberfläche ca. 1 nl Flüssigkeit übertragen wird, lassen sich mit einer Nadelfüllung theoretisch 100 Spots produzieren. Durch den Einsatz von „Prespots“ vor dem eigentlichen DNA-Auftrag wird diese Zahl jedoch minimiert. Dieser Schritt ist aber unumgänglich, da das Flüssigkeitsreservoir der Nadeln erst nach einigen Berührungen luftblasenfrei ist und die Spotting-Lösung dann kontinuierlich fließen kann.

### 4.1.3 Markierung der RNA

Die Markierung der RNA zu Hybridisierungszwecken kann auf unterschiedliche Arten erfolgen [86,87]. Die Herstellung der fluoreszenzmarkierten cDNA aus der, von mehreren Chargen gepoolten [88] Gesamt-RNA erfolgte über die indirekte Markierungsreaktion, bei der zuerst Aminoallyl-dUTP während der reversen Transkription in die synthetisierte cDNA eingebaut wird. In einem anschließenden zweiten Schritt erfolgt die Kopplung des monoreaktiven Fluoreszenzfarbstoffesters an das Aminoallyl-dUTP. Neben dieser Methode des indirekten Einbaus gibt es noch die Möglichkeit des direkten Einbaus der Farbstoffmoleküle [89] in die cDNA während der reversen Transkription. Diese Markierungsmethode ist schneller als die indirekte, da der Kopplungsschritt der Farbstoffe ausbleibt. Der Nachteil dieser Methode ist jedoch, daß der enzymatische Einbau der bereits markierten Nukleotide aufgrund ihrer Größe durch das assoziierte Farbstoffmolekül erschwert wird. Es wird dabei durchschnittlich ein Fluoreszenzfarbstoff auf 1000 Nukleotide eingebaut. Das indirekte Markierungsverfahren hingegen ermöglicht, aufgrund der geringen Größe des Aminoallyl-Restes wesentlich höhere Einbauraten von einem Farbstoffmolekül auf 30 – 150 Nukleotide [90]. Da sich die Farbstoffmoleküle Cy3 und Cy5 außerdem untereinander in ihrer Größe unterscheiden [91], besteht die Möglichkeit eines uneinheitlichen Einbaus der Farbstoffe während der reversen Transkription. Für die Markierungsexperimente wurde daher die indirekte Markierung verwendet. Um zu gewährleisten, daß auch Transkripte, die in geringer Anzahl vorliegen, markiert werden konnten, wurden jeweils 20 µg Gesamt-RNA pro Farbstoffreaktion eingesetzt. Setzt man bei der reversen Transkription eine Effizienz von 10% [5] voraus, dann ergeben sich daraus 2 µg markierte cDNA pro Reaktion. Dieser Wert korreliert gut mit den erzielten Ergebnissen (siehe 3.3.1.). Für die reverse Transkription der Gesamt-RNA in cDNA wurden Hexamer-Primer verwendet [92], denn dadurch werden auch längere Transkripte statistisch häufiger und damit gleichmäßiger markiert. Auf die Verwendung von Oligo(dT)-Primern, bei deren Einsatz gezielt mRNA in cDNA transkribiert wird, wurde verzichtet. Da bei Einsatz der Oligo (dT)-Primer die Transkription am 3'-Terminus der mRNA beginnt und die Transkriptionsreaktion im allgemeinen nach einer Länge von 2 kb abbricht, können sehr lange Transkripte nur unzureichend und nicht-polyadenylierte mRNA (z.B. Histon-mRNA) [93,94] gar nicht markiert werden. Durch den Einsatz der Hexamer-Oligonukleotide wird neben der mRNA auch ribosomale RNA markiert, die 95% der Gesamt-RNA darstellt. Das bedeutet, daß in den markierten 2 µg cDNA ein Anteil von 100 ng mRNA vertreten ist, der einen Querschnitt aller Transkripte repräsentiert. Durch Verwendung der beschriebenen Transkriptionsmethode konnte ein RNA-Verlust durch den Arbeitsschritt einer mRNA-Isolierung vermieden werden [95].

#### 4.1.4 Hybridisierung der Microarrays

Das nur 24 der 73 Hybridisierungsexperimente für die weitere Analyse genutzt werden konnte lag darin begründet, daß die Protokolle für die Markierung der Proben, Hybridisierung und Nachbehandlung der Microarrays erst noch etabliert bzw. optimiert werden mußten. Zwar wurden die meisten Methoden vor den eigentlichen Experimenten ausführlich getestet (wie z.B. Verwendung der indirekten Markierung mit Aminoallyl-dUTP), viele Fragestellungen oder Probleme ergaben sich aber erst während der eigentlichen Arbeit mit dem Heidelberg FlyArray und den Hybridisierungen der Proben (wie z.B. ein hohes Hintergrundsignal). Die Signalqualität der hybridisierten Microarrays ist ausschlaggebend für die weitere Genexpressionsanalyse. Ein unspezifisch hoher oder inhomogener Hintergrund führt zu einem schlechten Signal-zu-Hintergrund-Verhältnis. Das Problem eines hohen lokalen Hintergrundes an unterschiedlichen Stellen auf dem Array und unspezifische Bindung der Farbstoffe an die Glasoberfläche, anstatt an die DNA („black hole“-Phänomen), traten zeitweise und nicht reproduzierbar auf. Als Ursache dafür wurden nicht gänzlich deaktivierte, funktionelle Aminogruppen auf der Glasoberfläche vermutet. Deshalb wurden die Microarrays nach dem Spottingprozess für 4 h bei 80°C analog zu der Behandlung von Nylonmembranen erhitzt [96], was eine Dehydrierung der Aminogruppen bewirken sollte. Zusätzlich wurden die DNA-Chips vor jeder Hybridisierung mit 1% BSA prähybridisiert, wodurch eine Absättigung der Oberfläche erfolgen sollte [97]. Mit den angewendeten Maßnahmen konnte der unspezifisch hohe Hintergrund weitgehend unterdrückt werden. Die Hybridisierungen mit den markierten cDNAs erfolgten als Zweifarben-Experimente, wobei jeweils zwei unterschiedliche Cy3/Cy5-cDNAs auf einem DNA-Chip gleichzeitig hybridisiert wurden und jede Hybridisierung eines Entwicklungsstadiums dreimal wiederholt wurde. Durch den Austausch der Farbstoffe („dye swap“) zwischen den beiden co-hybridisierten cDNAs innerhalb dieser drei Wiederholungen wurde ein experimenteller Fehler aufgrund der Verwendung nur eines Farbstoffes für ein Stadium verhindert [98]. Um eine hohe Spezifität der Signale zu erreichen, wurde die Hybridisierung bei 55°C durchgeführt, um die Stringenz der Doppelstrangbildung zu erhöhen. Die Qualität der Hybridisierung wurde über die Bestimmung der Standardabweichung der Replikate eines Arrays ermittelt (siehe 2.2.13.). Die Signalqualität konnte erhöht werden durch Verwendung spezieller Deckgläser (LifterSlips™, Menzel Gläser), die durch einen Teflonstreifen („Spacer“) einen Abstand zwischen den Gläsern bildeten. Dadurch konnte ein größeres Volumen Hybridisierungspuffer mit markierter Probe auf den DNA-Chip pipettiert werden. Durch das höhere Volumen (40 µl) kann eine Diffusion der Probe auf der Array-Fläche stattfinden als mit herkömmlichen Deckgläsern, die für diese Fläche ein Volumen von nur 25 µl besitzen, da im Mittel mehr markierte cDNA an die komplementären immobilisierten Proben binden können.

#### 4.1.5 Auswertung der Expressionsdaten

Nach der Detektion und Quantifizierung der Signalintensitäten ist die Normalisierung [99] der Signalintensitätswerte zwischen den Farbstoffen und aus den wiederholten Hybridisierungen der Microarrayexperimente der erste wichtige Schritt der Datenauswertung [100]. Die Normalisierung dient dem Zweck, experimentelle Ungleichheiten während des gesamten Entstehungsprozesses Microarray-generierter Genexpressionsdaten auszugleichen, um aussagekräftige biologische Vergleiche auf Transkriptionsebene zu ermöglichen. Gründe für Ungleichheiten können z.B. Unterschiede in der eingesetzten RNA-Menge, unterschiedliche Effizienz in der Markierung oder Detektion zwischen den verwendeten Fluoreszenzfarbstoffen oder allgemein systematische Fehler sein [101,102]. Es gibt verschiedene Ansätze und Möglichkeiten, eine Normalisierung durchzuführen [103,104,105]. Es wurde bereits im Voraus das Experimentdesign der Hybridisierungen auf die in der von M-CHiPS verwendeten Normalisierungsmethode [9] der linearen Regression [106,107] abgestimmt, denn in allen Hybridisierungsexperimenten wurde eine gemeinsame Referenzkontrolle als Grundlage für die Normalisierung durch die lineare Regression verwendet (siehe 3.4.1.). Durch die Normalisierung der Hybridisierungsdaten wurden die unterschiedlichen Entwicklungsstadien von *Drosophila* auf einer gemeinsamen Basis untereinander vergleichbar. Auf dieser Grundlage wurden die Gene in M-CHiPS nach bestimmten Qualitätskriterien (siehe 2.2.13.) gefiltert, so dass die Expressionsraten der verbleibenden 13.812 Gene in die Korrespondenzanalyse eingehen konnten. Diese Methode aus der Statistik, die die Projektion der Daten in einen zweidimensionalen Raum nutzt, ähnlich der PCA („Principal component analysis“) [26,108,109], eignet sich gut für die explorative Analyse von Genexpressionsdaten aufgrund der Darstellung der Beziehung zwischen Hybridisierungen und den daran beteiligten Genen. Der Nachteil der Korrespondenzanalyse ist die präzise Definition von Grenzen zwischen den Clustern und damit die konkrete Zuordnung von Genen zu distinkten Clustern [110,111], wie es für andere Projektionsmethoden (z.B. PCA) [27] der Fall ist. Aus diesem Grund wurde die Hybridisierungsqualität mit Hilfe von M-CHiPS untersucht und das Clustering der Transkriptionsprofile mit SOM -und k-means-Algorithmen [112] durchgeführt (siehe 2.2.13.). Diesen beiden Algorithmen wurde der Vorzug gegenüber dem häufig verwendeten hierarchischen Clustering („Hierarchical clustering“) [23,113] gegeben, da sich diese Methode häufig als ungenau in der Definition der Cluster [114] erwies.

