

INAUGURAL – DISSERTATION  
zur  
Erlangung der Doktorwürde  
der  
Naturwissenschaftlich-Mathematischen Gesamtfakultät  
der  
Ruprecht-Karls-Universität  
Heidelberg

vorgelegt von  
Britta Koch  
aus Lemgo

Tag der mündlichen Prüfung:

Thema

**Entwicklung eines DNA Microarrays zur genomweiten  
Analyse Polycomb-vermittelter Genregulation  
in *Drosophila melanogaster***

Gutachter: Prof. Dr. Renato Paro  
Prof. Dr. Gabriele Petersen

**Freudig war, vor vielen Jahren  
Eifrig so der Geist bestrebt,  
Zu erforschen, zu erfahren,  
Wie Natur im Schaffen liegt.  
Und es ist das ewig Eine,  
Das sich vielfach offenbart,  
Klein das Große, Gross das Kleine,  
Alles nach der eigenen Art.  
Immer wechselnd, fest sich haltend,  
Nah und fern und fern und nah;  
So gestaltend, umgestaltend –  
Zum Erstaunen bin ich da.**

Johann Wolfgang von Goethe

## Danksagung

Diese Arbeit wurde am Zentrum für Molekulare Biologie in Heidelberg unter der Leitung von Prof. Dr. Renato Paro durchgeführt.

Mein besonderer Dank gilt Prof. Dr. Renato Paro für die Bereitstellung des Themas und die vielfältigen Anregungen während dieser Arbeit. Für die Übernahme des Zweitgutachtens möchte ich Prof. Dr. Gabriele Petersen danken.

Ganz herzlichen Dank an Marc Hild für die gute Betreuung und die ständige Diskussions- und Hilfsbereitschaft während unserer gemeinsamen Zeit am ZMBH und darüber hinaus.

Christian Beisel möchte ich vor allem für die gute Zusammenarbeit während der letzten Jahre, seine Diskussionsbereitschaft und gute Hybridisierungstipps danken. Ebenso danke ich Alexander Löwer für die vielen interessanten Gespräche, spontane Hilfe bei Computerproblemen, gute Literatur und un poco español – rayos! Allen dreien danke ich auch für das sorgfältige Gegenlesen dieser Arbeit.

Stefan Haas war vom Primerdesign bis zum Korrekturlesen dieser Arbeit in allen Bereichen unersetzlich – danke!

Die Durchführung der PCRs und das Entwicklungsprofil wären ohne die Unterstützung von Sylvia Hagar nicht denkbar gewesen, auch dafür ein großes Dankeschön.

Michael Boutros danke ich für die gute Zusammenarbeit und für gute Ideen bei den RNAi Experimenten, Leonie Ringrose für die Bereitstellung der Kartierungsdaten der Pc-Bindungsstellen, Maria Leptin für die Begutachtung der *in-situ* Hybridisierungen, Gundula Edenfeld für viele motivierende Diskussionen und der Gruppe Hoheisel für die Hilfestellung beim Spotten der Microarrays und die Durchführung der Microarray-Hybridisierungen für das Entwicklungsprofil.

Christian Beisel, Michaela Bilic, Bodo Brückner, Inhua Chen, Andrea Frank, Sylvia Hagar, Marc Hild, Heidi Ehret, Nara Lee, Alexander Löwer, Cedric Maurange, Gunter Merdes, Ana Monqaut, Melanie Müller, Tariq Muhammad, Christian Popp, Matthias Prestel, Gerhard Rank, Leonie Ringrose, Stefan Schönfelder und Sabine Schmitt – Euch allen ein großes Dankeschön für die hervorragende Atmosphäre im Labor!

---

**Danksagung**

---

**Zusammenfassung**

---

**Abstract**

---

---

**1 Einleitung**

---

**1**

1.1 Der Lebenszyklus von <i>Drosophila melanogaster</i>	1
1.2 Frühe Musterbildung in <i>Drosophila</i>	5
1.3 Die Aufrechterhaltung eines Expressionsmusters während der Entwicklung	7
1.4 DNA Microarrays	12
1.5 Genvorhersagen in <i>Drosophila</i>	14
1.6 Ziel dieser Arbeit	16

---

**2 Ergebnisse**

---

**17**

2.1 Die Herstellung eines PCR basierten DNA Transkriptom-Microarrays	17
2.2 Das Entwicklungsprofil von <i>Drosophila melanogaster</i>	20
2.3 Validierung der Microarrayanalysen mittels RT-PCR	22
2.4 Weitere Validierung und Analyse mittels <i>in-situ</i> Hybridisierung	23
2.5 Neue Gene, die mit bekannten Genen assoziiert sind	26
2.6 Neue Gene, die in der Nähe bekannter Gene liegen	32
2.7 HDC12511 zeigt ektodermale Expression	33
2.8 Neue Gene in isolierten Lokalisationen	40
2.9 Die HeidelbergFlyArray Web-Seite	49
2.10 Zielgene für Polycomb/Trithorax Gruppen Proteine	52
2.11 Untersuchung Polycomb-mutanter Embryonen und Wildtypembryonen mittels Microarrayanalysen	52
2.12 Vergleiche mit weiteren Daten über Polycomb-Bindungsstellen	56

---

**3 Diskussion**

---

**61**

3.1 Beiträge der einzelnen Mitarbeiter an der Herstellung und Validierung des Transkriptomarrays	61
3.2 Genvorhersagen für <i>Drosophila melanogaster</i>	62

---

3.3 Der Heidelberg Transkriptom-Microarray	64
3.4 Wie viele Gene gibt es in <i>Drosophila melanogaster</i> ?	67
3.5 In-situ Hybridisierungen	70
3.6 Zielgene von Polycomb	76
3.7 Ausblick	79
<b>4 Material</b>	<b>81</b>
<hr/>	
4.1 Laborausstattung	81
4.2 Software, Hardware und Internetadressen	82
4.3 Verbrauchsmaterial	83
4.4 DNA	85
4.5 Bakterienstämme ( <i>Escherichia coli</i> )	85
4.6 Fliegenstämme	86
4.7 Medien, Puffer und Stammlösungen	87
<b>5 Methoden</b>	<b>92</b>
<hr/>	
5.1 Molekularbiologische Methoden	92
5.2 Genetische Methoden	100
5.3 Histologische Methoden	102
5.4 Microarray-spezifische Methoden	104
<b>6 Referenzen</b>	<b>114</b>
<hr/>	
<b>7 Veröffentlichungen</b>	<b>127</b>
<hr/>	
<b>8 Abkürzungen</b>	<b>128</b>
<hr/>	
8.1 Genetik	128
8.2 Sonstige	129
<b>9 Anhang</b>	<b>130</b>
<hr/>	

## Zusammenfassung

Durch die Sequenzierungsprojekte der letzten Jahre konnte eine eindrucksvolle Menge an Daten über die Anzahl und Struktur der Gene sowie über die strukturelle und funktionelle Organisation von Genomen angesammelt werden. Diese Informationen bilden die Grundlage für Projekte, die es in Zukunft ermöglichen werden, die Rolle jeden Gens innerhalb des komplexen Zusammenspiels der Gene im lebenden System zu erforschen. Eine Methode, die auf solche Fragestellungen ausgerichtet ist, nutzt DNA Microarrays. Mit ihrer Hilfe kann die Expression von Genen eines gesamten Genoms zu bestimmten Zeitpunkten oder in verschiedenen Geweben untersucht werden. Auch epigenetische Mechanismen der Regulation können mit einem solchen System untersucht werden. Zwei Gruppen von Genen, die eine wichtige Rolle in diesem als „zelluläres Gedächtnis“ bezeichneten Prozess spielen, sind die Polycomb-Gruppe und die trithorax-Gruppe. Während die trithorax-Gruppe die Aufgabe erfüllt, das Expressionsmuster von aktiven Genen während der Entwicklung aufrecht zu erhalten, kommt der Polycomb-Gruppe die Aufgabe zu, den reprimierten Zustand von Genen über viele Zellteilungen sicherzustellen. Innerhalb dieser Arbeit wurden mit Hilfe der Microarraytechnik Gene untersucht, die der regulatorischen Funktion dieser beiden Gengruppen unterliegen

Hierzu wurde in Kollaboration mit anderen Gruppen ein PCR-basierter DNA-Microarray entwickelt, der ausgewählte Sequenzen für 20,948 ORFs enthält, die in einer neuen Annotation der *Drosophila* Genomsequenz identifiziert wurden. Um die in dieser Annotation enthaltenen 7,400 Gene zu validieren, die in vorangegangenen Genvorhersagen nicht berücksichtigt wurden, wurde zunächst mit Hilfe des Microarrays ein Entwicklungsprofil mit neun verschiedenen Stadien des Lebenszyklus von *Drosophila* durchgeführt. Durch diese Analysen und molekularbiologische Validierungen über RT-PCR und *in-situ* Hybridisierungen konnten etwa 2,000 neue Gene für *Drosophila* bestätigt werden.

Dieser Microarray wurde in weiteren Experimenten eingesetzt, um Hinweise auf Zielgene Polycomb-abhängiger Regulation zu bekommen. Dazu wurden Methoden entwickelt, um die Zielgene zu verschiedenen Zeitpunkten und in unterschiedlichen Geweben ermitteln zu können. Der Vergleich von cDNA Polycomb mutanter Fliegen mit Kontroll-cDNA auf dem Microarray gab erste Hinweise auf Zielgene in frühen embryonalen Stadien.

Insgesamt legt diese Arbeit eine wichtige Grundlage für die Analyse aller Gene in *Drosophila melanogaster*, die über den Mechanismus des „zellulären Gedächtnisses“ reguliert werden. Dieses Wissen wird dazu beitragen können, die komplexen Prozesse während der Entstehung eines Lebewesens besser zu verstehen.

## Abstract

The immense efforts invested in the sequencing projects of the last years provided an enormous amount of digital information about the number and structure of the genes, as well as of the structural and functional organization of entire genomes. This information set the basis for experiments, which in future will permit the analyses of the function of each gene in the context of the Transcriptome. DNA microarrays are effective tools for these investigations. They allow the examination of the expression of mRNA pools in the context of the entire organism, at different time points of development or in various tissues. During this work, microarrays were used to study the mechanisms of epigenetic regulation. The Polycomb group and the trithorax group of genes accomplish elemental functions in this process termed 'cellular memory', and genes that are subject to this regulatory process were identified using microarrays.

Therefore, a PCR-based DNA microarray was developed containing 20,948 ORFs that were identified in a new annotation of the *Drosophila* genomic sequence. In order to confirm the 7,400 genes that had not been predicted in genome annotations carried out by others, a developmental profiling approach was performed, which included nine different stages covering the complete life cycle of *Drosophila*. These analyses, together with RT-PCR and *in-situ* hybridisation experiments for validation, led to the identification of approximately 2,000 novel genes in the *Drosophila* genome.

The novel DNA Microarray was subsequently used to identify genes that are subject to the regulation by the epigenetic machinery. Thus, different methods were developed to allow the identification of target genes of the Polycomb/trithorax groups at different time points of development. Comparing the cDNA of Pc-mutant flies with wild type cDNA on Microarrays led to the first indication of such target genes.

Taken together this work provides the basis to further analyse all target genes of epigenetic mechanisms in a genome-wide context. This knowledge will eventually deepen the understanding of how the accurate development of an organism is attained.

# 1 Einleitung

Eine der zentralen Fragen der Biologie ist es, wie aus einer Zygote ein komplexer Organismus entstehen kann. Seit Beginn des 20. Jahrhunderts ist die Taufliege *Drosophila melanogaster* ein beliebter Modellorganismus um Gene zu erforschen, die an der Entwicklung und der Erhaltung eines Lebewesens beteiligt sind. Die Vorteile von *Drosophila* als Labortier sind neben ihrer geringen Größe und ihrer kurzen Generationszeit vor allem die Vielfalt zur Verfügung stehender genetischer Werkzeuge, die eine umfangreiche Manipulation des Genoms gewährleisten.

Viele der grundlegenden Erkenntnisse über die molekularen Prozesse während der Embryogenese, wie zum Beispiel die Achsendeterminierung oder die Transkriptionsregulation bei der Entstehung der Musterbildung, konnten mit Hilfe von Untersuchungen an *Drosophila* gewonnen werden. Die kürzlich erfolgte vollständige Sequenzierung des *Drosophila* Genoms lieferte eine weitere entscheidende Grundlage, um Regulationskaskaden in *Drosophila* zu untersuchen und einen tieferen Einblick in die Entstehung eines komplexen Organismus zu bekommen.

## **1.1 Der Lebenszyklus von *Drosophila melanogaster***

### **1.1.1 Die Embryogenese**

Während der Entwicklung durchläuft *Drosophila* eine komplette Metamorphose, der Lebenszyklus beinhaltet also embryonale Stadien, Larvalstadien und Puppenstadien, bevor am Ende das Adulttier schlüpft.

Die embryonale Entwicklung von *Drosophila* dauert etwa einen Tag, in dem sich der Embryo von der Zygote bis zur komplexen Larve entwickelt. Dabei durchlaufen die Kerne der Zygote zunächst 13 Teilungen. Die ersten 7 dieser Mitosen sind synchron (Zalokar 1976) und ergeben ein Syncytium von 128 Zellkernen, die ellipsenförmig im Dotter angeordnet sind. Während der nächsten drei Teilungen wandern die meisten Zellkerne an die Peripherie, und bilden das sogenannte syncytiale Blastoderm. 28 Kerne verbleiben jedoch in der Mitte des Embryos, um dort die Dotterkerne (Vitellogen) zu bilden, und weitere 2-3 wandern an den posterioren Pol des Eies und bilden dort die Polzellen, die später die Keimzellen bilden werden. Alle Kerne teilen sich drei weitere Male. Die zukünftigen Keimzellen

durchlaufen später weitere, von den somatischen Zellen unabhängige Teilungen, wohingegen sich die Vitellophagen nicht mehr teilen und polyploid werden. Nach der dreizehnten Teilung besteht das syncytiale Blastoderm aus etwa 5,000 Kernen. Die Bildung der somatischen Zellen geschieht durch die Invagination von Membranen zwischen diesen Zellkernen.

Nach etwa dreieinhalb Stunden durchläuft der Embryo die Gastrulation. Dabei stülpt sich das zukünftige Mesoderm ein und bildet entlang der ventralen Mittellinie eine Furche. Die Anlagen des anterioren Mitteldarms, des Mesoderms, des Proctodeums und des posterioren Mitteldarms werden in dieser Furche festgelegt. Im zweiten Schritt der Gastrulation beginnen Zellen am posterioren Pol, sich nach antero-dorsal zu bewegen, und damit den so genannten Keimstreifen zu bilden. Dieses Gewebe umfasst die Primordien der drei gnathalen, drei thorakalen und neun abdominalen Segmente. Während der Elongation des Keimstreifens kommt es zu starken Veränderungen der ektodermalen und mesodermalen Anlagen, durch die nach und nach das segmentale Muster sichtbar wird.

Bevor der Keimstreifen innerhalb von zwei Stunden seine maximale Elongation erreicht, invaginieren die Zellen, die später das zentrale Nervensystem bilden, in das Innere des Embryos und bilden zwischen dem Mesoderm und dem äußeren Ektoderm die Neuroblasten. Diese Invagination verläuft in drei Wellen, die drei unterschiedliche Subpopulationen von Neuroblasten bilden. Nach ihrer Segregation beginnen die Neuroblasten durch Teilung die Ganglionmutterzellen zu bilden, aus denen im Verlauf weiterer Teilungen die Neuronen und Gliazellen entstehen. Gleichzeitig zur Neurogenese entwickeln sich zwei Einstülpungen an beiden Seiten des zukünftigen anterioren- und posterioren Mitteldarms, die zum Entoderm des Mitteldarmes verwachsen.