Die genannten Methoden gehören alle zur Gruppe der sogenannten „unsupervised clustering methods“, die Muster in der Genexpression identifizieren sollen. Eine Alternative sind die „supervised methods“ [115], bei denen bereits vor der Analyse zusätzliche Informationen über die einzelnen Gene in den Clustering-Prozess miteinfließen. Beispiele dafür sind „Support Vector Machines“ (SVM) [116,117] und Klassifizierungsmethoden für

Gene [118,119]. Neben der eigentlichen Clusteranalyse zur Darstellung der Expressionsmuster ist die Regulation differentiell exprimierter Gene von großer Bedeutung [120,121]. Die statistische Auswertung der differentiell regulierten Gene [122] erfolgte anhand einer Varianzanalyse (ANOVA) (siehe 2.2.13.), deren abgewandelte Form („mixed-model ANOVA“) [123,124] durch die zusätzliche Definition von Zufallsfaktoren für die Analyse von Microarrayexperimenten gut geeignet ist. Für die jeweilige biologische Fragestellung sollte die Auswahl der Clustermethoden daher individuell angepaßt werden [125,126].

## **4.2 Untersuchung der Genexpression im Verlauf der Entwicklung von *Drosophila melanogaster***

### **4.2.1 Genannotationen und Transkriptom von *Drosophila melanogaster***

Der Einsatz von Microarrays zu Untersuchung des Transkriptoms von *Drosophila melanogaster* ist aufgrund der weiten Verbreitung dieses Organismus als Tiermodell in der Entwicklungsbiologie von großem Interesse. Frühere Transkriptionsanalysen der *Drosophila*-Entwicklung mit Microarrays zeigten starke morphologische Veränderungen während der Embryogenese oder der Metamorphose einhergehend mit massiven Änderung in der Regulation der Genexpression [127,128]. Weiterhin dienten Microarrays für die Analyse gewebespezifischer- [129,130,131], sowie Zeit- [132,133] oder geschlechtsspezifischer- [134,135] Genexpression. Einige Analysen verwendeten dabei Microarrays, die ungefähr ein Drittel der vorhergesagten Gene der BDGP Genomannotation repräsentierten [127,136]. Die bisherige Annahme bestand darin, daß für die Bildung des Transkriptoms in *Drosophila* ungefähr 14.000 proteinkodierende Gene verantwortlich sind [1]. In dieser Arbeit wurden mit Hilfe des Heidelberg FlyArray die Transkripte von 6.286 differentiell exprimierten- und 13.812 exprimierte Gene während des Lebenszyklus von *Drosophila* nachgewiesen. Für die differentielle Expression wurde in einer vorherigen Studie mit *Drosophila*-Unterstämmen, in der die Evolution der Metamorphose untersucht wurde, ein ähnlicher Wert gefunden [137]. Von den exprimierten Genen entstammten 10.339 der BDGP Genomannotation Version 2.0 und 3.473 Gene der eigenen Heidelberg Kollektion R1. Dies bedeutet einen experimentellen Nachweis von 74,2% der BDGP Annotation und 46,5% der Heidelberg Kollektion. Zur Validierung der Ergebnisse der Microarrayexperimente wurden am ZMBH, Uni Heidelberg, RT-PCR-Experimente und *in situ*-Hybridisierungen durchgeführt.



Bei den RT-PCRs stimmten 93,8% (136/145) der bekannten Gene und 74,4% (218/293) der unbekannt Genen mit den Microarray-Ergebnissen überein. Die Anzahl gefundener Gene aus der Microarrayanalyse weist starke Übereinstimmungen mit der Detektionsrate vorhergesagter Gene in einer Studie mit Teilen des menschlichen Genoms auf. Dabei konnten 80-85% bekannter Gene und 58% vorhergesagter Gene durch Microarray-Analysen validiert werden [138]. Diese Prozentzahlen lassen auf eine Limitierung der Microarrayanalyse in der Detektion kompletter Transkriptome schließen. Ebenfalls zeigt sich die Problematik der experimentellen Überprüfung von *in silico*- vorhergesagten Genen durch Softwareprogramme mit Microarray- bzw. anderen Genexpressionsanalysemethoden. Durch Änderungen z.B. von Exondefinitionen oder Exon-Intron-Grenzen in der Software kam es zu großen strukturellen Änderungen zwischen der BDGP Annotation Version 2.0 und der aktuellen Version 3.1. Dabei änderten sich 85% der Transkripte und 45% der vorhergesagten Proteine [139]. Da die ursprüngliche Heidelberg Kollektion auf der BDGP Version 2.0 basierte, war es notwendig, die Heidelberg Kollektion R1 zu überprüfen und eine neue Kombination aus der BDGP aktualisierten Version 3.1 und der originalen Heidelberg Prediction zu erstellen. Dadurch ergab sich die neue Heidelberg Kollektion R2 mit insgesamt 19.879 Genen, von denen 6.224 der eigenen Vorhersage und 13.050 der gemeinsamen Vorhersage entsprachen. Der Heidelberg FlyArray besteht demnach aus Amplikons für 92,7% der aktuellen BDGP Version 3.1-Gene, wobei einige Heidelberg eigene Genvorhersagen in die neue Version mit aufgenommen wurden. Durch die Eliminierung von Redundanzen, die verursacht durch mehrere Annotationen für ein einzelnes Gen zustande kamen, sank die Gesamtzahl der Gene angepaßt auf die Version 3.1 zwar quantitativ, jedoch wurde aufgrund der optimierten Exondefinitionen [139] eine qualitative Verbesserung der Genvorhersage erreicht. Die korrigierte Zahl der im Verlauf der Entwicklung von *Drosophila* exprimierten Gene sank daher auf 9.908 (78,3%) der BDGP Version 3.1 und 2.636 (42,9%) neu identifizierte Gene. Um erneuten Änderungen der Genbeschreibungen und kommender Versionen der BDGP Genomannotationen begegnen zu können, wurde eine FlyArray-ID eingeführt, die jeden Spot auf dem Heidelberg FlyArray eindeutig beschreibt. Zusätzlich wurde eine variable, anpassbare Gen-ID (HDC) etabliert, welche jedes Gen der gesamten, jeweils aktuellen Heidelberg Prediction, beschreibt. Alle Informationen sind unter der Heidelberg FlyArray Webseite (siehe 2.1.11.) abzufragen.

Durch die Korrespondenzanalyse war es möglich, einen Überblick auf das Transkriptom von *Drosophila melanogaster* während der untersuchten Entwicklungsstadien zu erhalten (siehe 3.4.1.). In Abb. 3.13 zeigt sich die Anordnung der Entwicklungsstadien der 13.812 gefilterten Gene. Die Embryonalstadien zeigen eine geringe Distanz zueinander und gleichzeitig höhere Distanzen gegenüber den Puppenstadien und dem Adultstadium. Dies bedeutet einen wesentlichen Unterschied in der Regulation der globalen Genexpression zwischen den frühen und späten Phasen in der Entwicklung von *Drosophila*. Die Larvenstadien zeigen ein

anderes Verhalten, sie sind zwischen den Embryonal- und Puppenstadien verteilt. Das bedeutet, daß sie sowohl Eigenschaften der Embryonalstadien wie auch der Puppenstadien aufweisen, wie bereits zuvor gezeigt wurde [99]. Der Grund für die verteilten Hybridisierungen (Abb. 3.13, türkise Quadrate) liegt vermutlich im experimentellen Ansatz, da für dieses Stadium die 3 Larvenstadien vereinigt („pool“) wurden und das Larvenstadium den Übergang (Metamorphose) vom embryonalen Zustand zur ersten Puppe darstellt [100]. Alle Entwicklungsstadien von *Drosophila* wurden in Hinblick auf die Regulation der Genexpression eingehender untersucht.

#### **4.2.2 Genexpression in den Entwicklungsstadien bei *Drosophila melanogaster***

Das Embryonalstadium von *Drosophila melanogaster* [140] wurde bereits detailliert untersucht [141,142] und Genexpressionsdaten für dieses Stadium stehen ebenfalls in einer Datenbank zur Verfügung [143]. Aus diesem Grund verwendeten wir das früheste Stadium, 0-4 h des Embryos als Kontrollbedingung in den Hybridisierungen. Von der Fliegenmutter werden bereits während der Oogenese mRNA und Proteine im Ei abgelegt. Da diese Gene von der Mutter exprimiert werden und nicht vom Embryo selbst, werden sie als maternale Gene bezeichnet. Sie sorgen für die erste Musterbildung im frühen Embryo, wobei der gesamte Körperbauplan zuerst als Achsen im Embryo festgelegt wird. Diese beiden Organisationsstrukturen sind die antero-posterio – und die dorso-ventrale Körperachse. Die Gene *nanos* und *oskar*, die beide in diesem Stadium hochreguliert vorliegen (siehe Abb.3.16), sind wesentlich am Aufbau des posterioren Gradienten beteiligt, der die Musterbildung am Hinterende des Eis definiert. Dabei fixiert das Produkt von *oskar* die *nanos*-mRNA am hinteren Pol des Eis, wo nach späterer Translation ein Konzentrationsgradient des *nanos*-Proteins entsteht [144,145]. Das maternale Gen *loki*, eine Serin/Threonin-Proteinkinase [146] ist an der Zellzykluskontrolle beteiligt [147] und wird ebenfalls während der Oogenese exprimiert. Von den vier erwähnten maternalen Genesystemen, die jeweils die Polarität des Eis determinieren, wurden neben Genen für die anterior-posterior Polarität auch Gene für die dorso-ventrale Musterbildung detektiert. Das Gen *Toll* (siehe Abb. 3.18.B), sowie die anderen *Toll*-Gene (*Toll-6*, *Toll-7* und *Tollo*, siehe Abb.3.17) kodieren für Transmembranrezeptorproteine, die in der Vitellinhülle aktiviert durch den *spätzle*-Liganden, den späteren Ventralbereich des Embryos determinieren [148]. Die maternalen Genprodukte legen die primären Körperachsen fest und erzeugen durch die räumliche Verteilung der RNA entlang dieser Achsen Bereiche, in deren Zellkernen die Aktivierung zygotischer Gene stattfindet. Diese werden vom Embryo selbst exprimiert und in verschiedene Klassen eingeordnet. Innerhalb der ersten 6 Stunden nach der Befruchtung und vor der Gastrulation werden die zygotischen Gene aktiv. Ein Beispiel ist das Lückengen

*Krüppel* (siehe Abb.3.18.B), das für ein DNA-bindendes Protein kodiert, an der frühen Einteilung des Embryos entlang der anterior-posterior-Achse beteiligt ist [149] und im Zusammenspiel mit anderen Lückengenen den Rahmen für die nachfolgenden Paarregelgene bildet. Die höchsten Genexpressionsraten von zygotischen Genen in der Embryonalentwicklung zwischen 4 und 12 Stunden zeigen auffälligerweise Gene, die mit der Neurogenese bzw. Axonogenese assoziiert sind. Die Gene *failed axon connections*, *slit* (siehe Abb.3.17.A) und *unc-5* (siehe Abb.3.17.B) spielen eine wichtige Rolle in der Lenkung des Axonwachstums [150,151,152]. Die Gene *dalmatian* und *optic ganglion reduced* sind am Aufbau des Nervensystems beteiligt [153,154]. Einen weiteren wichtigen Mechanismus der Neurogenese zeigen in der frühembryonalen Phase aktivierte Gene in den Abb.3.17.A und 3.18.B. Aus dem Neuroektoderm können sich neuronale Vorläuferzellen durch die Lateralinhibition zu Neuronen entwickeln. Bei der Spezifizierung der Neuronen des Zentralnervensystems und der sensorischen Organe sind die Gene *Notch*, *Delta* und *Scabrous* von zentraler Bedeutung für die Lateralinhibition. Beide Gene kodieren für Transmembranproteine, *Delta* [155] und *Scabrous* [156] für Liganden und *Notch* [157] für einen Rezeptor. Die Aktivierung von *Notch* durch *Delta* führt zur Repression proneuraler Gene, wodurch eine Entwicklung zum Neuroblasten verhindert wird [3]. In den Clustern finden sich weitere hochregulierte Gene, die mit dem *Notch*-Signalweg [158] und der neuronalen Zellspezifizierung direkt verbunden sind (*HLHmgamma*, *scalloped*, *big brain*, *twin of M4* und *bearded*) [159]. In den Embryonalstadien von *Drosophila* werden nicht nur Gengruppen aktiv, die spezifische morphogenetische Entwicklungsprozesse regulieren, sondern es werden auch in einem hohen Maße Gene transkribiert, die zellulär- und biochemisch wichtige Funktionen z. B. in der DNA-Replikation, der Transkription und im Proteinstoffwechsel erfüllen. Das Embryonalstadium ist aufgrund der Genexpression unterschiedlicher Gene und der hohen Anzahl der beteiligten Gene das dynamischste der Entwicklungsstadien. Gene, die für funktionell ähnliche Proteine kodieren, zeigen dabei eine Tendenz, zur gleichen Zeit exprimiert zu werden. In Abb. 3.18.A sind Gene repräsentiert, die mit Chromatin assoziiert sind. Die Gene *brahma* und *Imitation SW1* gehören zu Chromatinmodulierungskomplexen. *brahma* gehört zur *Trithorax*-Gruppe und spielt eine wichtige Rolle in der Genaktivierung [160]. *Imitation SW1* gehört ebenfalls zu einem Proteinkomplex, der durch Konformationsänderungen des Chromatins die DNA für Transkriptionsfaktoren zugänglich macht [161]. *Histon H3.3* ist im Nukleosom lokalisiert und spielt vermutlich eine wichtige Rolle in der chromatinregulierten Genaktivierung [162]. *Lamin* ist ein Strukturprotein der Zellkernlamina und verbindet den Zellkern mit Chromatinstrukturen [163]. Die Gene *Minichromosome maintenance 3*, *Minichromosome maintenance 6* [164], *spt4* und *spt6* sind ebenfalls an Konformationsänderungen des Chromatins beteiligt und sind in die Einleitung der DNA-Replikation involviert. Durch Veränderungen der Chromatinstruktur erhalten Transkriptionsfaktoren Zugang zu Enhancer- oder Promoterregionen der

betreffenden Gene; ermöglichen deren Expression und letztendlich erfolgt die Translation in das Protein. Die Reihenfolge dieses regulatorischen Prinzips läßt sich auch im Verlauf der Transkriptionsprofile in den Clustern nachvollziehen. Unmittelbar mit der Expression von Genen, die mit dem Chromatin assoziiert sind, zeigen sich in Abb. 3.18.D Gene, die an der Regulation der Transkription beteiligt sind. Die Trennung der Genfunktionen aufgrund der GO-Terminologie ist manchmal nicht möglich, da einem Gen jeweils mehrere GO-IDs (also Funktionen) zugeordnet sind (siehe 2.2.13.) und Gene auch in der Realität unterschiedliche Funktionen aufweisen. Ein Beispiel für die enge Verbindung zwischen Chromatinregulation und Transkriptionsregulation ist das Gen *enhancer of yellow 1*, das in beiden Bereichen eine wichtige Rolle spielt, da es sowohl eine Histondomäne besitzt [165] und zugleich ein Transkriptionsfaktor des TFIID-Komplexes ist. Generelle TBP-assoziierte Faktoren (TAFs) mit RNA-Polymerase II Transkriptionsfaktoraktivität sind ebenfalls vorhanden (*Taf110*, *Taf30alpha* und *Taf60*) [166]. Weitere mehr spezifische RNA-Polymerase II Transkriptionsfaktoren in diesem Cluster sind *dauhterless*, *modifier of mdg4*, *scute* und *Adh transcription factor 1*. Die gemeinsame Expression von Transkriptionsfaktoren wie in diesem Cluster ist sinnvoll, da die kodierten Proteine meistens gemeinsame Komponenten (auch in variabler Zusammensetzung) in Multiproteinkomplexen (z.B. TFIID) bilden [167]. Der Aufbau von Transportstrukturen für die Fertigung von Proteinen in Zellorganellen wie dem Golgi-Apparat läßt sich in Abb. 3.18.C nachvollziehen. Dem Profilverlauf des Clusters nach erfolgt die Expression der daran beteiligten Gene wie *CG2903*, *CG6359*, *CG8228*, *Kdel Receptor*, *Syntaxin 5* und *Tim10* sehr früh und ausschließlich in den ersten 4 Stunden der Entwicklung, da die Organellstrukturen vermutlich für die später folgende Proteinbiosynthese notwendig sind. Für *Syntaxin 5* wurde gezeigt, daß es essentiell ist für die Biogenese des Golgi-Apparates [168]. Der Beginn der Proteinbiosynthese ist gekennzeichnet durch die Transkription ribosomaler Gene (siehe Abb. 3.18.F). Die Gene repräsentieren kleine und große Untereinheiten der Ribosomen, deren Expression [169] in den ersten 4 bis 16 Stunden der Embryonalstadien erfolgt und beim Übergang zum Larvenstadium stark absinkt. Eine gemeinsame Expression von Genen mit ähnlicher Funktion zu diesem frühen Zeitpunkt zeigt auch Abb. 3.18.E, in der proteolytisch aktive Gene (z.B. *Dox-A2*, *Rpn11*, *Rpn2*, *Rpt3* und *Rpt5*) bzw. Gene des Proteasome 26-Komplexes (*Proteasome 26*, *Proteasome 54*) [170] exprimiert werden. Die Expression ist in den ersten acht Stunden der Embryogenese hoch, da wahrscheinlich in diesem Zeitraum massive morphologische und strukturelle Veränderungen mit der Degradation von Proteinen einhergehen. Vor dem Adultstadium steigt die Expression proteolytischer Gene ebenfalls an, vermutlich um puppenspezifische Strukturen abzubauen. Im letzten Embryonalstadium (siehe Abb. 3.17.C) werden vermehrt insektenspezifische Gene wie z.B. *aristaless*, *a methyl dopa resistant*, *CG8541* und *Larval visceral protein L* exprimiert. Gene wie *a methyl dopa resistant* und *CG8541* sind wesentlich an der Hautbildung der Larve beteiligt [171]. Die

erhöhte Anzahl von larvenspezifischen Genen am Ende der Embryogenese weist auf den Übergang vom Embryonal- zum Larvenstadium hin.

Das Larvenstadium von *Drosophila* gliedert sich in drei Stadien und wird charakterisiert durch entwicklungsspezifische Veränderungen in der gesamten Morphologie, wie auch durch Zellwachstum, Gewebeumbildung und Zellwanderungen. Dieser als Metamorphose bezeichnete Vorgang, der am Ende des dritten Larvenstadiums (Instar) beginnt und vor der Verpuppung endet, wird durch das Hormon 20-Hydroxyecdysol induziert [172]. Larvenspezifisches Gewebe, wie der Darm, die Speicheldrüsen und spezielle Muskeln reagieren auf das Ecdysol mit programmiertem Zelltod und nachfolgender Histolyse. Die Imaginalscheiben antworten auf den Ecdysoneinfluss mit der Restrukturierung und Differenzierung zu rudimentären Flügeln, Beinen, Augen oder Antennen [173]. Diese Umformung der Gewebe, verursacht durch eine Ecdysol-induzierte Transkriptionskaskade, wird begleitet von großen Veränderungen der gewebespezifischen Genexpression [174]. Die genetische Antwort auf das Ecdysolhormon ließ sich auch in dieser Arbeit nachvollziehen. In den Abbildungen 3.19.A und 3.19.C sind die Expressionsprofile von *Ecdysone-induced protein 28/29* und *Ecdysone-induced gene 71* dargestellt [175], die eine Aktivierung dieser Gene während des Larvenstadiums zeigen. Die Metamorphose setzt in dem letzten von drei Larvenstadien ein, und viele Gene werden zu diesem Zeitpunkt angeschaltet. Da alle drei Larvenstadien für die Genexpressionsanalysen vereinigt wurden, wird es daher als ein einzelnes Stadium repräsentiert. Dementsprechend zeigt sich in Abb. 3.19.C eine starke Expression durch die Metamorphose beeinflusster Gene, wie *Larval cuticle protein* (Hautbildung der Larve) [176] und *Salivary gland secretion*, die an der Sekretion der Speicheldrüsen beteiligt sind [177]. Die Gene *Larval serum protein* und *fat body protein 1* (siehe Abb. 3.19.D) kodieren für einen extrazellulären Proteinkomplex, die sogenannten Fettkörper, die als Reservoir für Aminosäuren zur Proteinherstellung für die adulte Fliege dienen [178]. Die erwähnte Histolyse von larvenspezifischem Gewebe, die durch die Metamorphose eingeleitet wird, äußert sich in der Hochregulierung einer großen Anzahl von spaltenden Enzymen wie Proteasen (z.B. *Serine protease 1*), Trypsinen (z.B. *alpha Trypsin*, *zeta Trypsin*) und Lysozymen (z.B. *Lysozym A*) (siehe Abb. 3.19.A und 3.19.B) [148,179]. Durch erneute Ecdysolinduktionen zu Beginn und im weiteren Verlauf des Puppenstadiums wird die Metamorphose koordiniert. In Abb. 3.20.A wird die Aktivierung der darauf reagierenden *Ecdysone-induced gene 71*-Gene durch den steilen Anstieg des Transkriptniveaus zu Beginn des ersten Puppenstadiums gekennzeichnet. Auffällig ist, über alle drei Puppenstadien verteilt, die Expression von Genen, die in den Kohlehydratstoffwechsel (z.B. *Aldolase*, *Malic enzyme* und *Malat dehydrogenases*) involviert sind. Im Zusammenhang damit zeigt die erhöhte Expression von Enzymgenen, die am Transport von Metabolismuskomponenten und Oxidation von Zuckern beteiligt sind (*Glutathione S-transferases*, siehe Abb.3.20.C, *Aconitase*, *Succinyl-coenzyme A*, siehe