Nach diesen Entwicklungen beginnt der Keimstreifen sich wieder zurückzuziehen. Während dieser Verkürzung kommt es zu einer Reihe morphologischer Prozesse, wie der Fusion der Darmanlagen, der Verbindung der verschiedenen Fragmente der Tracheen, und der Bildung der Gonaden. Bereits zu Beginn dieses Rückzuges kann man die erste Segmentierung des Embryos sehen. Eine Reihe gleichmäßig angeordneter Vertiefungen stellt die Grenzen der Parasegmente dar, aus denen sich später die Segmente der Larve und der Imagos entwickeln. Die Parasegmente und Segmente sind gegeneinander verschoben, wobei ein Segment jeweils aus dem hinteren Teil des anterioren und aus dem vorderen Teil des posterior gele-

genen Parasegmentes besteht. Insgesamt besteht der Embryo aus 14 Parasegmenten, von denen drei an der Bildung der gnathalen Segmente (Mandibel, Maxille und Labium) der Larve, drei an der Bildung der thorakalen (Pro- Meso- und Metathorax) und neun an der Bildung der abdominalen Segmente beteiligt sind.

Nach neuneinhalb Stunden hat sich der Keimstreifen des Embryos wieder vollständig zurückgezogen. Zu diesem Zeitpunkt ist der Embryo an der dorsalen Seite offen, die Amnioserosa ist hier im direkten Kontakt mit der Vitellinmembran. Durch die Ausdehnung der dorsalen Epidermis über die Amnioserosa kommt es zum dorsalen Schluss. Die Zellen der Amnioserosa werden dadurch in den Embryo aufgenommen und bilden später Zellen des Kopfes.

Zu der Zeit des Keimstreifrückzuges kommt es auch zu komplexen morphologischen Bewegungen, die die Kopfinvagination einleiten. Diese Bewegungen sind besonders für die gnathalen Segmente von Bedeutung. Dabei treffen die Speicheldrüsen an der ventralen Mittellinie aufeinander und fusionieren zu einem gemeinsamen Duktus. Die Mund- und Schlundpartien des Embryos werden ebenfalls durch diese Bewegungen gebildet.

Nach dem Rückzug des Keimstreifens und der Kopfinvagination ist die Morphogenese des Embryos abgeschlossen. Ab der 17. Stunde kann man Muskelbewegungen erkennen und die Tracheen füllen sich mit Luft, bevor kurz darauf die Larve schlüpft.

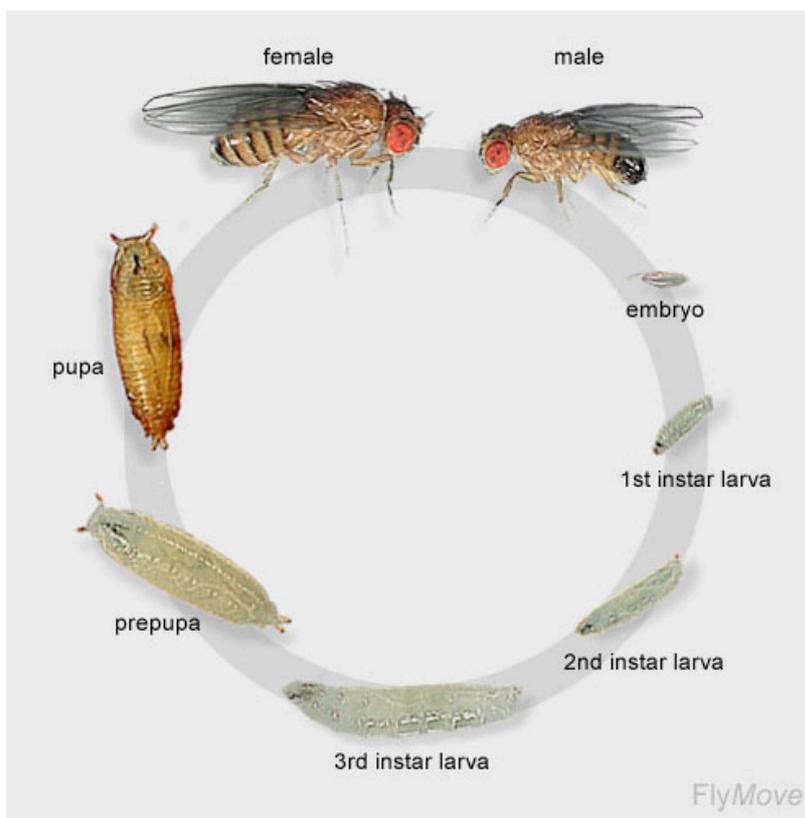
### **1.1.2 Die Larvalentwicklung**

Die larvalen Stadien dauern etwa 4-5 Tage, während derer sich die Larve zweimal häutet und an Größe und Gewicht zunimmt. Innerhalb der thorakalen und abdominalen Regionen der Larve sind bereits nach dem Schlüpfen kleine Zellgruppen erkennbar, die so genannten Imaginalscheiben. Sie wachsen während der gesamten larvalen Entwicklung und formen kleine epitheliale Säckchen, die undifferenzierte Vorläuferzellen adulter Gewebe darstellen. Am Ende des dritten Larvalstadiums beginnt die Larve, aus dem Futter herauszukriechen, um sich zu verpuppen, indem ihre Kutikula durch den Prozess der Apolyse die Hülle der Puppe bildet.

### 1.1.3 Die pupale Morphogenese

Unter hormoneller Kontrolle findet die pupale Metamorphose statt. Während dieser Stadien sterben die larvalen Zellen ab, und die adulten Organe und Gewebe bilden sich. Dabei differenzieren die Imaginalscheiben zu den Mundwerkzeugen, den Extremitäten, den Augen und Antennen sowie den Genitalien des Adulttieres. Auch Bereiche des dorsalen Thorax und des Kopfes entwickeln sich aus diesen Geweben.

Nach insgesamt 10 Tagen ist die Morphogenese abgeschlossen und das Adulttier schlüpft. Nach ungefähr acht bis zehn Stunden werden die Jungtiere fertil. Jedes Weibchen kann etwa 100 Eier pro Tag legen. Je nach Zuchtbedingungen leben die adulten Tiere etwa 50 Tage.



**Abb. 1 Der Lebenszyklus von *Drosophila melanogaster*.** Die embryonale Entwicklung und die ersten beiden Larvalstadien dauern je etwa einen Tag. Zum Ende des dritten Larvalstadiums klettert die Larve aus dem Futter, und bildet durch Ablösung der Kutikula die Hülle der Präpuppe. Innerhalb der folgenden vier bis fünf Tage bildet sich durch Absterben der larvalen Gewebe und Bildung von bereits in der Larve angelegten Organen der Imago. Nach etwa zehn Tagen schlüpft das Adulttier, und das Weibchen ist nach acht Stunden fertil (entnommen aus Weigmann et al. 2003)

## 1.2 Frühe Musterbildung in *Drosophila*

Die frühesten Gene, die an der Entwicklung beteiligt sind, werden benötigt, bevor der Embryo seine eigenen Gene exprimiert. Diese so genannten maternalen Gene werden während der Oogenese exprimiert und sorgen für eine korrekte Ausbildung des Embryos. Ihre Genprodukte bilden im Ei Gradienten, welche die Polarität und damit ein Muster für die weitere Entwicklung festlegen (Ingham 1988; St Johnston und Nüsslein-Volhard 1992). Dabei wird die Spezifizierung der antero-posterioren Achse von drei Systemen gesteuert, die die Lokalisation bestimmter Proteine und mRNAs in festgelegten Bereichen des Embryos kontrollieren, während die Festlegung der Dorsoventralachse dem Toll-Signalweg unterliegt. Jedes dieser Systeme wird bereits vor der Fertilisation festgelegt, und alle resultieren aus der Kommunikation zwischen der Oozyte und den umgebenden Follikelzellen (Ray und Schüpbach 1996).

Untersuchungen von larvalen Kutikulaphänotypen und mutanten Embryonen im Jahr 1980 führten zu der Annahme, daß drei verschiedene Klassen von zygotischen Genen für die Segmentierung des Embryos verantwortlich sind (Nüsslein-Volhard und Wieschaus 1980). So konnte im Rahmen dieser Untersuchungen gezeigt werden, daß Mutationen innerhalb der Lücken-Gene zum Verlust mehrerer zusammenhängender Segmente führen. Der Verlust von Paarregel-Genen resultiert in dem Fehlen jedes zweiten Segments, und einzelne Sektionen der Segmente sind in Mutationen der Segment-Polaritätsgene betroffen. Genauere Untersuchungen an diesen Gengruppen zeigten, daß sie in einer Kette von regulatorischen Prozessen fungieren, die den Embryo nach und nach in seine segmentalen Bereiche unterteilen (Akam 1987; Ingham 1988).

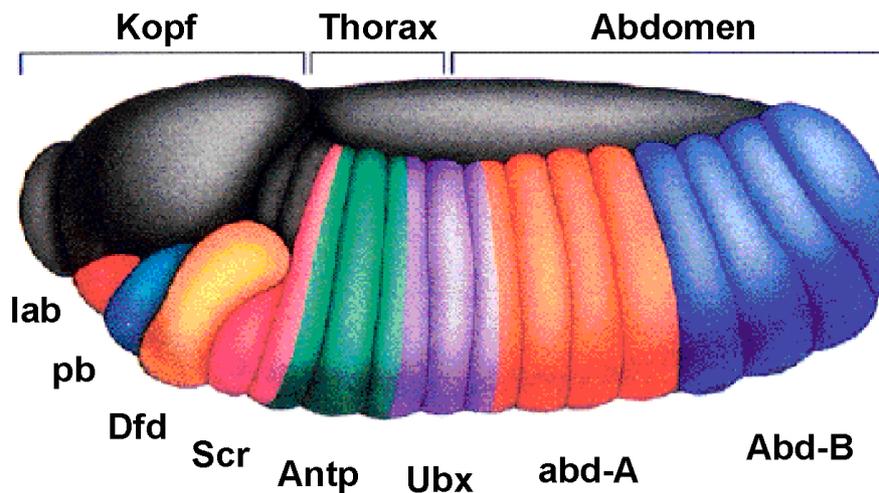
Die zygotischen Lücken-Gene sind in breiten, überlappenden Regionen im zentralen Bereich des Embryos exprimiert. Ihre Expression wird direkt von den maternalen Genen *bicoid* (*bcd*), *hunchback* (*hb*), *nanos* (*nos*) und *caudal* (*cad*) gesteuert (Gaul et al. 1987; Driever und Nüsslein-Volhard 1988; Tautz 1988; Driever und Nüsslein-Volhard 1989; Struhl 1989; Rivera-Pomar et al. 1995). Zusätzlich werden die räumlichen Grenzen der einzelnen Lücken-Gene durch Interaktionen mit den benachbarten Lücken-Genen definiert (Jäckle et al. 1986; Pankratz et al. 1989).

Paarregel-Gene werden in sieben oder acht Streifen entlang der antero-posterioren Achse des Embryos exprimiert. Sie werden in zwei Klassen aufgeteilt: die primären

Paarregel-Gene werden von einer Reihe von cis-regulatorischen Elementen in ihren Promotoren durch die Expression der maternalen Gene und der Lücken-Gene reguliert, während die sekundären Paarregel-Gene durch die Expression der primären Paarregel-Gene gesteuert werden (Ingham 1988). Jedes Element reguliert die Expression eines Streifens oder eines Teils dieses Streifens, und enthält eine bestimmte Anzahl an Aktivator- oder Repressor Bindungsstellen, die von den Transkriptionsfaktoren der maternalen oder Lücken-Gene erkannt werden (Howard et al. 1988; Goto et al. 1989; Harding et al. 1989; Pankratz et al. 1990). Zusammen mit den Lücken-Genen sind die Paarregel-Gene an der Kontrolle der Segmentierungs- und homöotischen Selektorgene beteiligt.

Die Expression der Segment-Polaritätsgene, der letzten Stufe der Segmentierungsgene, führt zu der Organisation des Embryos in 14 Segmente. Diese sind in engen Streifen in jedem Parasegment exprimiert und definieren einzelne Zellen in diesen Parasegmenten. Jedes einzelne Parasegment exprimiert eine bestimmte Anzahl an homöotischen Genen (Hox-Gene), welche die jeweilige Identität des Parasegments determinieren. Die fünf Hox-Gene des Antennapedia-Komplexes (*ANT-C*), *labial (lab)*, *proboscipedia (pb)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)* und *Antennapedia (Antp)*, werden für die Entwicklung der Kopf- und vorderen Thoraxsegmente benötigt, während die drei Hox-Gene des Bithorax-Komplexes (*BX-C*), *Ultrabithorax (Ubx)*, *abdominal-A (abd-A)* und *Abdominal-B (Abd-B)* für die Differenzierung der posterioren thorakalen und abdominalen Segmente verantwortlich sind (Kaufman et al. 1990).

Die jeweils spezifische Expression einer Anzahl an Hox-Genen in jedem Parasegment wird durch die gesamte Entwicklung benötigt, um die Identität der einzelnen Segmente zu definieren und aufrecht zu erhalten. In Abbildung 2 ist das spezifische Expressionsmuster der Hoxgene in den einzelnen Parasegmenten schematisch gezeigt.



**Abb. 2 Das Expressionsmuster der Hox-Gene.** Die Gene des Antennapedia Komplexes *labial*, *pb* und *Dfd* sind in einzelnen Segmenten im Kopf, *Scr* und *Antp* in zwei bzw. drei Parasegmenten des Thorax exprimiert. Die Hox-Gene des Bithorax Komplexes *Ubx*, *abd-A* und *Abd-B* sind für die korrekte Ausbildung der abdominalen Segmente verantwortlich (aus <http://www.blc.arizona.edu/courses/181gh/rick/development1/graphics/1717.gif>).

Da die frühen Expressionsmuster der Lücken- und der Paarregel-Gene jedoch nach etwa vier Stunden der Embryogenese verschwinden, werden weitere Regulatoren benötigt, um das Expressionsmuster der homöotischen Gene während der Entwicklung aufrecht zu erhalten.

### **1.3 Die Aufrechterhaltung eines Expressionsmusters während der Entwicklung**

#### **1.3.1 Das Konzept des zellulären Gedächtnisses**

Das Schicksal einer Zelle hängt von dem bereits früh angelegten Genexpressionsmuster ab. Ist dieses Muster einmal festgelegt, müssen bestimmte Mechanismen die Fortführung durch viele Zellteilungen hindurch vom Embryo bis zum Adulttier sichern.