Abb.3.20.D), ebenfalls den gesteigerten Energiebedarf. Gene des Energiestoffwechsels, die Reduktionsäquivalente in Form von ATP oder NADH bereitstellen (z.B. *Na pump alpha subunit*, *Oscp*, *CG3683* und *CG6020*, siehe Abb.3.20.E), sind gleichzeitig hochreguliert. Gene der Atmungskette (z.B. *Cytochrome P450 oxygen binding*, siehe Abb.3.20.B, *Cytochrome c oxidase*, *ATP synthases*, siehe Abb.3.20.E und *oxen*, siehe Abb.3.20.D), die im Mitochondrium lokalisiert ist, zeigen ebenfalls eine hohe Transkriptionshäufigkeit [180]. Dieser Anstieg von Komponenten des Energiestoffwechsels ist vermutlich eine Voraussetzung der Metamorphose zur adulten Fliege, da in diesen Stadien die endgültigen Organe und Gliedmaßen ausgebildet werden müssen. Es finden sich daher auch entwicklungspezifische Gene, wie z.B. *bride of sevenless*, *Jun-related antigen*, *bendless* und *chaoptic*, siehe Abb.3.20.C die für die Entwicklung der Photorezeptorzellen des Auges [181] wichtig sind. Interessanterweise fanden sich im dritten Puppenstadium Cluster, die eine hohe Zahl (87%, siehe 3.4.3.) der unbekanntenen Gene aus der Heidelberg Vorhersage (siehe 4.2.1.) aufwiesen. Wie in Abb.3.20.F zu sehen ist, sind diese Gene spezifisch im letzten Puppenstadium gleichmäßig hochreguliert. Es zeigten noch zwei weitere Cluster mit ähnlichem Profil eine hohe Zahl (70 – 80%) unbekannter Gene auf. In allen anderen Clustern lag der Anteil der unbekanntenen Gene jeweils zwischen 5 – 30 %. Das läßt darauf schließen, daß ein Teil der bisher unbekanntenen Gene spezifisch im dritten Puppenstadium ko-exprimiert wird. Möglich ist, daß die unbekanntenen Gene ebenfalls in den Energiestoffwechsel involviert sind oder mitochondrialen Ursprungs sind. Eine exakte Aussage ließe sich jedoch nur anhand weiterer funktioneller Untersuchungen treffen. Das Adultstadium von *Drosophila* weist ebenfalls eine hohe Expression von Genen auf, die mit dem Cytochrom-Komplex der Atmungskette assoziiert sind (siehe Abb. 3.21.B). Auffällig an diesem Stadium ist die Expression von Genen, die an der Sexualdifferenzierung von *Drosophila* beteiligt sind. Hauptsächlich handelt es sich dabei um die *Accessory gland peptides* (siehe Abb. 3.21.A – C) [182], sowie *Male-specific RNA 57 Dc* und *Yolk protein 3* (siehe Abb. 3.21.C). Gene wie das *Yolk protein 3* bewirken eine weitere Zelldifferenzierung genitaler Vorläuferzellen und Spezifizierung der Geschlechter [183] in diesem jungen Stadium der erwachsenen Fliege.

Die bisher dargestellten Cluster wiesen alle signifikant hohe Transkriptionsraten spezifisch für das jeweilige Entwicklungsstadium auf. Abhängig von der Funktion bildeten sich auch Cluster mit stark differentiellen und konstitutiv regulierten Genen. Abb. 3.22.A zeigt ein Cluster mit differentiellen Transkriptionsprofilen. Die Varianz der Genexpressionsprofile innerhalb des Clusters betrug 13,52. Viele Gene sind an der Muskelbildung beteiligt, wie z.B. *Myosin heavy chain*, *Muscle protein 20*, *Muscle specific protein 300*, *Paramyosin* und *Tropomyosin 2* und spielen eine wichtige Rolle in der Mesodermentwicklung [184]. Die Expression der Gene steigt gegen Ende des Embryonalstadiums stark an und schwankt dann auf hohem Niveau zwischen den späteren Stadien. Diese Beobachtung stimmt gut mit

vorherigen Genexpressionsstudien muskelspezifischer Gene überein [99]. Das Cluster ist außerdem ein Beweis, daß mit Hilfe von Microarrays auch gewebespezifische Genexpression nachgewiesen werden kann. Dagegen ist Abb. 3.22.B ein Beispiel für konstitutiv exprimierte Gene mit geringer differentieller Regulation, denn die Varianz der Genexpressionsprofile innerhalb des Clusters beträgt in diesem Fall nur 0,89. In diesem Cluster finden sich Gene, die für Signalrezeptorproteine wie *Calcium-dependent protein kinase 1*, *CG1976*, *CG4394* und *Calmodulin-binding protein related to Rab3* kodieren und die an der Kalzium-abhängigen Signaltransduktion beteiligt sind. Die Expression der Gene auf niedrigem Niveau ist vielleicht damit zu erklären, daß sie die Funktion von „Haushaltsgenen“ besitzen, da Kalzium- bzw. Calmodulin-abhängige Stoffwechselwege in vielen unterschiedlichen Bereichen (z.B. Signaltransduktion, Cytoskelett, Muskelaufbau, Reizleitung in Nerven) notwendig sind [185]. Die meisten dargestellten Cluster zeigten eine Amplitude in einem der untersuchten Stadien. Abb.3.22.C zeigt ein Beispiel für zwei Amplituden exprimierter Gene, die ihr Optimum jeweils im letzten Embryonalstadium und im ersten Puppenstadium besitzen. Die Gene gehören zu den *Ecdysone-inducible genes* und *Chitinases*, die wie oben diskutiert, eine essentielle Rolle während der Metamorphose spielen. Die Metamorphose setzt erst im dritten Larvenstadium ein, aber Gene wie die *Chitinasen* besitzen einen bivalenten Charakter, denn sie sind sowohl anfangs am Chitinmetabolismus, als auch gegen Ende des Larvenstadiums am Chitinkatabolismus beteiligt [186].

### 4.3 Ausblick

Die durchgeführten Genexpressionsanalysen zeigten, daß in jedem Entwicklungsstadium von *Drosophila* die Regulation bestimmter Gengruppen erfolgt. Die Expression auf molekularer Ebene stimmt mit den morphologischen Veränderungen zwischen den Entwicklungsstadien überein. Viele Gene, die eine biochemisch- oder physiologisch ähnliche Funktion erfüllen, werden koordiniert zur gleichen Zeit exprimiert.

Die vorgestellten Ergebnisse zeigen auf eindrucksvolle Weise die Möglichkeit, einen gesamten Organismus im Verlaufe seiner Entwicklung mit Hilfe Microarray-basierender Genexpressionsanalysen zu beobachten und die molekulargenetische Funktionsweisen besser zu verstehen. Durch den Einsatz von computergestützten *in silico*-Berechnungen von potentiell proteinkodierenden Genen in Kombination mit Microarrayexperimenten wurden mindestens 2.000 neue Gene von *Drosophila melanogaster* identifiziert und beweist die Möglichkeit, nahezu alle Gene eines Genoms zu entdecken. Dies bedeutet gleichzeitig, daß reine Computervorhersagen möglicher Gene (noch) nicht ausreichen, das komplette Genom eines Organismus zu ermitteln. Die Festlegung auf eine potentielle Genzahl sollte deshalb mit Vorsicht erfolgen. Dieses Prinzip läßt sich auch auf andere Organismen wie Mensch

oder Maus übertragen. Aufgrund der Resultate dieser Arbeit muss die Zahl der Gene von *Drosophila* von ca. 14.000 [1] auf mindestens 16.000 erhöht werden. Wegen der geringen Übereinstimmung der neuen Gene mit Genen der Genomsequenz von *Drosophila pseudoobscura* [37] handelt es sich eventuell um bisher unbekannte Proteinfamilien der Gattung *Drosophila melanogaster*.

Der für die Microarraykonstruktion verwendete Satz von PCR-Amplifikaten könnte ebenfalls für die Herstellung von doppelsträngigen RNA-Proben für genomweite RNAi-Studien eingesetzt werden. Damit ergibt sich die Möglichkeit, den gleichen Satz von Genen für genomweite RNAi-Experimente und anschließende Genexpressionsanalysen mit Microarrays effizient einzusetzen. Diese Kombination molekularbiologischer Instrumente eignet sich daher optimal für die funktionelle Analyse neu entdeckter Gene und weitere Studien genetischer Netzwerke.



## 5 Danksagung

Diese Arbeit wurde vom Juni 2000 bis Dezember 2003 unter der Leitung von Herrn Prof. Dr. Frank Sauer, Department of Biochemistry, University of California Riverside, USA und am Deutschen Krebsforschungszentrum Heidelberg in der Abteilung Funktionelle Genomanalyse unter der Leitung von Herrn Dr. Jörg. D. Hoheisel durchgeführt.

Mein besonderer Dank gilt:

Herrn Prof. Dr. Frank Sauer für die engagierte Betreuung meiner Promotionsarbeit.

Herrn Dr. Jörg D. Hoheisel für die Möglichkeit, diese Dissertation in seiner Arbeitsgruppe anzufertigen, seine stete Diskussionsbereitschaft und vielen Lösungsvorschläge bei Fragen, die im Laufe der Arbeit auftraten. Ebenso danke ich ihm für die Ermöglichung von Kongressbesuchen im In- und Ausland.

Jessica Werdermann für ihre engagierte und verlässliche Mitarbeit bei der Durchführung der Versuche und Unterstützung bei der Etablierung neuer Methoden zu den Genexpressionsanalysen.

Verena Beier und Andrea Bauer für die hilfreiche Unterstützung und Diskussionen rund um das Thema „Transcriptional profiling“. Kurt Fellenberg und Christian Busold für die Hilfe bei M-CHiPS.

Ole Brandt für die unermüdliche Hilfsbereitschaft bei Computerproblemen und tiefgehenden Diskussionen über Linux und Systemadministration im Allgemeinen.