Das zelluläre Gedächtnis bezeichnet den epigenetischen Prozess, der Zellen einer definierten Zelllinie ermöglicht, sich an ihr ursprüngliches Expressionsmuster zu „erinnern“ und dieses aufrecht zu erhalten. Im letzten Jahrzehnt konnte geklärt werden, daß die Chromatinstruktur den Transkriptionszustand eines Gens kontrollieren und diesen durch die DNA-Replikation und Mitose in die nächste Generation übernehmen kann. Die genauen Mechanismen, mit deren Hilfe die Chromatin-

struktur diese Kontrolle durch die Mitose hindurch beibehalten kann, sind aber noch nicht im Detail geklärt. Viele Studien konnten sogar zeigen, daß die Chromatinstruktur während der Mitose zu einer übergeordneten Chromatinstruktur kondensiert, und sich erst wieder in den Tochterzellen zurückbildet, um das Expressionsmuster wieder aufzubauen (Yamamoto et al. 1997; Buchenau et al. 1998; Dietzel et al. 1999). Aufgrund dieser Ergebnisse wird angenommen, daß Signale auf der Chromatinebene existieren, die während der Zellteilung verankert bleiben und auf die replizierten Chromosomen übertragen werden.

Veränderungen dieser Markierung würden zum Verlust des zellulären Gedächtnisses führen und damit zu einem anderen Zellschicksal, was drastische Konsequenzen für den sich entwickelnden Organismus hätte.

### 1.3.2 Die Proteine der Polycomb und trithorax Gruppen

Zwei Gruppen von Genen sind für die Aufrechterhaltung der Hox-Gen Expression nach der Gastrulation verantwortlich. Die Polycomb-Gruppe (PcG) von Genen ist für die Aufrechterhaltung des inaktiven Status verantwortlich (Jürgens 1985; Paro 1995), während die Trithorax-Gruppe (trxG) die Expression der Hox-Gene aufrechterhält (Kennison 1995). PcG und trxG Proteine sind ubiquitär exprimiert und müssen daher speziellen Mechanismen unterworfen sein, um ihre Zielgene während einer Zellteilung spezifisch, entweder im aktiven oder reprimierten Zustand zu halten.

#### 1.3.2.1 Die Polycomb Gruppe

Die ersten PcG Proteine, die aufgrund ihres homöotischen Phänotyps identifiziert wurden, waren *extra sex combs* (*esc*) und *Polycomb* (*Pc*) (Lewis, 1947; Slifer, 1942). Mutationen in diesen Genen führen zur Bildung von Geschlechtskämmen auf dem zweiten und dritten Beinpaar von *Drosophila* Männchen, die normalerweise nur auf dem ersten Beinpaar zu finden sind. Verursacht wird dieser Phänotyp durch den Verlust der korrekten Hox-Gen Expression in diesem Segment, so daß das betroffene Segment die Identität eines anderen bekommt (Kennison 1995; Kennison 2004). Durch molekulare Analysen konnten wichtige Erkenntnisse über den Mechanismus der Genregulation durch die PcG gewonnen werden. Obwohl die

identifizierten PcG Gene untereinander keine signifikanten Sequenzähnlichkeiten enthalten, verfügen sie alle über konservierte Proteinmotive, die potentielle Domänen für Protein-Protein Interaktionen darstellen. Weitere Experimente belegten später diese Theorie, indem sie zeigten, daß die PcG Proteine als multimeren Komplexe fungieren (DeCamillis et al. 1992; Franke et al. 1992).

In den letzten Jahren konnten unterschiedliche Komplexe der PcG isoliert werden. Der PRC1 Komplex hat eine Größe von 1-2 MDa und enthält die Proteine Pc, Polyhomeotic (Ph), Posterior Sex Combs (Psc) und Ring (Francis et al. 2001; Lavigne et al. 2004). Durch *in-vitro* Versuche konnte gezeigt werden, daß PRC1 in der Lage ist, Verschiebungen von Nukleosomen Remodeling zu verhindern (Shao et al. 1999). Dadurch könnte der Komplex an der Stabilisierung des Chromatins beteiligt sein. Allerdings konnte *in-vivo* keine Spezifität von DNA Sequenzen, die die Bindungsstellen der Polycomb Gruppe enthalten (PREs, Polycomb Response Elements), festgestellt werden (Simon et al. 1993; Chan et al. 1994). Im Gegensatz dazu binden die Genprodukte der Gene *trithorax-like (trl)* und *Pleiohomeotic (Pho)* an Sequenzen, die in PREs gefunden werden (Poux et al. 2001), und beide Proteine können mit den PcG Komplexen interagieren. Da *pho* ein Mitglied der PcG ist, wird angenommen, daß Pho die Funktion übernimmt, PcG Komplexe an die DNA zu binden und zu verankern (Fritsch et al. 1999). Vermutlich handelt es sich bei den Interaktionen zwischen diesen Proteinen und dem PRC1 Komplex um kurzfristige, schwache Bindungen, die bei der Aufreinigung der Komplexe verloren gehen.

Ein weiterer Komplex, der 600 kDa große PRC2 Komplex, konnte in den letzten Jahren isoliert werden (Tie et al. 2001; Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Müller et al. 2002; Tie et al. 2003). Er enthält die Produkte der Gene *Extra Sex Combs (esc)*, *Suppressor of zeste 12 (Su(z)12)* und *Enhancer of zeste (E(z))*, für das eine Histon-Methyltransferase-Aktivität nachgewiesen werden konnte (Jones et al. 1998; Cao et al. 2002; Czermin et al. 2002; Müller et al. 2002).

In verschiedenen Studien konnte festgestellt werden, daß bei der Aufreinigung der Komplexe auch andere Proteine, die eine Funktion in der Regulation der Genexpression haben, isoliert werden. Darunter sind Proteine, die in der Histonmodifizierung eine Rolle spielen, wie zum Beispiel die Histondeacetylase Rpd3, oder Transkriptionsfaktoren, wie TFIID. Dies legt eine direkte Verbindung zwischen

den PcG Proteinen und den histonmodifizierenden Enzymen oder Proteinen der Transkriptionsmaschinerie nahe (Breiling et al. 2001; Saurin et al. 2001; Tie et al. 2001). Insgesamt ist die PcG eine Familie, deren Komplexe auf vielen Ebenen der Genregulation beteiligt sind.

Protein Name	Gene	Komplex	Proteindomäne
Additional sex combs	<i>Asx</i>		WD40 repeats
Corto	<i>corto</i>		Chromodomäne
cramped	<i>crm</i>		PHD finger
DRING	<i>Sce</i>	PRC1	RING
Enhancer of Polycomb	<i>E(Pc)</i>	PRC2	
Enhancer of zeste	<i>E(z)</i>		SET, Histon-Methyltransferase
Extra sex combs	<i>esc</i>		
Lola like (batman)	<i>lolal</i>		BTB/POZ; (Faucheux et al. 2001)
Multi sex combs	<i>mxc</i>		WD40 repeats
Pho-like	<i>phol</i>		(Brown et al. 2003)
Pipsqueak	<i>psq</i>		(Jacobs und van Lohuizen 1999)
Pleiohomeotic	<i>pho</i>		Zinc finger (Fritsch et al. 1999)
Polycomb	<i>Pc</i>	PRC1	Carboxyl-terminale Domäne, Chromodomäne
Polycomb-like	<i>Pcl</i>		PHD finger
Polyhomeotic	<i>ph</i>	PRC1	SAM domain (Kim et al. 2002), zinc finger
Posterior sex combs	<i>Psc</i>	PRC1	RING
Sex comb on midleg	<i>Scm</i>	PRC1	SAM Domäne
Sex combs extra	<i>Sce</i>		
Suppressor of zeste (2)	<i>Su(z)2</i>		RING finger, HR Region
Suppressor of zeste 12	<i>Su(z)12</i>	E(z)/Esc	Zinc finger, VEPS box
Super sex combs (sxc)			WD40 repeats

**Tabelle 1 Die Mitglieder der PcG** (verändert nach Ringrose und Paro 2004).

### 1.3.2.2 Die Trithoraxgruppe

Die meisten *trxG* Gene wurden gefunden, da sie in der Lage sind, homöotische Phänotypen, wie sie bei *Pc*-Mutanten entstehen, zu unterdrücken. Die Tatsache, daß Mutationen in diesen Genen identisch zu solchen von Verlustmutationen der *Hox*-Gene sind (Shearn 1989; Ingham 1998), führte zu der Annahme, daß *trxG* die Aufgabe hat, die Expression der homöotischen Gene über den Zeitpunkt der Gastrulation hinaus stabil zu halten; eine Annahme, die wenige Jahre später bestätigt werden konnte (Breen und Harte 1991). Die *trxG* ist eine heterogene

Genfamilie, und einige ihrer Komplexe sind in generellen transkriptionellen Prozessen beteiligt. Ausnahmen bilden die Genprodukte von *trx* und *Ash1*, die eine spezifische Rolle in der Regulation an den PREs haben (Chinwalla et al. 1995; Rozovskaia et al. 1999).

Vier verschiedene Komplexe konnten für die *trxG* isoliert werden, und alle haben verschiedene, die Chromatinstruktur verändernde, Eigenschaften. Der BRM Komplex, ein etwa 2 MDa umfassender Komplex, der die *trxG* Proteine Brahma, Moira und Osa umfasst, zeigt hohe Homologie zu dem Hefe SWI/SNI Komplex. Dieser Komplex ist an der Regulation von Chromatinstrukturen beteiligt und scheint nicht mit spezifischen chromosomalen Regionen assoziiert zu sein (Peterson und Herskowitz 1992; Elfring et al. 1998).

Die Proteine *Ash1* und *Ash2* konnten in zwei weiteren Komplexen isoliert werden (Papoulas et al. 1998). *Ash1* war das erste Mitglied der *trxG* Familie, für das eine Methyltransferase-Aktivität gezeigt werden konnte (Beisel et al. 2002). Der vierte und letzte bisher bekannte *trxG* Komplex ist der so genannte TAC1 Komplex (Petruk et al. 2001), der das *Trx* Protein enthält.

Das *Trx*-Protein enthält drei Zinkfinger-ähnliche Domänen, die PHD-Finger (Mazo et al. 1990; Aasland und Stewart 1995), und interagiert vermutlich spezifisch mit den PREs. Darüber hinaus besitzt *Trx* mehrere konservierte Proteinmotive, die ebenfalls in anderen chromosomalen Proteinen gefunden wurden und wahrscheinlich an Veränderungen der Chromatinstruktur beteiligt sind.

Fluoreszenz-*in-situ*-Hybridisierungen an Polytänchromosomen sowie X-ChIP (Formaldehyde Crosslinking Chromatin Immunoprecipitation) Analysen zeigten, daß *trxG* und *PcG* Proteine an den gleichen, zellgedächtnis-vermittelnden, regulatorischen Sequenzen co-lokalisieren können (Chang et al. 1995; Chinwalla et al. 1995; Strutt et al. 1997; Orlando et al. 1998).

Zusätzlich zu den homöotischen Genen werden zahlreiche weitere Gene über die *PcG/trxG* Proteine kontrolliert. Viele Ansätze werden verwendet, um diese Gene zu identifizieren. Eine der Möglichkeiten, um das Gesamtgenom nach solchen Zielgenen zu untersuchen, ist die Expressionsanalyse mit Hilfe von DNA Microarrays.

Name	Gen	Komplex	Domäne oder Funktion
Trithorax	<i>trx</i>	TAC	Histon-Methyltransferase
Absent, small, or homeotic discs 1	<i>ash-1</i>	Ash-1	Histon-Methyltransferase
Absent, small, or homeotic discs 2	<i>ash-2</i>	Ash-2	PHD finger, SPRY Domäne
Brahma	<i>brm</i>	Brahma	ATPase Untereinheit
Osa	<i>osa</i>	Brahma	
Moira	<i>mor</i>	Brahma	
Zeste	<i>z</i>		
GAF (GAGA factor)	<i>Trl</i>		DNA bindendes Protein
Kismet	<i>kis</i>		DNA helicase, Chromodomäne, BRK
Tonalli	<i>tna</i>		SP-RING Zn Finger
breathless	<i>dev (btl)</i>		FGF-Rezeptor, Tyrosinkinase
Kohtalo	<i>kto</i>	Mediator Komplex (Homolog des TRAP240 Komplexes )	
Skuld	<i>skd</i>	Mediator Komplex (Homolog des TRAP240 Komplexes )	

**Tabelle 2 Die Mitglieder der *trxG*** (verändert nach Ringrose und Paro 2004).

## 1.4 DNA Microarrays

### 1.4.1 Herstellung von Microarrays mit vorab synthetisierter DNA

Die Methode, DNA Proben auf eine feste Oberfläche wie Nylon zu verankern, wurde bereits 1975 von Ed Southern entwickelt (Southern 1975). Radioaktiv markierte Proben wurden auf die hergestellten Membranen hybridisiert, um zu untersuchen, ob komplementäre DNAs zu den markierten Oligonukleotiden unter den fixierten Molekülen vorhanden waren. Später wurden Filter mit ganzen fixierten cDNA-Bibliotheken und sogar den Genomen kleiner Prokaryoten hergestellt, so daß erste genomweite Experimente durchgeführt werden konnten.

In den neunziger Jahren wurden dann nach und nach die Nylonfilter durch Glas ersetzt. Verschiedene Labore entwickelten Möglichkeiten, vorab synthetisierte DNAs auf Glasarrays zu fixieren (Schena et al. 1996; Shalon et al. 1996). Diese festen Oberflächen ermöglichten es, geringere Mengen der Probenmoleküle auf

einer kleineren Fläche zu fixieren. Die aufgetragenen Mengen je Spot lagen hier inzwischen im Pico- bis Nanoliterbereich, während sie bei den porösen Membranen um einige Zehnerpotenzen größer waren. Mitentscheidend für das Aufbringen solcher kleiner Flüssigkeitsmengen ist das Verhalten der unterschiedlichen Oberflächen. Im Gegensatz zu den Nylon- oder Zellulosemembranen saugen die festen Glasarrays keine Flüssigkeit auf. Allerdings muß der Glasarray mit einer Beschichtung versehen werden, damit die DNA-Proben optimal binden. Diese Beschichtungen besitzen dabei hauptsächlich silanisierte Aminogruppen, die über das Silan kovalent an die Oberfläche des Glases gebunden werden. Die positive Ladung der Aminogruppe auf der Glasoberfläche ermöglicht dann die Bindung des negativ-geladenen Rückgrades der DNA. Weitere Oberflächenbeschichtungen sind silanisierte Epoxide und Aldehyde, über welche die DNA kovalent an den Array gebunden wird.

Gespottet werden zumeist cDNAs, PCR-Produkte oder Oligonukleotide. Die DNA wird in 384-well Platten bereitgestellt, und mit Hilfe von einem Roboter in einem vorgegebenen Muster auf den Array gespottet. Diese Methode ist flexibel und die verwendete DNA kann jederzeit variiert werden. Daher wird diese Methode hauptsächlich in Laboren verwendet, wo die Microarrays je nach Versuchsaufbau variiert werden sollen.