Verena Beier und Ole Brandt für die Hilfe bei der Durchsicht dieser Dissertation.

Allen Mitarbeitern der Abteilung Funktionelle Genomanalyse für die angenehme und entspannte Arbeitsatmosphäre.

Meiner Freundin Kathrin für ihre ausdauernde Geduld und für ihr Verständnis während der Durchführung dieser Arbeit. Unserem Sohn Carl für den „Beschleunigungseffekt“ zum Abschluß der Dissertation, sowie meiner und Kathrins Familie für die Unterstützung während der Arbeit und darüber hinaus.

## 6 Eigene Publikationen

Hild, M., Beckmann, B., Haas, S. A., Koch, B., Solovyev, V., Busold, C., Fellenberg, K., Boutros, M., Vingron, M., Sauer, F., Hoheisel, J. D. and Paro, R. (2003) An integrated Gene Annotation and Transcriptional Profiling Approach towards the full Gene Content of the Drosophila Genome. *Genome Biology*, in Press.

Bauer, A., Beckmann, B., Busold, C., Brandt, O., Kusnezow, W., Pullat, J., Aign, V., Fellenberg, K., Fleischer, R., Jacob, A., Frohme, M., Hoheisel, J.D. (2003) Use of complex DNA and antibody microarrays as tools in functional analyses. *Comp Funct Genom.*,**4**: 520-524.

Diehl, F., Beckmann, B., Kellner, N., Hauser, N., Diehl, S. and Hoheisel, J. (2002 Aug 15) Manufacturing DNA microarrays from unpurified PCR products. *Nucleic Acids Res*, **30**(16), e79.

---

## Literaturverzeichnis

1. Adams, M., Celniker, S., Holt, R., Evans, C., Gocayne, J., Amanatides, P., Scherer, S., Li, P., Hoskins, R., Galle, R., George, R., Lewis, S., Richards, S., Ashburner, M., Henderson, S., Sutton, G., Wortman, J., Yandell, M., Zhang, Q., Chen, L., Brandon, R., Rogers, Y., Blazej, R., Champe, M., Pfeiffer, B., Wan, K., Doyle, C., Baxter, E., Helt, G., Nelson, C., Gabor, G., Abril, J., Agbayani, A., An, H., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E., Beeson, K., Benos, P., Berman, B., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M., Bouck, J., Brokstein, P., Brottier, P., Burtis, K., Busam, D., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J., Cawley, S., Dahlke, C., Davenport, L., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A., Dew, I., Dietz, S., Dodson, K., Doup, L., Downes, M., Dugan-Rocha, S., Dunkov, B., Dunn, P., Durbin, K., Evangelista, C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A., Garg, N., Gelbart, W., Glasser, K., Glodek, A., Gong, F., Gorrell, J., Gu, Z., Guan, P., Harris, M., Harris, N., Harvey, D., Heiman, T., Hernandez, J., Houck, J., Hostin, D., Houston, K., Howland, T., Wei, M., Ibegwam, C., et al. (2000 Mar 24) The genome sequence of *Drosophila melanogaster*. *Science*(5461), 2185 - 95.
2. Celniker, S., Wheeler, D., Kronmiller, B., Carlson, J., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S., Frise, E., Hodgson, A., George, R., Hoskins, R., Lavery, T., Muzny, D., Nelson, C., Pacleb, J., Park, S., Pfeiffer, B., Richards, S., Sodergren, E., Svirskas, R., Tabor, P., Wan, K., Stapleton, M., Sutton, G., Venter, C., Weinstock, G., Scherer, S., Myers, E., Gibbs, R. and Rubin, G. (2002) Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol*, **3**(12), RESEARCH0079.
3. Wolpert, L. (1999) Entwicklungsbiologie, *Spektrum*, Berlin.
4. Lawrence, P. A. (1992) The Making of a Fly, *Blackwell Scientific Publications*, Oxford.
5. Granjeaud, S., Bertucci, F. and Jordan, B. (1999 Sep) Expression profiling: DNA arrays in many guises. *Bioessays*, **21**(9), 781 - 90.
6. DeRisi, J. and Iyer, V. (1999 Jan) Genomics and array technology. *Curr Opin Oncol*, **11**(1), 76 - 9.
7. (1999) The Chipping Forecast. *Nature Genet.*, **21**(suppl.), 1-60.
8. Kane, M., Jatkoe, T., Stumpf, C., Lu, J., Thomas, J. and Madore, S. (2000 Nov 15) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res*, **28**(22), 4552 - 7.
9. Fodor, S., Read, J., Pirrung, M., Stryer, L., Lu, A. and Solas, D. (1991 Feb 15) Light-directed, spatially addressable parallel chemical synthesis. *Science*, **251**(4995), 767 - 73.
10. Maskos, U. and Southern, E. (1992 Apr 11) Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. *Nucleic Acids Res*, **20**(7), 1679 - 84.
11. Hieter, P. and Boguski, M. (1997 Oct 24) Functional genomics: it's all how you read it. *Science*, **278** (5338), 601 - 2.
12. Brown, P. and Botstein, D. (1999 Jan) Exploring the new world of the genome with DNA microarrays. *Nat Genet*, **21**(1 Suppl), 33 - 7.
13. Lockhart, D. and Winzeler, E. (2000 Jun 15) Genomics, gene expression and DNA arrays. *Nature*, **405**(6788), 827 - 36.
14. Freeman, W., Robertson, D. and Vrana, K. (2000 Nov) Fundamentals of DNA hybridization arrays

- 
- for gene expression analysis. *Biotechniques*, **29**(5), 1042 - 6, 1048-55.
15. Schadt, E., Li, C., Su, C. and Wong, W. (2000 Oct 20) Analyzing high-density oligonucleotide gene expression array data. *J Cell Biochem*, **80**(2), 192 - 202.
  16. Tseng, G., Oh, M., Rohlin, L., Liao, J. and Wong, W. (2001 Jun 15) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*, **29**(12), 2549 - 57.
  17. Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H. and Herzog, H. (2000 May 15) Normalization strategies for cDNA microarrays. *Nucleic Acids Res*, **28**(10), E47.
  18. Lee, M., Kuo, F., Whitmore, G. and Sklar, J. (2000 Aug 29) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A*, **97**(18), 9834 - 9.
  19. DeRisi, J., Penland, L., Brown, P., Bittner, M., Meltzer, P., Ray, M., Chen, Y., Su, Y. and Trent, J. (1996 Dec) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet*, **14**(4), 457 - 60.
  20. Chen, Y., Kamat, V., Dougherty, E., Bittner, M., Meltzer, P. and Trent, J. (2002 Sep) Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics*, **18**(9), 1207 - 15.
  21. Fellenberg, K., Hauser, N., Brors, B., Hoheisel, J. and Vingron, M. (2002 Mar) Microarray data warehouse allowing for inclusion of experiment annotations in statistical analysis. *Bioinformatics*, **18**(3), 423 - 33.
  22. Greenacre, M. J. (1984) Theory and application of Correspondence Analysis, 1 Ed., *Academic Press*, London.
  23. Eisen, M., Spellman, P., Brown, P. and Botstein, D. (1998 Dec 8) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, **95**(25), 14863 - 8.
  24. Tavazoie, S., Hughes, J., Campbell, M., Cho, R. and Church, G. (1999 Jul) Systematic determination of genetic network architecture. *Nat Genet*, **22**(3), 281 - 5.
  25. Kohonen, T. (1995) Self Organizing Maps, *Springer*, Berlin.
  26. Raychaudhuri, S., Stuart, J. and Altman, R. (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput*, 455 - 66.
  27. Karlin, S., Bergman, A. and Gentles, A. (2001 May 17) Genomics. Annotation of the *Drosophila* genome. *Nature*, **411**(6835), 259 - 60.
  28. Gopal, S., Schroeder, M., Pieper, U., Sczyrba, A., Aytikin-Kurban, G., Bekiranov, S., Fajardo, J., Eswar, N., Sanchez, R., Sali, A. and Gaasterland, T. (2001 Mar) Homology-based annotation yields 1,042 new candidate genes in the *Drosophila melanogaster* genome. *Nat Genet*, **27**(3), 337 - 40.
  29. Andrews, J., Bouffard, G., Cheadle, C., Lü, J., Becker, K. and Oliver, B. (2000 Dec) Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res*, **10**(12), 2030 - 43.
  30. Posey, K., Jones, L., Cerda, R., Bajaj, M., Huynh, T., Hardin, P. and Hardin, S. (2001) Survey of transcripts in the adult *Drosophila* brain. *Genome Biol*, **2**(3), RESEARCH0008.
-

- 
31. Reese, M., Kulp, D., Tammanna, H. and Haussler, D. (2000 Apr) Gene--gene finding in *Drosophila melanogaster*. *Genome Res*, **10**(4), 529 - 38.
  32. Burge, C. and Karlin, S. (1997 Apr 25) Prediction of complete gene structures in human genomic DNA. *J Mol Biol*, **268**(1), 78 - 94.
  33. Holt, R., Subramanian, G., Halpern, A., Sutton, G., Charlab, R., Nusskern, D., Wincker, P., Clark, A., Ribeiro, J., Wides, R., Salzberg, S., Loftus, B., Yandell, M., Majoros, W., Rusch, D., Lai, Z., Kraft, C., Abril, J., Anthouard, V., Arensburger, P., Atkinson, P., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G., Chrystal, M., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M., Hladun, S., Hogan, J., Hong, Y., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J., Lobo, N., Lopez, J., Malek, J., McIntosh, T., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S., O'Brochta, D., Pfannkoch, C., Qi, R., Regier, M., Remington, K., Shao, H., Sharakhova, M., Sitter, C., Shetty, J., Smith, T., Strong, R., Sun, J., Thomasova, D., Ton, L., Topalis, P., Tu, Z., Unger, M., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K., Wortman, J., Wu, M., Yao, A., Zdobnov, E., Zhang, H., Zhao, Q., et al. (2002 Oct 4) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, **298**(5591), 129 - 49.
  34. Bergman, C., Pfeiffer, B., Rincón-Limas, D., Hoskins, R., Gnirke, A., Mungall, C., Wang, A., Kronmiller, B., Pacleb, J., Park, S., Stapleton, M., Wan, K., George, R., de Jong, P., Botas, J., Rubin, G. and Celniker, S. (2002) Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome. *Genome Biol*, **3**(12), RESEARCH0086.
  35. Rubin, G., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M. and Harvey, D. (2000 Mar 24) A *Drosophila* complementary DNA resource. *Science*, **287**(5461), 2222 - 4.
  36. Salamov, A. and Solovyev, V. (2000 Apr) Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res*, **10**(4), 516 - 22.
  37. Hild, M., Beckmann, B., Haas, S. A., Koch, B., Solovyev, V., Busold, C., Fellenberg, K., Boutros, M., Vingron, M., Sauer, F., Hoheisel, J. D. and Paro, R. (2003) An integrated Gene Annotation and Transcriptional Profiling Approach towards the full Gene Content of the *Drosophila* Genome. *Genome Biology*, in Press.
  38. Haas, S., Hild, M., Wright, A., Hain, T., Talibi, D. and Vingron, M. (2003 Oct 1) Genome-scale design of PCR primers and long oligomers for DNA microarrays. *Nucleic Acids Res*, **31**(19), 5576 - 81.
  39. Mullis, K. and Faloona, F. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*, **155**, 335 - 50.
  40. Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. and Erlich, H. (1988 Jan 29) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**(4839), 487 - 91.
  41. Yang, I., Chen, E., Hasseman, J., Liang, W., Frank, B., Wang, S., Sharov, V., Saeed, A., White, J., Li, J., Lee, N., Yeatman, T. and Quackenbush, J. (2002 Oct 24) Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol*, **3**(11), research0062.
  42. Kerr, M., Martin, M. and Churchill, G. (2000) Analysis of variance for gene expression microarray data. *J Comput Biol*, **7**(6), 819 - 37.
-