#### **1.4.2 *In-situ* synthetisierte Oligonukleotid-Arrays**

Firmen bedienen sich meistens einer anderen Technik, in der die Proben nicht im Vorfeld, sondern *in-situ* direkt auf den Array synthetisiert werden. Die erste der für diesen Ansatz verwendete Methoden war die fotolithografische Methode (Fodor et al. 1991) und wurde zuerst von der Firma Affymetrix Inc. vermarktet. Dabei werden zunächst spezifische Oligonukleotide für jedes Fragment, das auf dem Array vertreten sein soll, anhand ihrer Hybridisierungseigenschaften definiert. Für das Design einer fotolithografischen Maske werden Computeralgorithmen verwendet. Ausgewählte Bereiche einer mit lichtempfindlichen Linkermolekülen versehenen Glasoberfläche werden durch Löcher in der Maske belichtet, und die Linkermoleküle dadurch an diesen Stellen aktiviert. Die Oberfläche wird mit dem ersten Nukleotid überschichtet und kann an den belichteten Stellen über eine foto-

chemische Bindung synthetisiert werden. Der Array wird danach gewaschen, um die nicht gebundenen Nukleotide zu entfernen. Dieser Vorgang wird für die drei weiteren Nukleotide wiederholt, bis an jeder Stelle des Arrays, an der eine DNA synthetisiert werden soll, das erste Nukleotid bindet. Jedes dieser Nukleotide ist über eine Phosphoramitid-Kappe geschützt. Im zweiten Schritt werden wiederum durch eine Maske die entsprechenden Stellen durch Licht angeregt und die Phosphoramitid-Kappen dadurch entfernt, und die zweite Base hinzugefügt. Dieser Prozess wird so lange durchgeführt, bis alle Oligonukleotide die berechnete Länge erreicht haben. Diese Art der Produktion ist aufgrund der verwendeten Masken sehr unflexibel und eignet sich nur für die Herstellung sehr großer Mengen an Arrays. Andere Firmen konnten z.B. durch Modifizierung der piezoelektrischen Druckprozesse, ähnlich einem Tintenstrahldrucker, oder durch die Verwendung von digitalen Lichtprozessoren, wie sie in Beamern verwendet werden und die aus vielen kleinen Spiegeln bestehen, die gleichzeitig Licht auf bestimmte Stellen lenken können, andere Arten der *in-situ* Probensynthese entwickeln.

### **1.5 Genvorhersagen in *Drosophila***

Um genomweite Studien mittels Microarrays zu machen, ist es zunächst notwendig herauszufinden, welche Bereiche auf der DNA einer Spezies Gene kodieren. Die Genomsequenzen vieler verschiedener Arten, so auch die von *Drosophila melanogaster*, wurden in den letzten Jahren vollständig sequenziert und stellen so die Voraussetzung dar, Microarrays zu entwickeln, die alle transkribierten Bereiche enthalten.

Obwohl sich die Sequenzen noch in einem Prozess der Reevaluierung und Maturierung befinden, so sind sie doch so ausgereift, daß Genvorhersagen getroffen werden können. Die Probleme einer Genvorhersage liegen auf der Hand: Gene selbst sind auf der DNA diskontinuierlich, d.h. immer wieder von Introns unterbrochen. Obwohl die theoretischen Begrenzungen zwischen Exon und Intron durch spleißen sowie des Genanfangs durch den Promoter bekannt sind, so sind sie doch in *in-silico* Genvorhersage nur begrenzt zu verwenden. So haben zum Beispiel ca. 70% der Promoterregionen eine TATA-Box, die verbleibenden 30% jedoch nicht, wodurch man die TATA-Box selbst nicht als Signal für einen Promoter nehmen kann (Baxevanis 2001). Ein weiteres Problem ist die Tatsache, daß ein offenes

Leseraster zwar benötigt aber nicht ausreichend ist, um eine Region als Exon vorherzusagen. Um diese Probleme zu umgehen, bedienen sich Genvorhersagen verschiedener Methoden. Eine solche Methode, die unter anderem von der Vorhersagesoftware fgenesh verwendet wird, ist die „discriminant analysis“. Mit Hilfe dieser Methode kann man verschiedene Informationen über Sequenzabschnitte zusammenfassen und danach über einen dynamischen Algorithmus entscheiden, ob dieser Abschnitt zum Beispiel ein Exon ist oder nicht. Andere Vorhersageprogramme, wie zum Beispiel Genscan, verwenden das „probabilistic model“, über dessen Algorithmus die Wahrscheinlichkeit berechnet wird, mit der eine Sequenz ein Exon, einen Promoter oder ähnliches enthält (Claverie 1997; Claverie et al. 1997). Zusätzlich zu solchen Methoden werden Genvorhersagen oft nur dann als gesichert angesehen und verwendet, wenn es mehrere Hinweise auf das Vorhandensein des entsprechenden Gens gibt. So sind die Identifikation eines ESTs (Expressed Sequence Tag), Homologien zu anderen Proteinen, Konserviertheit der Sequenz zu anderen Organismen oder eine Mindestlänge der als Gen identifizierten Sequenz eine Möglichkeit, die Anzahl der falsch vorhergesagten Gene gering zu halten. Durch unterschiedliche Auswahlkriterien kommt es vor, daß verschiedene Genvorhersagen, die auf unterschiedlichen Methoden beruhen, zwar auf eine ähnliche Gesamtzahl von Genen kommen, diese Annotationen aber wenig Übereinstimmung miteinander haben (Hogenesch et al. 2001; Daly 2002). Die Vorhersagen für das Drosophila-Genom beziehen sich in der Zwischenzeit nur noch zu 6% auf rein bioinformatische Vorhersagen (Misra et al. 2002). Über einen solchen Ansatz können bekannte Gene bzw. solche mit großer Ähnlichkeit zu bekannten Proteinfamilien mit einer großen Sicherheit vorhergesagt werden. Gene, die keiner bekannten Proteinfamilie zugehörig sind, oder solche, die nicht hoch konserviert sind und über die keine weiteren Daten vorliegen, werden dagegen mit hoher Wahrscheinlichkeit übersehen.

### **1.6 Ziel dieser Arbeit**

Mehrere Studien belegen, daß viele Gene der Regulation der PcG/trxG unterliegen. Während einzelne dieser Gene durch biologische Ansätze identifiziert werden konnten, sind doch die meisten bisher unbekannt. Ziel dieser Arbeit war es daher, Methoden zu entwickeln, die eine genomweite Analyse dieser Zielgene ermöglichen. Dieser Ansatz beinhaltete die Entwicklung eines *Drosophila* Microarrays, der einen möglichst kompletten Satz an Genen enthält. Um die beschriebenen Restriktionen der Genvorhersagen zumindest teilweise zu überwinden, sollte die publizierte Genannotation von *Drosophila* mit weiteren, weniger stringenten Methoden zu einer neuen Genkollektion kombiniert, und auf dieser Basis ein Microarray produziert werden. Nachfolgende biologische Validierungen der in diesen Experimenten erzielten Ergebnisse sollten zum einen dazu dienen, die Qualität des Microarrays zu überprüfen, zum anderen sollte die Existenz eventuell neu vorhergesagter Gene bestätigt werden.

Im weiteren Verlauf der Arbeit sollte der Microarray dazu verwendet werden, Zielgene Polycomb-abhängiger Regulation zu identifizieren. Vergleiche von wildtypischer cDNA mit solcher mutanter Embryonen, die kein funktionsfähiges Pc-Protein besitzen, sollten auf den neu entwickelten Microarrays durchgeführt werden, um beispielhaft erste Hinweise auf Zielgene für Pc in *Drosophila melanogaster* zu finden. Zusammenfassend stellt diese Arbeit die methodische Grundlage dar, um mit Hilfe eines neuen DNA Microarrays genomweit nach Genen zu suchen, die epigenetischen Regulationsmechanismen unterliegen.

## 2 Ergebnisse

### 2.1 Die Herstellung eines PCR basierten DNA Transkriptom-Microarrays

Durch die Sequenzierungsprojekte der letzten Jahre konnten die Genome von *Drosophila melanogaster*, Maus, Mensch sowie vieler weiterer Vertebraten, Invertebraten und Prokaryoten entschlüsselt werden. Diese Sequenzinformation ist die Grundvoraussetzung für die vollständige Vorhersage der Gene des jeweiligen Organismus und eröffnet damit die Möglichkeit, das Zusammenspiel von Genen genomweit zu erforschen. Die Komplexität einer genomweiten Analyse erfordert die Verwendung neuer, auf hohen Durchsatz optimierter Methoden. Daher stellen DNA-Microarrays die geeignete Methode dar, um diese Netzwerke der Genexpression zu untersuchen.

#### 2.1.1 Die Genkollektion für den HeidelbergFlyArray

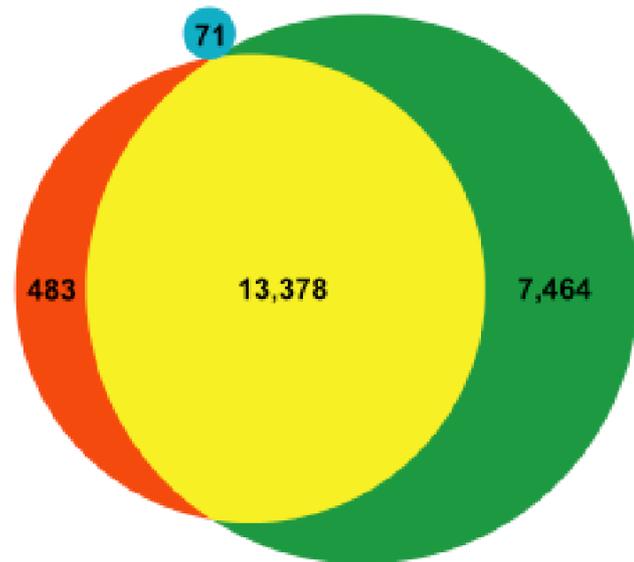
Der erste Teil dieser Arbeit bestand in der Herstellung eines PCR basierenden Transkriptom-Microarrays, der eine möglichst komplette Anzahl an Genen enthalten sollte. Zunächst wurden dafür die Genvorhersagen der Version 2 des Berkeley *Drosophila* Genom Projektes (BDGP) mit der weniger stringenten Vorhersage der fgenesh Software sowie den Genen der Berkeley *Drosophila* Gene Collection (Stapleton et al. 2002) vereint.

Das Berkeley *Drosophila* Genome Project beinhaltet zusammen mit der *Drosophila* Gene Collection (DGC) 13,861 vorhergesagte Gene, während die fgenesh Software 20,622 Gene vorhersagt. Unter Berücksichtigung der beschriebenen Überschneidungskriterien (siehe 5.6.1) ergibt die Fusion beider Vorhersagen 21,325 Gene, von denen 13,377 von beiden Programmen vorhergesagt wurden. 483 Vorhersagen entstammen nur der Berkeley-Annotation, während 7,464 Vorhersagen nur von fgenesh gemacht wurden (Abb. 3).

Zusätzlich zu diesen 21,325 Genen wurden weitere 71 Gene, die zu dem Zeitpunkt nicht in der Genomsequenz beinhaltet waren, manuell dem Satz hinzugefügt. Daraus ergibt sich die neue HeidelbergCollection mit insgesamt 21,396 potentiellen Genen. Im Folgenden werden alle Gene, die bereits von Berkeley vorhergesagt wurden, mit

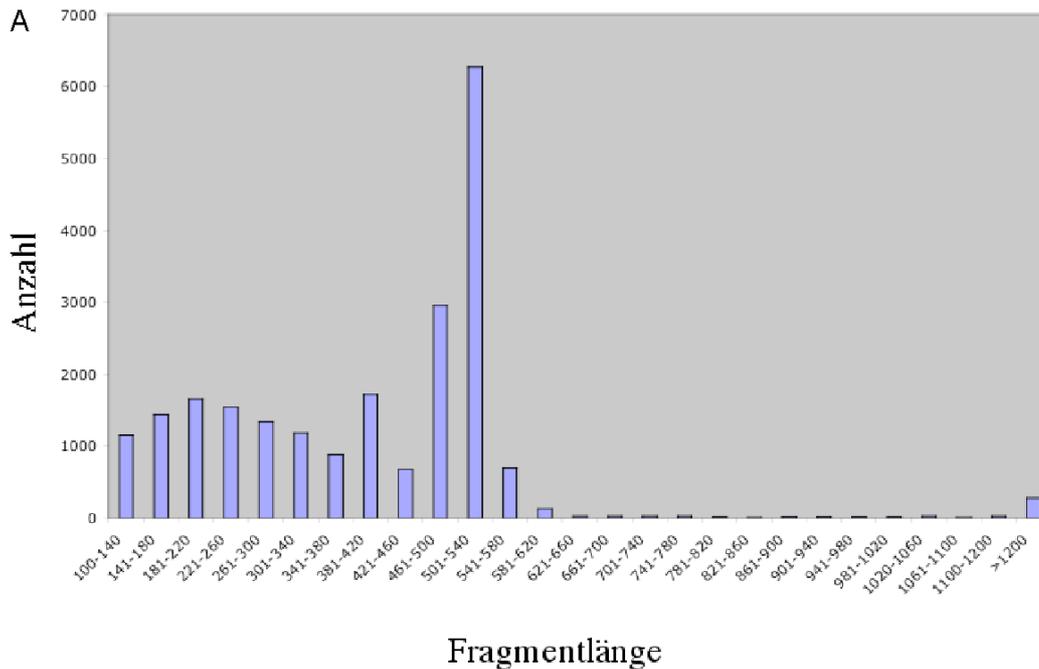
dem in der Flybase verwendeten Namen (z.B. CG12369) bezeichnet werden, und für die 7,464 potentiell neuen Gene wird die Nummer der HeidelbergCollection (z.B. HDC06000) verwendet werden.

**Abb. 3 Die Heidelberg Annotation.** Die Kombination der BDGP cDNA Collection R1 mit der BDGP Genannotation enthalten zusammen 13,861 Gene (rot/gelb). In der Heidelberg Annotation sind insgesamt 20,622 Gene enthalten (gelb/grün). Ausgehend von der Annahme, daß zwei Gene, die mit mehr als 30% der exonischen Bereiche überlappen das gleiche Gen darstellen, wurden beide Annotationen kombiniert. Zusätzlich wurden 71 Gene, deren Sequenz nicht in der Genomsequenz enthalten war, manuell dem Satz an Genen hinzugefügt (blau). Die resultierende Heidelberg Annotation enthält 21,396 potentielle Gene.

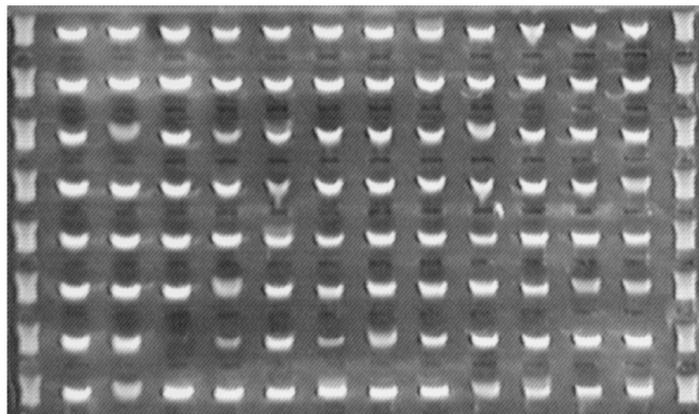


### 2.1.2 Das Primerdesign und die Synthese der Fragmente für den HeidelbergFlyArray

Zur Herstellung des HeidelbergFlyArrays wurde für jedes der 21,396 vorhergesagten Gene ein Primerpaar für die PCR synthetisiert. Mit Hilfe der GenomePride Software konnten spezifische Primer für 21,306 der vorhergesagten Gene berechnet werden. Die Primer wurden so gewählt, daß Homologieregionen und Introns vermieden wurden und das Amplikon auf einem mittleren Exon lag. Aufgrund der unterschiedlichen Länge der einzelnen Exons und der stringenten Kriterien für die Platzierung der Primer konnte die angestrebte Fragmentlänge von 500 Basenpaaren nicht immer eingehalten werden; insgesamt beträgt der Median der Fragmentlängen daher 456 bp (Abb.4, A). Für insgesamt 90 vorhergesagte Gene konnten aufgrund repetitiver Sequenzen bzw. nicht ausreichender Länge der einzelnen Exons keine Primer berechnet werden.



**B**



**Abb. 4 Die Größenverteilung der PCR Fragmente und die Auswertung der PCRs.** Mit der GenomePride Software werden Primer unter der Bedingung berechnet, daß für die entstehenden PCR Fragmente Introns, Homologieregionen und Repeat-Sequenzen vermieden werden. Die optimale Länge der Fragmente beträgt dabei 500bp, wobei Abweichungen durch die Länge der Exons, bzw. durch die Beschaffenheit der Sequenz verursacht werden (A). Die durchschnittliche Länge der Fragmente beträgt 408bp. Mittels Gelelektrophorese auf 1%igen Agarosegelen wurden die PCR-Fragmente auf Anwesenheit und Länge überprüft (B, Platte 973130).

Die Synthese der PCR-Produkte erfolgte in zwei Schritten, um mögliche Kontamination mit genomischer DNA auszuschließen und über die Verwendung von Tag-Primern eine einfache Reamplifikation der Produkte zu ermöglichen (siehe auch 5.4.3). Die erste PCR erfolgte in Kollaboration mit der Gruppe Hoheisel am DKFZ,

die zweite PCR wurde sowohl am DKFZ als auch am ZMBH für unterschiedliche Applikationen durchgeführt. Beide PCR Schritte wurden mittels Gelelektrophorese auf ihre Produktlänge überprüft (Abb.4, B), und nicht erfolgreiche PCRs wurden einzeln wiederholt. Insgesamt konnten 20,948 (97,9%) Fragmente in der PCR amplifiziert werden.

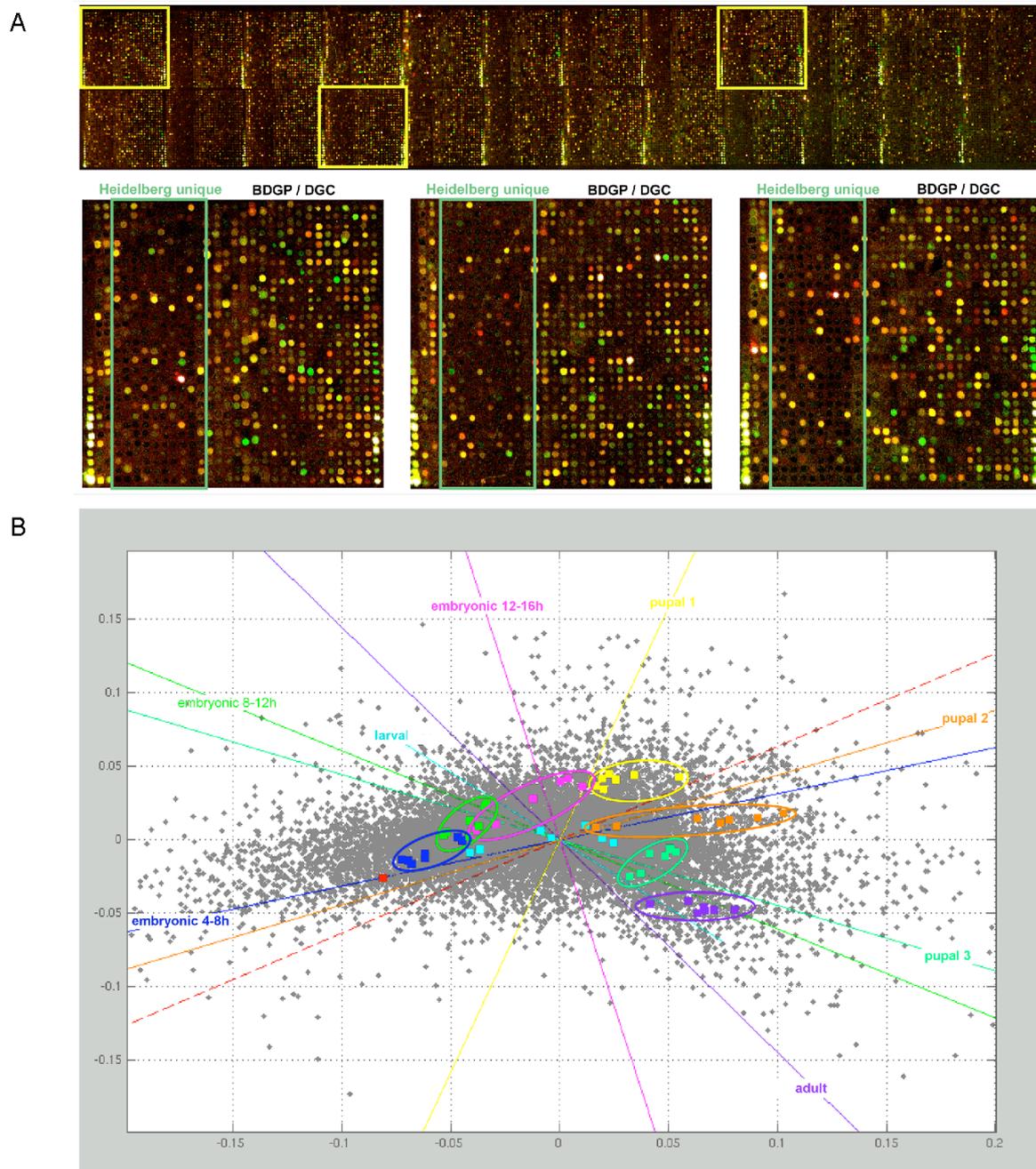
Jeweils die Hälfte der PCR Produkte wurde eingetrocknet, in Spottingpuffer aufgenommen und mit dem Biogrid II auf Aminosilan-Glaslides gespottet (siehe 5.4.7). Dabei wurden die Gene, die ausschließlich von fgenesh vorhergesagt wurden, in abgegrenzten Bereichen gespottet.

## **2.2 Das Entwicklungsprofil von *Drosophila melanogaster***

Der gewählte Ansatz zur Erstellung einer neuen Genkollektion beinhaltet den Vergleich zweier unterschiedlich stringenter Genvorhersagen. Die vom BDGP verwendeten Genvorhersageprogramme beruhen auf einem rein informatischen Ansatz und müssen dementsprechend sehr stringente Kriterien genügen, um falsch vorhergesagte Gene weitestgehend auszuschließen. Die in der Heidelberg Collection mit fgenesh vorhergesagten Gene unterliegen weit weniger stringente Kriterien, was die Wahrscheinlichkeit, daß Gene falsch vorhergesagt wurden, erhöht. Der Vorteil der Heidelberg Collection liegt darin, einen möglichst kompletten Satz an Genen auf einem Microarray zu repräsentieren. Allerdings ist die ausschließlich informatische Vorhersage bei diesen Kriterien nicht ausreichend, um zu entscheiden, welche Gene tatsächlich existieren und welche falsch vorhergesagt wurden. Daher sollten die auf dem Transkriptomarray enthaltenen Gene in einem biologischen Ansatz validiert werden.

Ausgehend von der Annahme, daß jedes Gen während des Lebenszyklus von *Drosophila* mindestens einmal exprimiert wird, wurde ein Entwicklungsprofil von *Drosophila* erstellt, das Aufschluß über die Zahl der exprimierten Gene geben sollte. Dazu wurde der Lebenszyklus von *Drosophila* in neun Stadien aufgeteilt und jedes Stadium mit dem HeidelbergFlyArray auf seine spezifische Genexpression untersucht. Für alle Stadien wurden Zweifarbenexperimente durchgeführt, wobei das Embryonalstadium 0-4h als Kontrollstadium für die acht weiteren Stadien diente. Für jedes der acht Stadien wurden mindestens drei unabhängige Wiederholungen durchgeführt. Aufgrund der unterschiedlichen Einbaurate der beiden Farbstoffe

wurde in wenigstens einer der Wiederholungen ein Farbstofftausch (Dye-swap) vorgenommen.



**Abb. 5 Hybridisierung des HeidelbergFlyArray.** In einem Testarray wurden indirekt markierte cDNAs von 4-8h alten Embryonen und Adulttieren auf den Microarray hybridisiert (A). Die gelben Quadrate in A) sind in darunter vergrößert dargestellt. Bereits in diesen beiden Stadien kann man in den Bereichen, in denen die neu vorhergesagten Gene liegen, in etwa 20% der Spots Expression erkennen (grüne Rechtecke in A). Alle Experimente wurden mit Hilfe der Korrespondenzanalyse auf ihre Qualität überprüft (B). Alle Wiederholungen der einzelnen Stadien sind farbig dargestellt und liegen eng zusammen, wobei larvale Experimente eine Ausnahme bilden (türkise Rechtecke ohne farbige Umrandung).

Nach eingehender Datenprozessierung konnte mit Hilfe der Microarrays die Expression von 9,908 Genen der BDGP Vorhersage sowie 2,636 neu identifizierter Gene während des gesamten Lebenszyklus von *Drosophila* nachgewiesen werden. Als Qualitätskontrolle der Hybridisierungen dienten dabei die Ergebnisse der Korrespondenzanalyse. Alle Wiederholungsexperimente wurden in dieser Analyse miteinander verglichen, wobei das Clustering der einzelnen Stadien in abgegrenzte Bereiche die Reproduzierbarkeit der Hybridisierungen widerspiegelt. Die gute Korrelation der einzelnen Hybridisierungen für ein Stadium zeigt eine zufriedenstellende Qualität der Experimente. Nur die Experimente mit larvaler cDNA variieren voneinander (Abb. 5, B). Dies kann darauf zurückzuführen sein, daß die verschiedenen Larvenstadien für diese Experimente gemischt wurden, und Larven im L1 Stadium eine ähnliche Expression wie Adultstadien aufweisen, wogegen die späteren Larvalstadien eher mit der Expression in Embryonen korrelieren (Arbeitman et al. 2002). Ein leicht unterschiedliches Mischverhältnis der larvalen RNAs kann also zu der beobachteten Diskrepanz in den einzelnen Hybridisierungen geführt haben.

### **2.3 Validierung der Microarrayanalysen mittels RT-PCR**

Obwohl die Tatsache, daß 73,7% der BDGP Gene auf unserem Microarray detektiert werden, gut mit veröffentlichten Daten der Sensitivität von Microarrayexperimenten übereinstimmt (Shoemaker et al. 2001), und die Filterkriterien der Arrayauswertung sehr stringent waren, sollten weitere biologische Validierungen Aufschluß über mögliche falsch-positive Gene geben. Dazu wurden RT-PCR Analysen von einer ausgewählten Anzahl an Genen gemacht. Aus dem originalen Primersatz wurden mehrere Platten, in denen die Anzahl an neu vorhergesagten Genen sehr hoch war, ausgewählt. Da die Amplikons in den Platten entsprechend ihrer Lage auf dem Chromosom sortiert sind, war die Auswahl nur bedingt zufällig. Die RT Reaktion wurde mit RNA aus einem gleichen Mischungsverhältnis aller Entwicklungsstadien durchgeführt. Die Ergebnisse der RT-PCRs können die Resultate der Microarrayexperimente für 98% (136/145) der ausgewählten BDGP Gene bestätigen. Ebenso konnte für 74,4% (218/293) der neu vorhergesagten Gene, die auf dem Microarray exprimiert waren, über die RT-PCR eine Expression bestätigt werden. Diese Ergebnisse deuten auf die Existenz von ~2,000 neuen Genen in *Drosophila* hin. Allerdings ist nicht auszuschließen, daß aufgrund des verwendeten RNA-

Gemischtes Gene, die sehr schwach oder nur während einer kurzen Zeitspanne exprimiert sind, nicht nachweisbar sind.

#### **2.4 Weitere Validierung und Analyse mittels *in-situ* Hybridisierung**

Während RT-PCRs Aufschluß über die Expression eines Gens geben, kann man über *in-situ* Hybridisierungen an wildtypischen Embryonen von *Drosophila* Informationen über das Expressionsmuster bekommen und damit Rückschlüsse auf eine mögliche Funktion ziehen. Des Weiteren kann man über Vergleiche der Expressionsmuster feststellen, ob benachbarte bekannte Gene das gleiche Muster aufweisen wie neu vorhergesagte, und so Aussagen über die Wahrscheinlichkeit machen, ob diese Genmodelle dasselbe Gen repräsentieren.

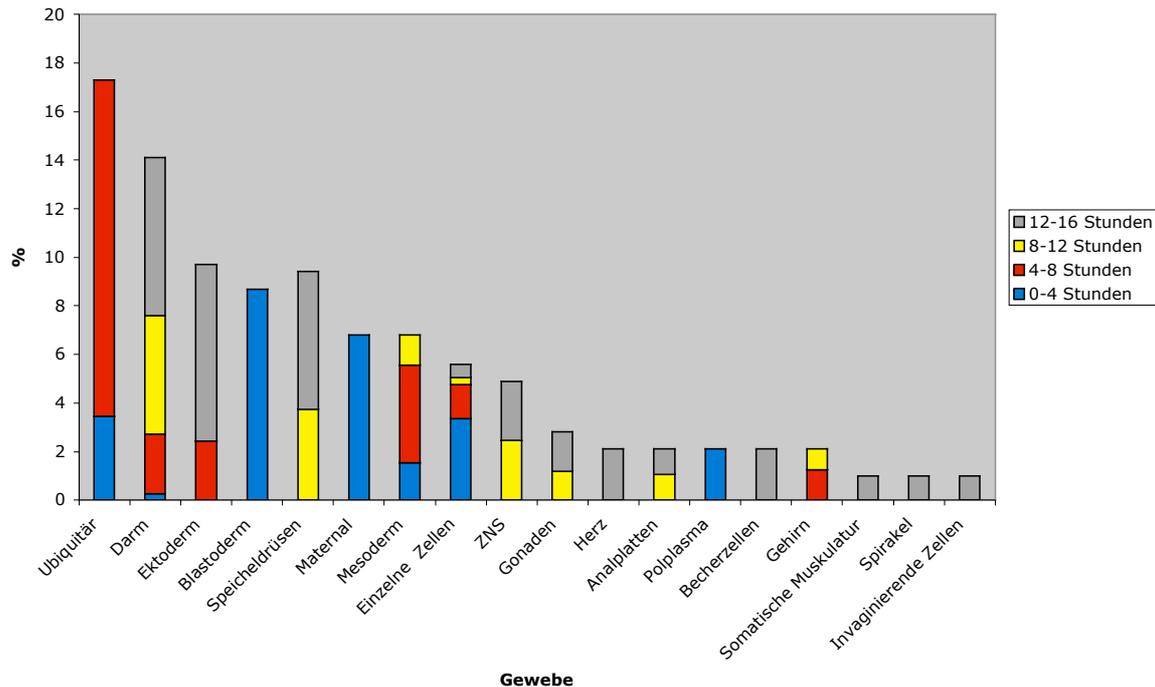
Es wurden von einer Anzahl der neuen Gene *antisense*-RNA Proben hergestellt und für *in-situ* Hybridisierungen verwendet. Es wurden nur Gene ausgewählt, die auf dem Microarray in embryonalen Stadien exprimiert waren. Für die Herstellung der antisense-RNA wurden die gleichen Fragmente verwendet, die auch als Proben auf den Microarray gespottet wurden. 213 der neuen Gene wurden über *in-situ* Hybridisierungen getestet, wobei für 82 (38,5%) ein Expressionsmuster detektiert werden konnte. Kontrollexperimente mit bekannten Genen ergaben die gleiche Erfolgsrate. Da die durchschnittliche Länge der PCR Fragmente 405 Basenpaare beträgt ist es möglich, daß diese für eine Hybridisierung sehr kurzen Proben zu der geringen Erfolgsrate geführt haben. Auch schwache oder ubiquitäre Expressionen lassen sich mit dieser Methode schwer nachweisen, und könnten daher als negativ gewertet werden. Um letzteres zu überprüfen, wurden von einigen Genen, die in der Hybridisierung als nicht exprimiert eingestuft wurden, RT-PCRs gemacht. Tatsächlich konnte hierdurch die Expression für 14 der 17 getesteten Gene nachgewiesen werden (Tabelle 3).