- 
43. Cui, X. and Churchill, G. A. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol*, **4**(4), 210.
  44. Dysvik, B. and Jonassen, I. (2001 Apr) J-Express: exploring gene expression data using Java. *Bioinformatics*, **17**(4), 369 - 70.
  45. Knudsen, S. (2002) A Biologist's Guide to Analysis of DNA Microarray Data, *John Wiley & Sons*, New York.
  46. Azuaje, F. (2003 Mar) Clustering-based approaches to discovering and visualising microarray data patterns. *Brief Bioinform*, **4**(1), 31 - 42.
  47. Dougherty, E., Barrera, J., Brun, M., Kim, S., Cesar, R., Chen, Y., Bittner, M. and Trent, J. (2002) Inference from clustering with application to gene-expression microarrays. *J Comput Biol*, **9**(1), 105 - 26.
  48. Nikkilä, J., Törönen, P., Kaski, S., Venna, J., Castrén, E. and Wong, G. (2002 Oct-Nov) Analysis and visualization of gene expression data using self-organizing maps. *Neural Netw*, **15**(8-9), 953 - 66.
  49. Törönen, P., Kolehmainen, M., Wong, G. and Castrén, E. (1999 May 21) Analysis of gene expression data using self-organizing maps. *FEBS Lett*, **451**(2), 142 - 6.
  50. Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J., Richardson, J., Ringwald, M., Rubin, G. and Sherlock, G. (2000 May) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1), 25 - 9.
  51. Lewis, S., Ashburner, M. and Reese, M. (2000 Jun) Annotating eukaryote genomes. *Curr Opin Struct Biol*, **10**(3), 349 - 54.
  52. Diehl, F., Grahlmann, S., Beier, M. and Hoheisel, J. (2001 Apr 1) Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. *Nucleic Acids Res*, **29**(7), E38.
  53. Diehl, F., Beckmann, B., Kellner, N., Hauser, N., Diehl, S. and Hoheisel, J. (2002 Aug 15) Manufacturing DNA microarrays from unpurified PCR products. *Nucleic Acids Res*, **30**(16), e79.
  54. Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. (2001 Jan 1) Quantitative assessment of DNA microarrays--comparison with Northern blot analyses. *Genomics*, **71**(1), 34 - 9.
  55. Beissbarth, T., Fellenberg, K., Brors, B., Arribas-Prat, R., Boer, J., Hauser, N., Scheideler, M., Hoheisel, J., Schütz, G., Poustka, A. and Vingron, M. (2000 Nov) Processing and quality control of DNA array hybridization data. *Bioinformatics*, **16**(11), 1014 - 22.
  56. Fellenberg, K., Hauser, N., Brors, B., Neutzner, A., Hoheisel, J. and Vingron, M. (2001 Sep 11) Correspondence analysis applied to microarray data. *Proc Natl Acad Sci U S A*, **98**(19), 10781 - 6.
  57. Jansen, R., Greenbaum, D. and Gerstein, M. (2002 Jan) Relating whole-genome expression data with protein-protein interactions. *Genome Res*, **12**(1), 37 - 46.
  58. Qin, Z., McCue, L., Thompson, W., Mayerhofer, L., Lawrence, C. and Liu, J. (2003 Apr) Identification of co-regulated genes through Bayesian clustering of predicted regulatory binding sites. *Nat Biotechnol*, **21**(4), 435 - 9.
  59. Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. and
-

- Golub, T. (1999 Mar 16) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A*, **96**(6), 2907 - 12.
60. Spellman, P. and Rubin, G. (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol*, **1**(1), 5.
61. Stathopoulos, A. and Levine, M. (2002) Whole-genome expression profiles identify gene batteries in *Drosophila*. *Dev Cell*, **3**(4), 464-5.
62. Quackenbush, J. (2001 Jun) Computational analysis of microarray data. *Nat Rev Genet*, **2**(6), 418 - 27.
63. Schena, M., Shalon, D., Davis, R. and Brown, P. (1995 Oct 20) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**(5235), 467 - 70.
64. Wrobel, G., Schlingemann, J., Hummerich, L., Kramer, H., Lichter, P. and Hahn, M. (2003 Jun 15) Optimization of high-density cDNA-microarray protocols by 'design of experiments'. *Nucleic Acids Res*, **31**(12), e67.
65. Rickman, D., Herbert, C. and Aggerbeck, L. (2003 Sep 15) Optimizing spotting solutions for increased reproducibility of cDNA microarrays. *Nucleic Acids Res*, **31**(18), e109.
66. McQuain, M., Seale, K., Peek, J., Levy, S. and Haselton, F. (2003 Sep 15) Effects of relative humidity and buffer additives on the contact printing of microarrays by quill pins. *Anal Biochem*, **320**(2), 281 - 91.
67. Zammateo, N., Jeanmart, L., Hamels, S., Courtois, S., Louette, P., Hevesi, L. and Remacle, J. (2000 Apr 10) Comparison between different strategies of covalent attachment of DNA to glass surfaces to build DNA microarrays. *Anal Biochem*, **280**(1), 143 - 50.
68. Chiu, S., Hsu, M., Ku, W., Tu, C., Tseng, Y., Lau, W., Yan, R., Ma, J. and Tzeng, C. (2003 Sep 15) Synergistic effects of epoxy- and amine-silanes on microarray DNA immobilization and hybridization. *Biochem J*, **374**(Pt 3), 625 - 32.
69. Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P. and Jordan, B. (1995 Sep 1) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics*, **29**(1), 207 - 16.
70. Yue, H., Eastman, P., Wang, B., Minor, J., Doctolero, M., Nuttall, R., Stack, R., Becker, J., Montgomery, J., Vainer, M. and Johnston, R. (2001 Apr 15) An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. *Nucleic Acids Res*, **29**(8), E41 - 1.
71. Wetmur, J. (1991) DNA probes: applications of the principles of nucleic acid hybridization. *Crit Rev Biochem Mol Biol*, **26**(3-4), 227 - 59.
72. Santoro, M., Liu, Y., Khan, S., Hou, L. and Bolen, D. (1992 Jun 16) Increased thermal stability of proteins in the presence of naturally occurring osmolytes. *Biochemistry*, **31**(23), 5278 - 83.
73. Rees, W., Yager, T., Korte, J. and von Hippel, P. (1993 Jan 12) Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry*, **32**(1), 137 - 44.
74. Henke, W., Herdel, K., Jung, K., Schnorr, D. and Loening, S. (1997 Oct 1) Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res*, **25**(19), 3957 - 8.

- 
75. Chakrabarti, R. and Schutt, C. (2001 Jun 1) The enhancement of PCR amplification by low molecular weight amides. *Nucleic Acids Res*, **29**(11), 2377 - 81.
76. Henke, W. and Loening, S. (1998 Jan 15) Recently, betaine has been introduced as an additive in different PCR strategies. *Nucleic Acids Res*, **26**(2), 687.
77. Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J., Snesrud, E., Lee, N. and Quackenbush, J. (2000 Sep) A concise guide to cDNA microarray analysis. *Biotechniques*, **29**(3), 548 - 50, 552-4, 556 passim.
78. Lemeshko, S., Powdrill, T., Belosludtsev, Y. and Hogan, M. (2001 Jul 15) Oligonucleotides form a duplex with non-helical properties on a positively charged surface. *Nucleic Acids Res*, **29**(14), 3051 - 8.
79. Peterson, A., Heaton, R. and Georgiadis, R. (2001 Dec 15) The effect of surface probe density on DNA hybridization. *Nucleic Acids Res*, **29**(24), 5163 - 8.
80. Shi, S., Scheffer, A., Bjeldanes, E., Reynolds, M. and Arnold, L. (2001 Oct 15) DNA exhibits multi-stranded binding recognition on glass microarrays. *Nucleic Acids Res*, **29**(20), 4251 - 6.
81. Nakano, S., Fujimoto, M., Hara, H. and Sugimoto, N. (1999 Jul 15) Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res*, **27**(14), 2957 - 65.
82. Schena, M. (2000) Microarray Biochip Technology, *Eaton Publishing*, Natick.
83. Iacobas, A., Urban, M. and Spray, D. (2000 Jan-Dec) New protocol in spotting microarray technique. *Rom J Physiol*, **37**(1-4), 69 - 80.
84. Benes, V. and Muckenthaler, M. (2003 May) Standardization of protocols in cDNA microarray analysis. *Trends Biochem Sci*, **28**(5), 244 - 9.
85. Heller, M. (2002) DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng*, **4**, 129 - 53.
86. Manduchi, E., Scarce, L., Brestelli, J., Grant, G., Kaestner, K. and Stoeckert, C. (2002 Sep 3) Comparison of different labeling methods for two-channel high-density microarray experiments. *Physiol Genomics*, **10**(3), 169 - 79.
87. Richter, A., Schwager, C., Hentze, S., Ansorge, W., Hentze, M. and Muckenthaler, M. (2002 Sep) Comparison of fluorescent tag DNA labeling methods used for expression analysis by DNA microarrays. *Biotechniques*, **33**(3), 620 - 8, 630.
88. Kendzioriski, C., Zhang, Y., Lan, H. and Attie, A. (2003 Jul) The efficiency of pooling mRNA in microarray experiments. *Biostatistics*, **4**(3), 465 - 77.
89. Gupta, V., Cherkassky, A., Chatis, P., Joseph, R., Johnson, A., Broadbent, J., Erickson, T. and DiMeo, J. (2003 Feb 15) Directly labeled mRNA produces highly precise and unbiased differential gene expression data. *Nucleic Acids Res*, **31**(4), e13.
90. Schroeder, B., Peterson, L. and Fleischmann, R. (2002 Mar) Improved quantitation and reproducibility in Mycobacterium tuberculosis DNA microarrays. *J Mol Microbiol Biotechnol*, **4**(2), 123 - 6.
91. Mujumdar, R., Ernst, L., Mujumdar, S., Lewis, C. and Waggoner, A. (1993 Mar-Apr) Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug Chem*, **4**(2), 105 - 11.
-



- 
92. Feinberg, A. and Vogelstein, B. (1983 Jul 1) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem*, **132**(1), 6 - 13.
93. Winkles, J. and Grainger, R. (1985 Jul) Polyadenylation state of abundant mRNAs during *Drosophila* development. *Dev Biol*, **110**(1), 259 - 63.
94. Lanzotti, D., Kaygun, H., Yang, X., Duronio, R. and Marzluff, W. (2002 Apr) Developmental control of histone mRNA and dSLBP synthesis during *Drosophila* embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo. *Mol Cell Biol*, **22**(7), 2267 - 82.
95. Hauser, N., Vingron, M., Scheideler, M., Krems, B., Hellmuth, K., Entian, K. and Hoheisel, J. (1998 Sep 30) Transcriptional profiling on all open reading frames of *Saccharomyces cerevisiae*. *Yeast*, **14**(13), 1209 - 21.
96. Sambrook, J. u. R., D. W. (2001) Molecular cloning, a laboratory manual, 3 Ed., *Cold Spring Harbor Laboratory Press*, New York.
97. Taylor, S., Smith, S., Windle, B. and Guiseppi-Elie, A. (2003 Aug 15) Impact of surface chemistry and blocking strategies on DNA microarrays. *Nucleic Acids Res*, **31**(16), e87.
98. Liang, M., Briggs, A., Rute, E., Greene, A. and Cowley, A. (2003 Aug 15) Quantitative assessment of the importance of dye switching and biological replication in cDNA microarray studies. *Physiol Genomics*, **14**(3), 199 - 207.
99. Quackenbush, J. (2002 Dec) Microarray data normalization and transformation. *Nat Genet*, **32 Suppl**, 496 - 501.
100. Tsodikov, A., Szabo, A. and Jones, D. (2002 Feb) Adjustments and measures of differential expression for microarray data. *Bioinformatics*, **18**(2), 251 - 60.
101. Yang, Y., Dudoit, S., Luu, P., Lin, D., Peng, V., Ngai, J. and Speed, T. (2002 Feb 15) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, **30**(4), e15.
102. Colantuoni, C., Henry, G., Zeger, S. and Pevsner, J. (2002 Jun) Local mean normalization of microarray element signal intensities across an array surface: quality control and correction of spatially systematic artifacts. *Biotechniques*, **32**(6), 1316 - 20.
103. Chen, Y., Kodell, R., Sistare, F., Thompson, K., Morris, S. and Chen, J. (2003 Feb) Normalization methods for analysis of microarray gene-expression data. *J Biopharm Stat*, **13**(1), 57 - 74.
104. Park, T., Yi, S., Kang, S., Lee, S., Lee, Y. and Simon, R. (2003 Sep 2) Evaluation of normalization methods for microarray data. *BMC Bioinformatics*, **4**(1), 33.
105. Wilson, D., Buckley, M., Helliwell, C. and Wilson, I. (2003 Jul 22) New normalization methods for cDNA microarray data. *Bioinformatics*, **19**(11), 1325 - 32.
106. Chatterjee, S. a. P., B. (1991) Regression Analysis by Example, *John Wiley & Sons*, New York.
107. Kepler, T., Crosby, L. and Morgan, K. (2002 Jun 28) Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biol*, **3**(7), RESEARCH0037.
108. Yeung, K. and Ruzzo, W. (2001 Sep) Principal component analysis for clustering gene expression data. *Bioinformatics*, **17**(9), 763 - 74.
-

- 
109. Peterson, L. (2003 Feb) Partitioning large-sample microarray-based gene expression profiles using principal components analysis. *Comput Methods Programs Biomed*, **70**(2), 107 - 19.
110. Dudoit, S. and Fridlyand, J. (2002 Jun 25) A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol*, **3**(7), RESEARCH0036.
111. Dudoit, S. and Fridlyand, J. (2003 Jun 12) Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, **19**(9), 1090 - 9.
112. Wang, J., Delabie, J., Aasheim, H., Smeland, E. and Myklebost, O. (2002 Nov 24) Clustering of the SOM easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study. *BMC Bioinformatics*, **3**(1), 36.
113. Weinstein, J., Myers, T., O'Connor, P., Friend, S., Fornace, A., Kohn, K., Fojo, T., Bates, S., Rubinstein, L., Anderson, N., Buolamwini, J., van Osdol, W., Monks, A., Scudiero, D., Sausville, E., Zaharevitz, D., Bunow, B., Viswanadhan, V., Johnson, G., Wittes, R. and Paull, K. (1997 Jan 17) An information-intensive approach to the molecular pharmacology of cancer. *Science*, **275**(5298), 343 - 9.
114. Gibbons, F. and Roth, F. (2002 Oct) Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res*, **12**(10), 1574 - 81.
115. Dettling, M. and Bühlmann, P. (2002) Supervised clustering of genes. *Genome Biol*, **3**(12), RESEARCH0069.
116. Brown, M., Grundy, W., Lin, D., Cristianini, N., Sugnet, C., Furey, T., Ares, M. and Hausler, D. (2000 Jan 4) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A*, **97**(1), 262 - 7.
117. Qian, J., Lin, J., Luscombe, N., Yu, H. and Gerstein, M. (2003 Oct 12) Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. *Bioinformatics*, **19**(15), 1917 - 26.
118. Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C. and Lander, E. (1999 Oct 15) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**(5439), 531 - 7.
119. von Heydebreck, A., Huber, W., Poustka, A. and Vingron, M. (2001) Identifying splits with clear separation: a new class discovery method for gene expression data. *Bioinformatics*, **17** **Suppl 1**, S107 - 14.
120. Baggerly, K., Coombes, K., Hess, K., Stivers, D., Abruzzo, L. and Zhang, W. (2001) Identifying differentially expressed genes in cDNA microarray experiments. *J Comput Biol*, **8**(6), 639 - 59.
121. Tsai, C., Chen, Y. and Chen, J. (2003 May 1) Testing for differentially expressed genes with microarray data. *Nucleic Acids Res*, **31**(9), e52.
122. Kerr, M. and Churchill, G. (2001 Apr) Statistical design and the analysis of gene expression microarray data. *Genet Res*, **77**(2), 123 - 8.
123. Wernisch, L., Kendall, S., Soneji, S., Wietzorrek, A., Parish, T., Hinds, J., Butcher, P. and Stoker, N. (2003 Jan) Analysis of whole-genome microarray replicates using mixed models. *Bioinformatics*, **19**(1), 53 - 61.
124. Yang, Y., Hoh, J., Broger, C., Neeb, M., Edington, J., Lindpaintner, K. and Ott, J. (2003) Statistical
-

- 
- methods for analyzing microarray feature data with replications. *J Comput Biol*, **10**(2), 157 - 69.
125. Yeung, K., Haynor, D. and Ruzzo, W. (2001 Apr) Validating clustering for gene expression data. *Bioinformatics*, **17**(4), 309 - 18.
126. De Smet, F., Mathys, J., Marchal, K., Thijs, G., De Moor, B. and Moreau, Y. (2002 May) Adaptive quality-based clustering of gene expression profiles. *Bioinformatics*, **18**(5), 735 - 46.
127. Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W. and White, K. P. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science*, **297**(5590), 2270-5.
128. White, K., Rifkin, S., Hurban, P. and Hogness, D. (1999 Dec 10) Microarray analysis of *Drosophila* development during metamorphosis. *Science*, **286**(5447), 2179 - 84.
129. Furlong, E., Andersen, E., Null, B., White, K. and Scott, M. (2001 Aug 31) Patterns of gene expression during *Drosophila* mesoderm development. *Science*, **293**(5535), 1629 - 33.
130. Egger, B., Leemans, R., Loop, T., Kammermeier, L., Fan, Y., Radimerski, T., Strahm, M., Certa, U. and Reichert, H. (2002 Jul) Gliogenesis in *Drosophila*: genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. *Development*, **129**(14), 3295 - 309.
131. Butler, M., Jacobsen, T., Cain, D., Jarman, M., Hubank, M., Whittle, J., Phillips, R. and Simcox, A. (2003 Feb) Discovery of genes with highly restricted expression patterns in the *Drosophila* wing disc using DNA oligonucleotide microarrays. *Development*, **130**(4), 659 - 70.
132. McDonald, M. and Rosbash, M. (2001 Nov 30) Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell*, **107**(5), 567 - 78.
133. Etter, P. and Ramaswami, M. (2002 Jun) The ups and downs of daily life: profiling circadian gene expression in *Drosophila*. *Bioessays*, **24**(6), 494 - 8.
134. Jin, W., Riley, R., Wolfinger, R., White, K., Passador-Gurgel, G. and Gibson, G. (2001 Dec) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet*, **29**(4), 389 - 95.
135. Meiklejohn, C., Parsch, J., Ranz, J. and Hartl, D. (2003 Aug 19) Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A*, **100**(17), 9894 - 9.
136. Klebes, A., Biehs, B., Cifuentes, F. and Kornberg, T. (2002 Jul 24) Expression profiling of *Drosophila* imaginal discs. *Genome Biol*, **3**(8), RESEARCH0038.
137. Rifkin, S., Kim, J. and White, K. (2003 Feb) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet*, **33**(2), 138 - 44.
138. Shoemaker, D., Schadt, E., Armour, C., He, Y., Garrett-Engle, P., McDonagh, P., Loerch, P., Leonardson, A., Lum, P., Cavet, G., Wu, L., Altschuler, S., Edwards, S., King, J., Tsang, J., Schimmack, G., Schelter, J., Koch, J., Ziman, M., Marton, M., Li, B., Cundiff, P., Ward, T., Castle, J., Krolewski, M., Meyer, M., Mao, M., Burchard, J., Kidd, M., Dai, H., Phillips, J., Linsley, P., Stoughton, R., Scherer, S. and Boguski, M. (2001 Feb 15) Experimental annotation of the human genome using microarray technology. *Nature*, **409**(6822), 922 - 7.
139. Misra, S., Crosby, M., Mungall, C., Matthews, B., Campbell, K., Hradecky, P., Huang, Y., Kaminker, J., Millburn, G., Prochnik, S., Smith, C., Tupy, J., Whitfield, E., Bayraktaroglu, L., Berman, B., Bettencourt, B., Celniker, S., de Grey, A., Drysdale, R., Harris, N., Richter, J., Russo, S.,
-