Gencode	Peptidlänge (Anzahl der Aminosäuren)	Fragmentlänge (Anzahl der Basenpaare)
HDC00784	184	519
HDC00804	159	331
HDC00888	136	373
HDC01523	179	506
HDC01644	193	511
HDC02683	180	377
HDC02718	210	484
HDC02755	111	280
HDC02901	565	510
HDC03202	132	304
HDC03292	151	388
HDC03331	51	703
HDC03334	51	419
HDC03347	181	315

**Tabelle 3** Gene, deren Expression über RT-PCR, nicht aber über *in-situ* Hybridisierungen nachweisbar war. Dargestellt sind die Gennamen, die vorhergesagte Peptidlänge und die Länge des Fragments, das für die *in-situ* Hybridisierung verwendet wurde.

Die neuen Gene, für die ein Expressionsmuster beobachtet werden konnte, wurden dahingehend untersucht, ob für sie eine bevorzugte Expression in bestimmten Geweben oder zu bestimmten Zeitpunkten erkennbar ist. Die Analyse ergab, daß dies nicht der Fall ist. Die Expression kann zu vielen verschiedenen Zeitpunkten und in den meisten embryonalen Geweben beobachtet werden (Abb. 6).

Eine große Anzahl der neuen Gene ist in den frühen Embryonalstadien ubiquitär oder in einzelnen Keimblättern exprimiert, so daß für diese Gene nicht auf eine Funktion geschlossen werden kann. In älteren Stadien kann man jedoch die Expression einiger neuer Gene in differenzierten Geweben oder einzelnen Regionen erkennen, wie zum Beispiel im zentralen Nervensystem, den Gonaden oder den Anallplatten. Eine Rolle der neuen Gene für die Entwicklung oder Funktion dieser Organe ist damit naheliegend.



**Abb. 6 Die Häufigkeit der Expression in verschiedenen Geweben zu unterschiedlichen Zeitpunkten der Embryogenese.** Über *in-situ* Hybridisierung an wildtypischen Embryonen wurde eine Auswahl der neuen Gene auf ihr Expressionsmuster untersucht. Die Gene sind in vielen Geweben und zu unterschiedlichen Zeitpunkten exprimiert. Die Anzahl der beobachteten Expression innerhalb eines Gewebes in Relation zu der Expression in allen Geweben ist in Prozent angegeben. Gene, die in unterschiedlichen Geweben exprimiert werden, wurden mehrfach berücksichtigt.

Einige der neu vorhergesagten Gene wurden im Verlauf der Arbeit näher untersucht. Dazu wurden die Gene, die über *in-situ* Hybridisierungen validiert werden konnten, entsprechend ihrer Lage zu bereits bekannten Genen, in drei Gruppen unterteilt.

Gene der 1. Gruppe – überlappen mit exonischen Bereichen bekannter Gene  
und/oder – liegen in der gleichen Transkriptionsrichtung in Introns

Gene der 2. Gruppe – liegen in der gleichen Transkriptionsrichtung innerhalb von  
10 kb zu anderen Genen

Gene der 3. Gruppe – liegen mehr als 10 kb von anderen Genen entfernt  
oder – liegen in entgegengesetzter Transkriptionsrichtung zu be-  
nachbarten Genen

## 2.5 Neue Gene, die mit bekannten Genen assoziiert sind

Die Gene der ersten Gruppe überlappen mit anderen Genen oder liegen in Introns anderer Gene, jedoch mit weniger als 30% der exonischen Bereiche, da ansonsten die von uns definierten Kriterien für ein eigenes Gen nicht erfüllt wurden (Tabelle 4).

Name	Genbank-ID	Chrom	Lokalisation	Synonym	Expression	Evidenz
HDC00966	BK003326	2L	Intron CG11030	-	zelluläres Blastoderm	-
HDC02494	BK003424	2L	Intron CG15288	wing blister (wb)	Mesoderm- anlage	<i>Dp</i>
HDC02764	BK003493	2L	Intron CG4838	-	zelluläres Blastoderm	<i>Dp</i> , EST
HDC03057	BK003539	2L	Intron CG5803 überlappend	Fasciclin3 (fas3)	maternal, Blastoderm	<i>Dp</i> , EST
HDC03282	-	2L	Intron CG10443	Leukocyte-anti- gen-related-like (Lar)	maternal	-
HDC06000	BK003754	2R	Intron CG12369	Lachesin (Lac)	Zelluläres Blastoderm	<i>Dp</i>
HDC07276	-	2R	Intron CG15112 überlappend	Enabled (ena)	Zellen im Kopfbereich	-
HDC09414	-		Intron CG32057	-	schwach, Mesoderm- anlage	-
HDC10019	BK002122	3L	Intron CG10741	-	Speicheldrü- sensvorläufer	-
HDC10028	BK002123	3L	Intron CG12478	bruno-3	ZNS	<i>Dp</i>
HDC10120	BK002139	3L	Intron CG17697	frizzled	Mesoderm- anlage	-
HDC03960	BK003614	2R	CG17921	-	Mesoderm- anlage	<i>Dp</i>
HDC14231	BK002626	3R	überlappend TE19396	-	maternal, Speicheldrü- sensvorläufer, Speichel- drüsen	-
HDC13283	-	3R	Intron CG14390	Beat-VC	Streifen im Blastoderm, Ektoderm	-
HDC14493	BK002672	3R	Intron CG31175	Dystrophin	Blutgefäße	<i>Ag Dp</i>
HDC18326	BK003102	X	Intron CG1691	-	schwach, ubiquitär	<i>Dp</i>

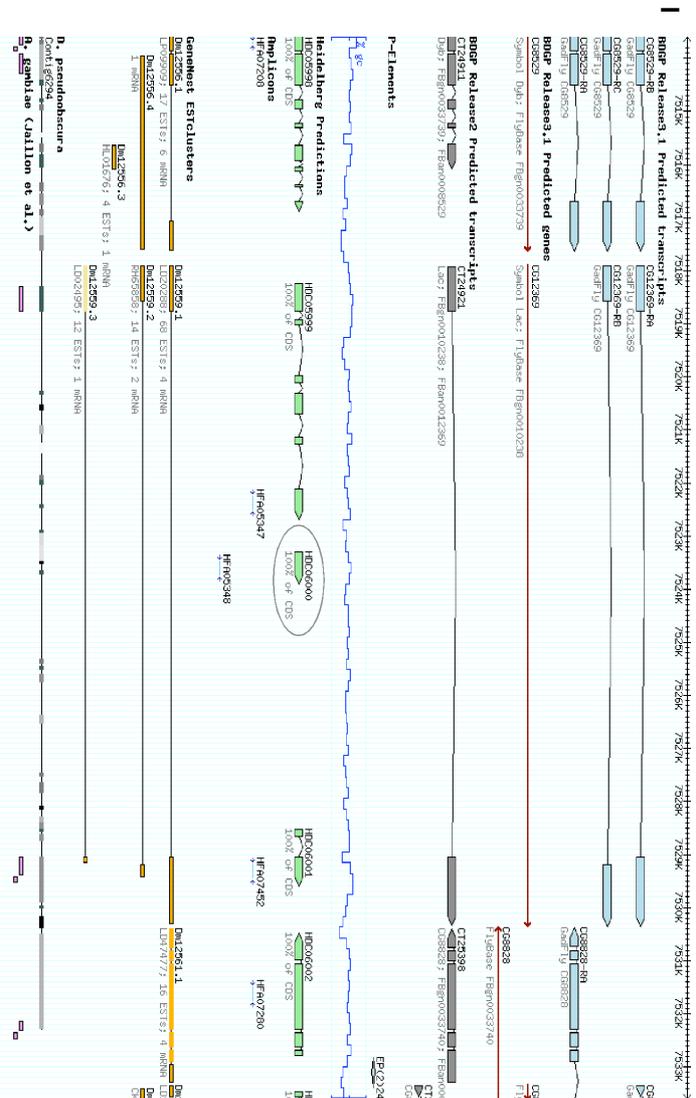
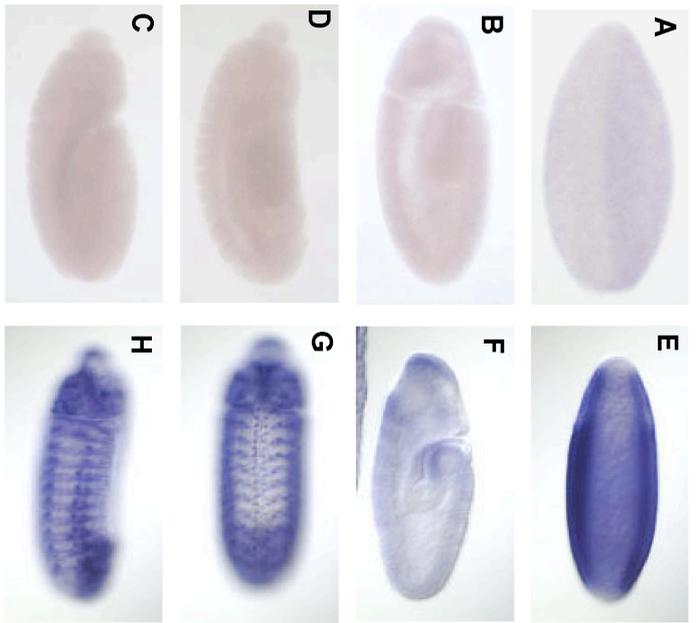
**Tabelle 4 Gene, die mit anderen Genen assoziiert sind.** Die Gennamen, die zugehörigen Genbank IDs sowie die Lokalisation auf dem Chromosom und die Lage in Relation zu bekannten Genen sowie die Synonyme sind dargestellt. Homologiebereiche zu *Drosophila pseudoobscura* (*Dp*) oder *Anopheles gambiae* (*Ag*), die zusätzliche Hinweise auf eine Funktion dieses genomischen Bereiches sind, sind ebenfalls aufgeführt, ebenso wie in diesen Regionen liegende ESTs. Mögliche Überlappung mit exonischen Bereichen bekannter Gene beträgt weniger als 30%.

Die Wahrscheinlichkeit, daß es sich bei solchen Genen um nur ein Gen mit eventuell alternativen Exons handelt, erscheint relativ hoch. Über Vergleiche des Expressionsmusters mit den benachbarten Genen sowie über die Ergebnisse der Microarrayexperimente wurden einige dieser Gene näher validiert und charakterisiert.

### **2.5.1 HDC06000, ein zusätzliches Exon von *Lachesin*?**

HDC06000 ist auf dem rechten Arm des 2. Chromosoms lokalisiert. Es ist mit einer Peptidlänge von 197 Aminosäuren vorhergesagt und liegt im einzigen Intron des Gens *Lachesin* (CG12369), ohne mit exonischen Regionen dieses Gens zu überlappen (Abb. 7, I). *Lachesin*, ein Mitglied der Immunoglobulin Superfamilie zeigt frühe mesodermale Expression sowie Expression in den Neuroblasten des sich entwickelnden Nervensystems (Karlstrom et al. 1993). des Weiteren wurde kürzlich die Rolle von *Lachesin* im trachealen System beschrieben (Llimargas et al. 2004). Vergleicht man die Expression von HDC06000 in *in-situ* Hybridisierungen mit den veröffentlichten Bildern von *Lachesin in-situs*, so kann man eine identische Expression in frühen Embryonen beobachten. In beiden Färbungen ist eine starke, ubiquitäre Expression in jungen Embryonen sichtbar, wobei jeweils die ventralen und die ventrolateralen Bereiche von der Expression ausgeschlossen sind. In späteren Stadien kann man in Embryonen, die mit der HDC06000 Probe hybridisiert wurden, keine Expression mehr detektieren (Abb. 7, B-D), wohingegen in den Embryonen, die mit der *Lachesin* Probe hybridisiert wurden, deutliche Expression im Nervensystem und in den Tracheen sichtbar ist (Abb. 7, F-H).

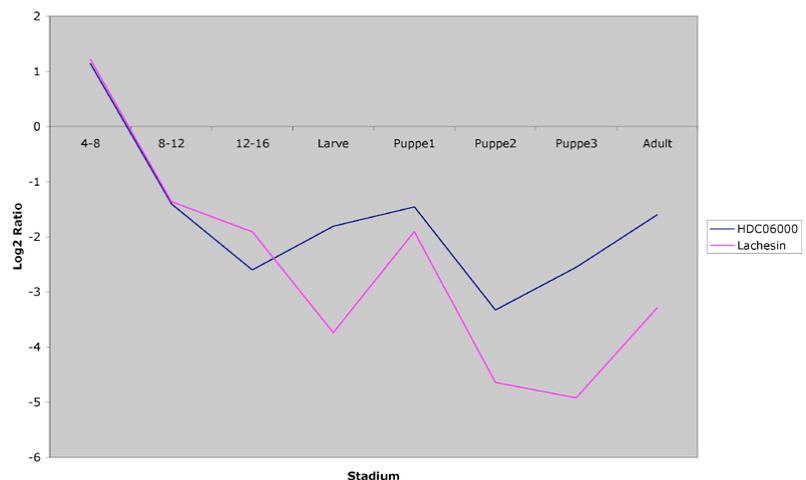
Untersuchungen des Gens HDC05999, welches ebenfalls innerhalb des gleichen Introns vorhergesagt wurde, zeigen das gleiche Expressionsmuster wie HDC06000 (Daten nicht gezeigt). Aufgrund der ähnlichen Expression in Embryonen ergibt sich die Frage, ob es sich bei HDC06000 um ein zusätzliches Exon von *Lachesin* handelt, oder ob es ein unabhängiges, eventuell über den gleichen Promoter reguliertes Gen ist.



**Abb. 7 Vergleich der Expression der Gene HDCC06000 und Lachesin.** *In-situ* Hybridisierungen mit einer RNA-Probe gegen HDCC06000 zeigen deutliche Expression im zellulären Blastoderm (A), wogegen in späteren Stadien keine Expression mehr beobachtet werden kann (B-D). HDCC06000 ist im einzigen Intron des *Lachesin* (*Lac*) Gens lokalisiert und zeigt Homologien zu *Drosophila pseudoobscura* (1). *In-situ* Hybridisierungen gegen eine Probe des *Lac* Gens zeigen identische Expression zu HDCC06000 in frühen Stadien; in späteren Stadien ist eine starke Expression in Teilen des zentralen und peripheren Nervensystems sowie in den Tracheen zu beobachten (F-H). Abbildungen der *Lac* Expression (E-H) entstammen der Berkeley Embryonic Expression Pattern Webseite (Tomancak et al. 2002). Die Darstellung der 20 kb flankierenden Region des Gens sind der HDFlyArray Webseite entnommen (hdflyarray.zmbh.uni-heidelberg.de).

Der Vergleich der Microarray-basierenden Entwicklungsprofile beider Gene kann Aufschluß über Unterschiede in der Expression während der gesamten Entwicklung geben. Sind HDC06000 und *Lachesin* während des Lebenszyklus identisch reguliert, kann dies ein Hinweis darauf sein, daß es sich um das gleiche Gen handelt. Tatsächlich erkennt man bei der Analyse der Entwicklungsprofile, daß *Lachesin* und HDC06000 sehr ähnlich reguliert sind (Abb. 8). Während die Expressionsmuster in der Larvalentwicklung leicht variieren, sind sie in der embryonalen und der pupalen Entwicklung sowie im Adultstadium sehr ähnlich. Insgesamt können die Microarraydaten durch die geringen Unterschiede in der Expression keinen eindeutigen Hinweis darauf geben, ob es sich bei HDC06000 und *Lachesin* um zwei benachbarte Gene oder um alternative Formen desselben Gens handelt.

**Abb. 8 Entwicklungsprofile von HDC06000 und *Lachesin*.** Über die gesamte Entwicklung bleiben die Entwicklungsprofile beider Gene sehr ähnlich. Eine leichte Abweichung ist nur in der larvalen Entwicklung zu beobachten. Insgesamt können über den Vergleich der Entwicklungsprofile keine eindeutigen Aussagen über die Abhängigkeit der Gene voneinander gemacht werden.



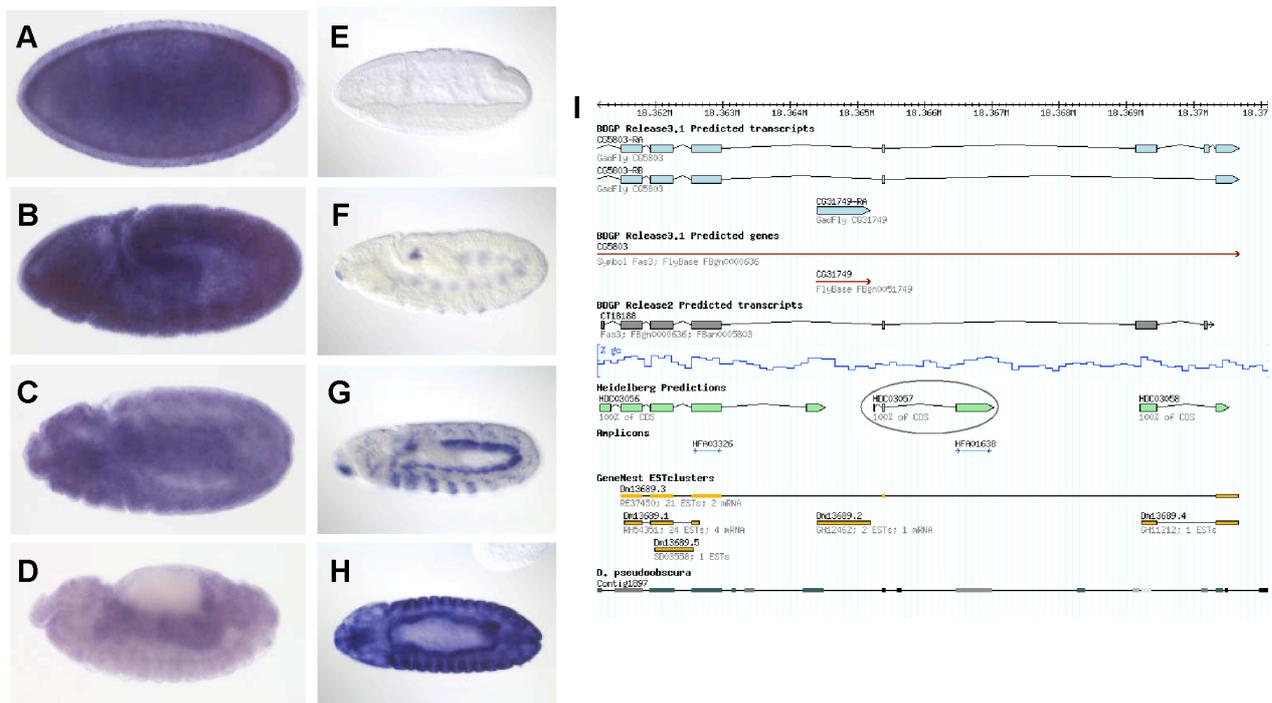
Für *Lachesin* existieren 94 ESTs sowie 6 volle Längen cDNAs (Abb. 7, I). Es ist mit einer langen (CG12369-RA) und einer kurzen Form (CG12360-RB) vorhergesagt. In keiner dieser Sequenzen ist ein zusätzliches Exon präsent, was gegen HDC06000 oder HDC05999 als zusätzliche Exons spricht. Allerdings stammt ein Großteil der in der Literatur verwendeten ESTs aus einer einzigen embryonalen cDNA Bibliothek, so daß man nur bedingt Aussagen über die Genstruktur aus diesen Daten machen darf. Handelt es sich bei HDC06000 um ein alternatives Exon von *Lachesin*, so sollte man es mit geeigneten Primern in einer PCR auf cDNA aus verschiedenen Stadien nachweisen können. Daher wurde RNA aus den in 5.4.8 beschriebenen 16 Stadien isoliert und für RT-PCRs verwendet. Der 5'-Primer für die RT-PCRs lag auf dem

ersten Exon von *Lac*, während der 3'-Primer am 3'-Ende von HDC06000 lokalisiert war. Tatsächlich konnte mit diesen Primerpaar das erwartete 500bp Fragment erzeugt werden. Die bisherigen Untersuchungen sprechen insgesamt dafür, daß es sich bei HDC06000 nicht um ein eigenes, bisher nicht vorhergesagtes Gen, sondern um eine alternative, bislang nicht beschriebene Form des *Lac* Gens handelt.

### 2.5.2 HDC03057 ist im Intron von *Fasciclin3* lokalisiert

HDC03057 wurde auf dem Chromosom 2L innerhalb eines Introns des Gens *Fasciclin3* (*Fas3*) vorhergesagt. Es hat eine vorhergesagte Länge von 606 Basenpaaren, und besteht aus zwei kurzen und einem längeren Exon, wobei das zweite, etwa 30 Basenpaare umfassende Exon von HDC03057 komplett mit einem mittleren Exon von *Fas3* überlappt. Die genomische Sequenz ist größtenteils in *Drosophila pseudoobscura* konserviert (Abb. 9, I), ein weiterer Hinweis auf die Existenz dieses Gens. Des Weiteren kann man das 515 bp lange Fragment, das auf dem Microarray gespottet wurde, in einer RT-PCR mit cDNA aus embryonalen Stadien erfolgreich amplifizieren (Abb. 10 B).

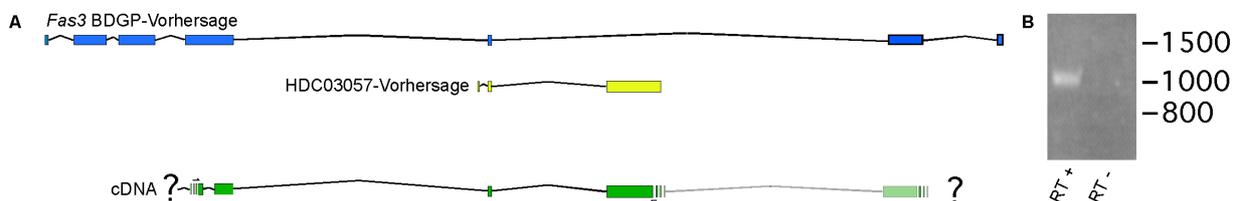
Interessanterweise kann man in *in-situ* Hybridisierungen mit einer Probe gegen HDC03057 keine Ähnlichkeit zu einer Hybridisierung gegen *Fas3* erkennen. HDC03057 zeigt starke maternale, ubiquitäre Expression, die bis zum Stadium 10 konstant bleibt und dann langsam bis zum Stadium 16 hin abnimmt (Abb. 9, B-D). Im Gegensatz dazu zeigen *in-situ* Hybridisierungen mit einer RNA-Probe gegen *Fas3* keine maternale Komponente. Erst ab Stadium 9 kann die erste Expression von *Fas3* beobachtet werden. Von diesem Stadium an bis zum Stadium 17 kann man in dieser Hybridisierung Expression in der embryonalen viszeralen Muskulatur, in einem Teil der Neuronen sowie in den polaren Follikelzellen beobachten (Abb. 9, G & H). Eine Funktion von *Fas3* ist unter anderem für die korrekte Axonausbildung sowie für die Erkennung der synaptischen Ziele beschrieben (Snow et al. 1989).



**Abb. 9** Vergleiche der Expression der Gene *HDC03057* und *Fas3*. Während *HDC03057* eine starke maternale Komponente und ubiquitäre Expression zeigt (A-D), kann man *Fas3* Expression in spezifischen Zellen des Nervensystems ab Stadium 10 (F), sowie in der viszeralen Muskulatur und in polaren Follikelzellen sehen (F-H). In früheren Embryonen bis zur Gastrulation kann keine Expression beobachtet werden (E). *HDC03057* liegt in einem Intron von *Fas3* in der Nähe eines weiteren, bisher nicht charakterisierten Gens, *CG91749* (I). Der größte Teil des dritten Exons von *HDC03057* ist in *Drosophila pseudoobscura* konserviert (I). Embryonen wurden von lateral (A-G) oder dorsal (H) aufgenommen, anterior ist links. Abbildungen von Embryonen, die mit einer antisense-Probe gegen *Fas3* hybridisiert wurden, stammen vom Berkeley Gene Expression Project (Tomancak et al. 2002). Die Darstellung der 20 kb flankierenden Region des Gens mit den BDGP Transkript Vorhersagen der Version 2 (grau) und 3.1 (blau) sowie der entsprechenden Genvorhersage (rot), der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplicons (blaue Linien), vorhandenen ESTs und cDNAs sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) sind der HDFlyArray Webseite entnommen ([hdflyarray.zmbh.uni-heidelberg.de](http://hdflyarray.zmbh.uni-heidelberg.de)).

Auch für die Gene *Fas3* und *HDC03057* wurde untersucht, ob es sich um alternative Produkte eines Gens handelt. Eine Mischung der oben beschriebenen cDNAs wurde als Template für eine RT-PCR mit dem 5'-Primer von *Fas3* und dem 3'-Primer von *HDC03057* verwendet. Aufgrund des erhaltenen PCR Produktes kann auch hier davon ausgegangen werden, daß es sich bei *HDC03057* um eine alternative Spliceform von *Fas3* handelt. Da *Fas3* ein sehr gut untersuchtes Gen ist und diese Variante bisher nicht beschrieben wurde, wurde das erhaltene PCR Produkt in den pCR2-TOPO Vektor kloniert und sequenziert, um die genaue Exon-Exon Grenze dieser Variante zu analysieren. *Fas3* und *HDC03057* überlappen in dem zweiten, 32 Basenpaare umfassenden Exon von *HDC03057* (Abb. 9, I und Abb. 10). Dieses Exon kommt in allen Varianten vor, wobei die folgenden Exons durch alternatives

Spleißen voneinander abweichen. An dieser Stelle kommt durch alternatives Spleißen auch die Variante mit HDC03057 zustande (Abb. 10). Allem Anschein nach handelt sich also auch bei HDC03057 um eine alternative, bisher nicht vorhergesagte Form eines bekannten Gens. Die Tatsache, daß die vorhergesagte Proteinsequenz beider Gene im Leseraster bleibt, stützt diese Vermutung. Im Gegensatz zu den oben beschriebenen Varianten von *Lachesin* kann hier aufgrund des unterschiedlichen Expressionsmusters während der Embryonalentwicklung davon ausgegangen werden, daß diese alternative Form eine andere Funktion als die bisher beschriebenen Varianten ausübt.



**Abb. 10 Überprüfung der Genmodelle von *Fas3* und *HDC03057*.** Das vorhergesagte Genmodell von *Fas3* (blau) ist durch die Existenz von cDNAs bestätigt. Eine RT-PCR mit Primern, die Exons beider Vorhersagen (blau und gelb) abdecken und anschließende Sequenzierung des Produktes belegte darüberhinaus die Existenz einer weiteren Spleißvariante (grün), welche das neu vorhergesagte Gen *HDC03057* mit einschließt. Das genaue Genmodell dieser alternativen Form wurde nicht untersucht.

## 2.6 Neue Gene, die in der Nähe bekannter Gene liegen

Unter den Genen, die über die *in-situ* Hybridisierungen auf ihr Expressionsmuster untersucht wurden, lagen 24 weniger als 10 kb von bekannten Genen entfernt (Tabelle 5). Drei dieser Gene wurden näher auf ihre Expression und ihre eventuelle Funktion überprüft.