- Schroeder, A., Shu, S., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W., Rubin, G. and Lewis, S. (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol*, **3**(12), RESEARCH0083.
140. Lawrence, P. A. (1992) The Making of a Fly, *Blackwell Scientific Publications*, Oxford.
141. Tomancak, P., Beaton, A., Weiszmamm, R., Kwan, E., Shu, S., Lewis, S., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. and Rubin, G. (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol*, **3**(12), RESEARCH0088.
142. Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M. and Levine, M. (2002 Nov 27) Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell*, **111**(5), 687 - 701.
143. Montalta-He, H. and Reichert, H. (2003) Impressive expressions: developing a systematic database of gene-expression patterns in *Drosophila* embryogenesis. *Genome Biol*, **4**(2), 205.
144. Wang, C. I. and Lehmann, R. (1991) Nanos is the localized posterior determinant in *Drosophila*. *Cell*, **66**(4), 637--647.
145. Wilson, J. E., Connell, J. E., Schlenker, J. D. and Macdonald, P. M. (1996) Novel genetic screen for genes involved in posterior body patterning in *Drosophila*. *Dev. Genet.*, **19**(3), 199--209.
146. Oishi, I., Sugiyama, S., Otani, H., Yamamura, H., Nishida, Y. and Minami, Y. (1998) A novel *Drosophila* nuclear protein serine/threonine kinase expressed in the germline during its establishment. *Mech. Dev.*, **71**(1,2), 49--63.
147. Xu, J., Xin, S. and Du, W. (2001) *Drosophila* Chk2 is required for DNA damage-mediated cell cycle arrest and apoptosis. *FEBS Lett.*, **508**(3), 394--398.
148. Anderson, K. V., Jurgens, G. and Nusslein-Volhard, C. (1985) Establishment of dorsal-ventral polarity in the *Drosophila* embryo. Genetic studies on the role of the Toll gene product. *Cell*, **42**, 779--789.
149. Pankratz, M. J., Hoch, M., Seifert, E. and Jaekle, H. (1989) Kruppel requirement for knirps enhancement reflects overlapping gap gene activities in the *Drosophila* embryo. *Nature*, **341**, 337--340.
150. Hill, K. K., Bedian, V., Juang, J. L. and Hoffmann, F. M. (1995) Genetic interactions between the *Drosophila* Abelson, Abl, tyrosine kinase and failed axon connections (Fax), a novel protein in axon bundles. *Genetics*, **141**(2), 595--606.
151. Brose, K., Bland, K. S., Wang, K. H., Arnott, D., Henzel, W., Goodman, C. S., Tessier-Lavigne, M. and Kidd, T. (1999) Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell*, **96**(6), 795--806.
152. Keleman, K. and Dickson, B. J. (2001) Short- and long-range repulsion by the *Drosophila* Unc5 Netrin receptor. *Neuron*, **32**(4), 605--617.
153. Holmes, A. L., Raper, R. N. and Heilig, J. S. (1998) Genetic analysis of *Drosophila* larval optic nerve development. *Genetics*, **148**(3), 1189--1201.
154. Watanabe, T. and Kankel, D. R. (1992) The l(1)ogre gene of *Drosophila melanogaster* is expressed in postembryonic neuroblasts. *Dev. Biol.*, **152**, 172--183.
155. Alton, A. K., Fechtel, K., Kopczynski, C. C., Shepard, S. B., Kooh, P. J. and Muskavitch, M. A.

- 
- (1989) Molecular genetics of Delta, a locus required for ectodermal differentiation in *Drosophila*. *Dev. Genet.*, **10**, 261--272.
156. Mlodzik, M., Baker, N. E. and Rubin, G. M. (1990) Isolation and expression of scabrous, a gene regulating neurogenesis in *Drosophila*. *Genes Dev.*, **4**, 1848--1861.
157. Hoppe, P. E. and Greenspan, R. J. (1986) Local function of the Notch gene for embryonic ectodermal pathway choice in *Drosophila*. *Cell*, **46**, 773--783.
158. Artavanis-Tsakonas, S., Matsuno, K. and Fortini, M. (1995 Apr 14) Notch signaling. *Science*, **268** (5208), 225 - 32.
159. Brennan, K., Tateson, R., Lewis, K. and Martinez Arias, A. (1997) A functional analysis of Notch mutations in *Drosophila*. *Genetics*, **147**(1), 177--188.
160. Tamkun, J. (1995 Aug) The role of brahma and related proteins in transcription and development. *Curr Opin Genet Dev*, **5**(4), 473 - 7.
161. Tsukiyama, T., Daniel, C., Tamkun, J. and Wu, C. (1995 Dec 15) ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, **83**(6), 1021 - 6.
162. Ahmad, K. and Henikoff, S. (2002 Dec 10) Histone H3 variants specify modes of chromatin assembly. *Proc Natl Acad Sci U S A*, **99 Suppl 4**, 16477 - 84.
163. Riemer, D., Stuurman, N., Berrios, M., Hunter, C., Fisher, P. and Weber, K. (1995 Oct) Expression of *Drosophila* lamin C is developmentally regulated: analogies with vertebrate A-type lamins. *J Cell Sci*, **108 ( Pt 10)**, 3189 - 98.
164. Ohno, K., Hirose, F., Inoue, Y. H., Takisawa, H., Mimura, S., Hashimoto, Y., Kiyono, T., Nishida, Y. and Matsukage, A. (1998) cDNA cloning and expression during development of *Drosophila melanogaster* MCM3, MCM6 and MCM7. *Gene*, **217**(1-2), 177--186.
165. Baxevanis, A. D., Arents, G., Moudrianakis, E. N. and Landsman, D. (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res.*, **23**(14), 2685--2691.
166. Orphanides, G., Lagrange, T. and Reinberg, D. (1996 Nov 1) The general transcription factors of RNA polymerase II. *Genes Dev*, **10**(21), 2657 - 83.
167. Aoyagi, N. and Wassarman, D. A. (2000) Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIIID components contributes to gene-specific transcriptional regulation. *J. Cell Biol.*, **150**(2), F45--F49.
168. Xu, H., Brill, J., Hsien, J., McBride, R., Boulianne, G. and Trimble, W. (2002 Nov 15) Syntaxin 5 is required for cytokinesis and spermatid differentiation in *Drosophila*. *Dev Biol*, **251**(2), 294 - 306.
169. Brogna, S., Sato, T. and Rosbash, M. (2002 Oct) Ribosome components are associated with sites of transcription. *Mol Cell*, **10**(4), 93 - 104.
170. Hölzl, H., Kapelari, B., Kellermann, J., Seemüller, E., Sümegi, M., Udvardy, A., Medalia, O., Sperling, J., Müller, S., Engel, A. and Baumeister, W. (2000 Jul 10) The regulatory complex of *Drosophila melanogaster* 26S proteasomes. Subunit composition and localization of a deubiquitylating enzyme. *J Cell Biol*, **150**(1), 119 - 30.
-

- 
171. Marsh, J. L., Erfle, M. and Leeds, C. A. (1986) Molecular localization, developmental expression and nucleotide sequence of the  $\alpha$ -methyl dopa hypersensitive gene of *Drosophila*. *Genetics*, **114**, 453--467.
172. Riddiford, L. (1993 Fall) Hormone receptors and the regulation of insect metamorphosis. *Receptor*, **3**(3), 203 - 9.
173. Thummel, C. (1996 Aug) Files on steroids--*Drosophila* metamorphosis and the mechanisms of steroid hormone action. *Trends Genet*, **12**(8), 306 - 10.
174. Li, T. and White, K. (2003 Jul) Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. *Dev Cell*, **5**(1), 59 - 72.
175. Cherbas, P. (1993) Ecdysone-responsive genes (The IVth Karlson Lecture). *Insect Biochem. Molec. Biol.*, **23**(1), 3--11.
176. Kimbrell, D. A., Tojo, S. J., Alexander, S., Brown, E. E., Tobin, S. L. and Fristrom, J. W. (1989) Regulation of larval cuticle protein gene expression in *Drosophila melanogaster*. *Dev. Genet.*, **10**(4), 198--209.
177. Prohaska, C., Karlan, A. and Andres, A. J. (2003) , A. Dros. Res. Conf. 44, pp. 240C.
178. Deutsch, J., Laval, M., Lepesant, J. A., Maschat, F., Pourrain, F. and Rat, L. (1989) Larval fat body-specific gene expression in *Drosophila melanogaster*. *Dev. Genet.*, **10**(3), 220--231.
179. Birr, C., Fristrom, D., King, D. and Fristrom, J. (1990 Sep) Ecdysone-dependent proteolysis of an apical surface glycoprotein may play a role in imaginal disc morphogenesis in *Drosophila*. *Development*, **110**(1), 239 - 48.
180. Caggese, C., Ragone, G., Perrini, B., Moschetti, R., De Pinto, V., Caizzi, R. and Barsanti, P. (1999) Identification of nuclear genes encoding mitochondrial proteins: isolation of a collection of *D. melanogaster* cDNAs homologous to sequences in the Human Gene Index database. *Molec. gen. Genet.*, **261**(1), 64--70.
181. Freeman, M. (1997 Jan) Cell determination strategies in the *Drosophila* eye. *Development*, **124** (2), 261 - 70.
182. Begun, D. J., Whitley, P., Todd, B. L., Waldrip-Dail, H. M. and Clark, A. G. (2000) Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics*, **156**(4), 1879--1888.
183. Bownes, M. (1994) The regulation of the yolk protein genes, a family of sex differentiation genes in *Drosophila melanogaster*. *BioEssays*, **16**(10), 745--752.
184. Furlong, E., Andersen, E., Null, B., White, K. and Scott, M. (2001 Aug 31) Patterns of gene expression during *Drosophila* mesoderm development. *Science*, **293**(5535), 1629 - 33.
185. James, P., Vorherr, T. and Carafoli, E. (1995 Jan) Calmodulin-binding domains: just two faced or multi-faceted? *Trends Biochem Sci*, **20**(1), 38 - 42.
186. de la Vega, H., Specht, C. A., Liu, Y. and Robbins, P. W. (1998) Chitinases are a multi-gene family in *Aedes*, *Anopheles* and *Drosophila*. *Insect Molec. Biol.*, **7**(3), 233--239.