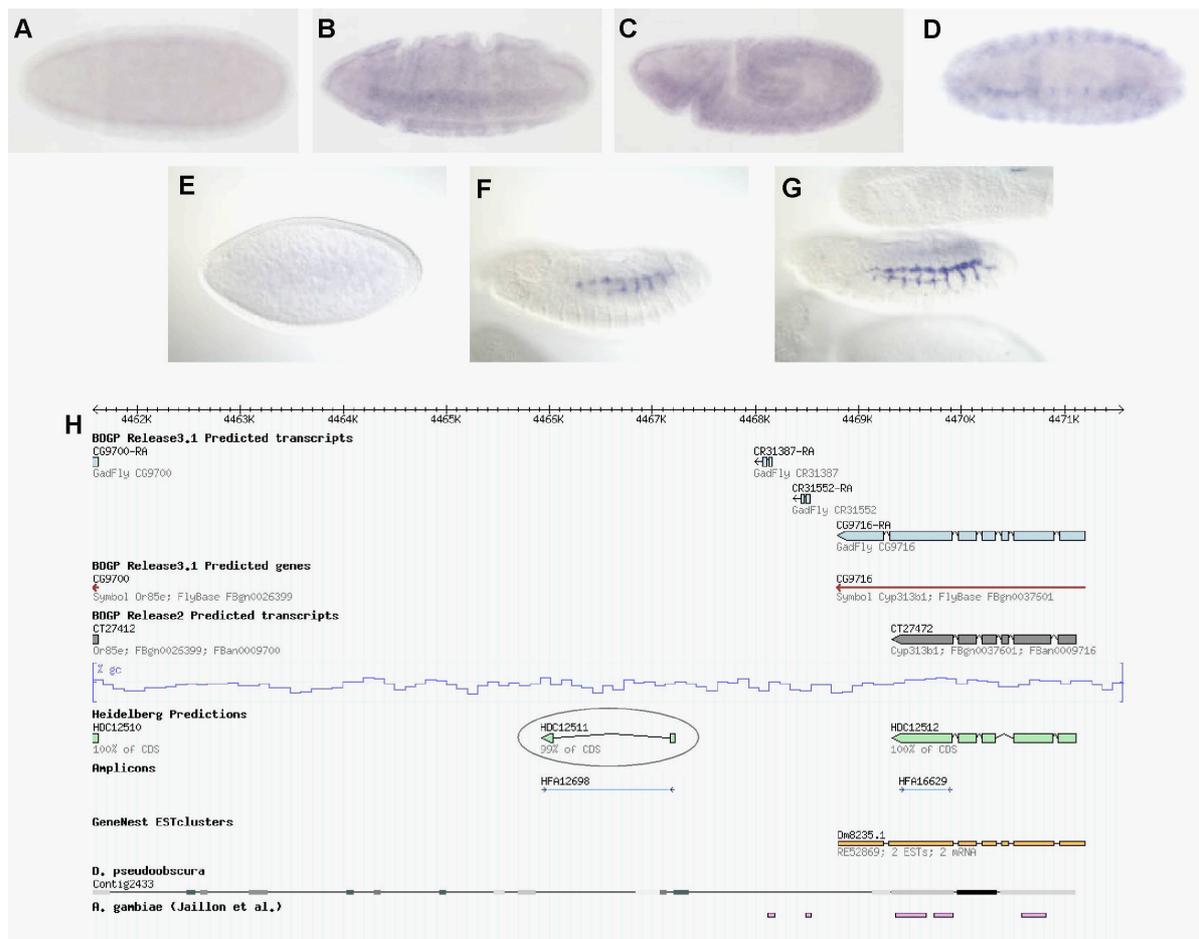
Genname	Genbank-ID	Chrom.	ben. Gen	Expression	Weitere Evidenz
HDC00027	BK003260	2L	CG11620	Ektoderm	<i>Ag</i>
HDC02527	BK003429	2L	CR31977	Speicheldrüsen	<i>Dp</i>
HDC05573	BK003699	2R	CR30009	Mitteldarm, ZNS	<i>Dp</i> ,EST
HDC04256	BK003630	2R	CG3092	Mesoderm, Gonaden	<i>Dp</i>
HDC05693	-	2R	CG7722	ZNS	
HDC06241	BK003785	2R	CG17716	ventrale Ektoderanlage, Mesoderanlage	<i>Dp</i>
HDC07651	-	2R	TE18887	Teile des Blastoderms	EST
HDC08265	BK001908	3L	CG1921/ CG10840	Verschiedene Zellen	-
HDC09080	BK002002	3L	CG32356	Speicheldrüsen	<i>Dp</i>
HDC10292	BK002156	3L	CG7945/ CG7973	zweite Welle von Neuroblasten	<i>Ag</i>
HDC11908	BK002321	3R	CG10609	ZNS	<i>Dp</i>
HDC12511	BK002404	3R	CG9700/ CG9716	Ektoderm	<i>Dp</i>
HDC12925	BK002446	3R	CG4509/ CG4565	schwach, ubiquitär 4-8h	EST
HDC15681	BK002831	3R	CG14548	schwach, ubiquitär 4-8h	<i>Dp</i> ,EST
HDC16874	BK002959	X	CG14635	Dotterkerne, anteriore Endoderanlage, Mitteldarm	<i>Ag</i>
HDC16879	BK002961	X	CG11664/ CG14633	invaginierende Zellen	-
HDC17351	BK003012	X	CG15576	Darm	-
HDC17498	-	X	CG4209/ CG10706	schwach, ubiquitär 0-4h	<i>Dp</i>
HDC17958	-	X	CG32716	Darm	<i>Ag</i>
HDC18148	BK003079	X	CG12650	schwach, ubiquitär 4-8h	-
HDC18410	BK003108	X	CG15198	schwach, ubiquitär 4-8h	-
HDC19378	BK003172	3R	CG4599	schwach, ubiquitär 4-8h	<i>Dp</i>

**Tabelle 5 Neu vorhergesagte Gene, die in der Nähe bekannter Gene lokalisiert sind.** Gennamen, die zugehörigen Genbank IDs, die Chromosomenlokalisierung und die benachbarten Gene sind dargestellt. Die beobachteten Expressionsmuster, Homologie zu *Drosophila pseudoobscura* (*Dp*) und *Anopheles gambiae* (*Ag*), sowie vorhandene cDNAs und ESTs geben weitere Hinweise auf die Existenz der potentiellen Gene.

## 2.7 HDC12511 zeigt ektodermale Expression

HDC12511 liegt auf dem rechten Arm des dritten Chromosoms, stromabwärts von Cyp313b1. Es hat weder konservierte Domänen noch zeigt es Homologien zu *Drosophila pseudoobscura* oder *Anopheles gambiae*. In den öffentlichen Datenbanken sind keine Hinweise auf cDNAs oder ESTs aus dieser Region vorhanden. Für das in 5'-Richtung liegende Gen Cyp313b1 existieren 2 ESTs sowie 7 cDNAs. Dies und die Tatsache, daß dieses Gen in der Literatur molekular untersucht wurde (Tijet et al. 2001) sprechen dafür, daß die Genstruktur für Cyp313b1 korrekt ist. Cyp313b1 kodiert ein Protein, welches eine Funktion im Elektronentransport hat und mög-

licherweise am Steroidmetabolismus beteiligt ist. Die Aminosäuresequenz des Gens enthält zwei E-class P450 Domänen. Die Expression wurde durch das Berkeley Gene Expression Projekt untersucht. Dabei wurde die Expression dieses Gens im Tracheensystem, besonders in den viszeralen und dorsalen Ästen der Tracheen beschrieben (Abb. 11, F-G), wogegen in den Stadien 1-12 keine Expression beobachtet wurde.



**Abb. 11 Expression von HDC12511 ist in anderen Geweben lokalisiert als die des benachbarten Gens Cyp313b1.** In frühen Stadien ist für dieses Gen keine bzw. nur eine sehr schwache Expression in der Mesodermanlage zu beobachten (B und C). Ab dem Stadium 13 ist HDC12511 jedoch in der embryonalen/larvalen Muskulatur exprimiert (D). Das in 5'-gelegene Gen Cyp313b1 (I) zeigt Expression im trachealen System (F und G); in jüngeren Embryonen ist keine Expression erkennbar (E). Die Embryonen wurden von lateral (A-C, E-G) oder ventral (D) aufgenommen, anterior ist links. Abbildungen F-H entstammen dem Berkeley Expression Projekt (Tomancak et al. 2002). Die Darstellung der 20 kb flankierenden Region des Gens mit den BDGP Transkript Vorhersagen der Version 2 (grau) und 3 (blau), der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplikons (blaue Linien), vorhandenen ESTs und cDNAs sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) und *Anopheles gambiae* (pink) sind der HDFlyArray Webseite entnommen (hdflyarray.zmbh.uni-heidelberg.de).

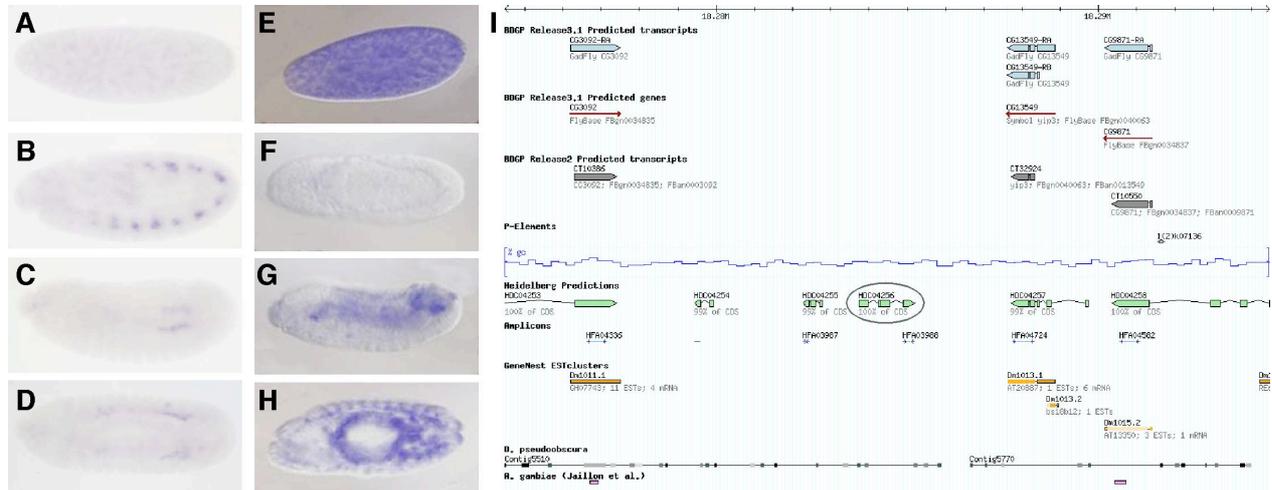
Die mRNA von HDC12511 wird in der somatischen Muskulatur und deren Vorläuferzellen ab dem Stadium 13 exprimiert (Abb. 11, D). In früheren Stadien ab der Gastrulation scheint das Gen in der Mesodermanlage exprimiert zu sein, allerdings ist das erkennbare Signal sehr schwach (Abb. 11, B und C). Die unterschiedliche Expression dieser Gene impliziert, daß es sich um zwei unabhängige Gene handelt. Die fehlende Konserviertheit des Gens kann ein Hinweis darauf sein, daß HDC12511 eine Funktion ausübt, die spezifisch für *Drosophila melanogaster* ist.

### 2.7.1 Expression von HDC04256 ist im Mesoderm lokalisiert

Anders als HDC03960 liegt HDC04256 nicht in direkter Nachbarschaft zu anderen bekannten Genen. 7 kb stromaufwärts, in der zytologischen Region 75E4, liegt das Gen CG3092, welches mit einer Länge von 5,171 Nukleotiden vorhergesagt wurde und eine Domäne für eine Acyl-CoA Dehydrogenase aufweist. 11 ESTs und vier cDNAs decken die Region von CG3092 vollständig ab, und die Vorhersage des Genmodells hat sich in den letzten beiden Versionen des Berkeley Drosophila Genome Projektes nicht verändert (Abb. 12, I). Genetische oder molekulare Studien von CG3092 sind nicht veröffentlicht, dennoch kann man die Genvorhersage aufgrund der cDNAs, der Anzahl an ESTs und der bekannten Proteindomäne als gesichert ansehen.

Der Vergleich der Expression von HDC04256 und CG3092 zeigt kaum Übereinstimmungen. Für HDC04256 kann man zuerst im Stadium 10 eine Expression in distinkten Zellen im Mesoderm erkennen (Abb. 12, A-B). Ab Stadium 12 erkennt man die Gonaden, und in späteren Stadien sind die Gonaden und der Gonadenschafft deutlich sichtbar (Abb. 12, C-D). Im Gegensatz dazu zeigt CG3092 maternale Expression (Abb. 12, E), die bereits im Stadium 4 nicht mehr erkennbar ist (Daten nicht gezeigt). Bis zum Stadium 10 ist keine Expression beobachtbar (Abb. 12, F). Im Stadium 13 ist das Transkript der CG3092 RNA im Darm und den Analplatten lokalisiert, und später in Teilen des Mitteldarms und des Ektoderms. Bei genauerer Analyse ist jedoch auch Expression in den Gonaden und im Gonadenschafft zu beobachten (Abb. 12, G & H). Insgesamt ist es sowohl aufgrund der unterschiedlichen Expression, die in den *in-situ* Hybridisierungen und auch in den Microarrayanalysen erkennbar ist, als auch hinsichtlich der Entfernung der beiden Gene un-

wahrscheinlich, daß es sich bei CG3092 und HDC04256 um das gleiche Gen handelt.



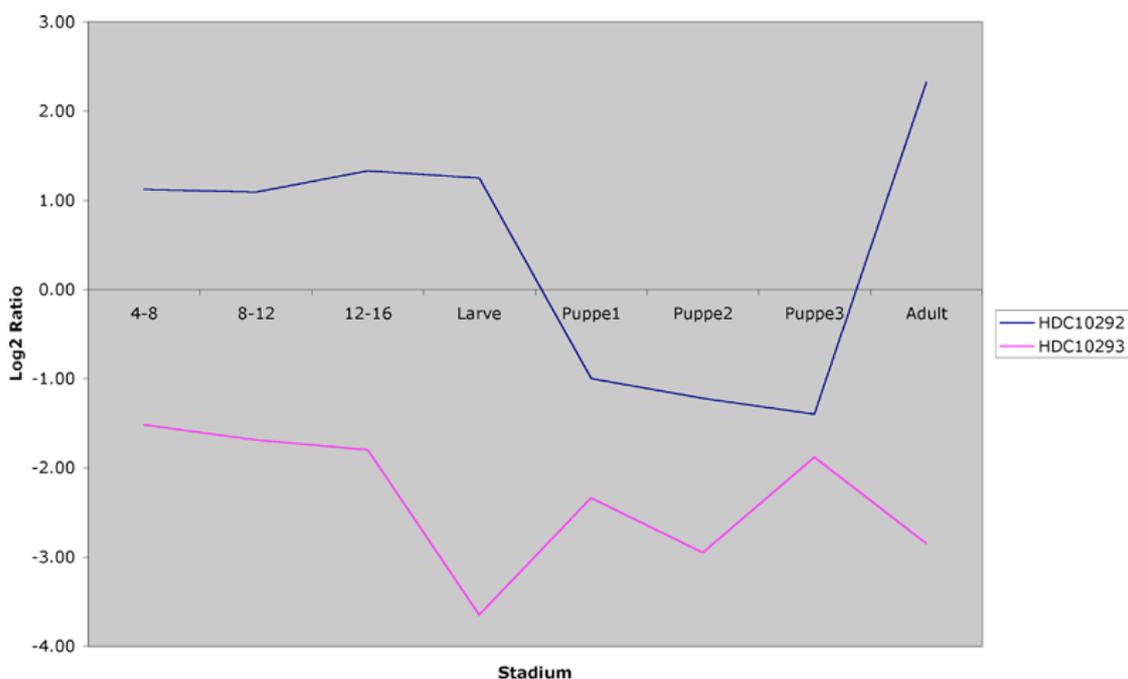
**Abb. 12 Expression der Gene HDC04256 und CG3092.** Ein Vergleich der Expression in den Stadien 0-13 zeigt kaum Übereinstimmung. HDC04256 ist ab Stadium 10 in einigen Mesodermzellen exprimiert (B), während in späteren Stadien die Gonaden und der Gonadenschaf Expression zeigen. CG3092 liegt 7 kb in 5'-Richtung von HDC04256 (I). Es zeigt maternale Expression (E), die nach sehr kurzer Zeit verschwindet (F). Im Stadium 12 ist Expression im Darm zu erkennen (G), die ab Stadium 15 stärker wird. Zusätzlich kann man im Stadium 15 die Gonaden erkennen (H). Die Abbildungen E-H entstammen dem Berkeley Expression Projekt (Tomancak et al. 2002). Die Darstellung der 20 kb flankierenden Region des Gens mit den BDGP Transkript Vorhersagen der Version 2 (grau) und 3.1 (blau) sowie der entsprechenden Genvorhersage (rot), der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplikons (blaue Linien), existierenden P-Elementen (blaue Raute), vorhandenen ESTs und cDNAs sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) sind der HDFlyArray Webseite entnommen (hdflyarray.zmbh.uni-heidelberg.de).

### 2.7.2 HDC10292 ist in der zweiten Welle von Neuroblasten und in einzelnen Neuronen exprimiert

HDC10292 wurde in der Version zwei des BDGP vorhergesagt, in der neueren Version allerdings verworfen. Es handelt sich um ein Gen auf dem dritten Chromosom, in der direkten Nähe der Gene CG7945 und CG7973. In der Version 2.1 des BDGP wurden die vier Exons von HDC10292 und das Gen CG7973 als zusätzliche Exons von CG7945 vorhergesagt. Gleichzeitig wurde CG7945 auch als eigenständiges Gen vorhergesagt (Abb. 13, I). Der Vergleich der Entwicklungsprofile von HDC10292 und dem benachbarten CG7973 (HDC10293) zeigt fast gegensätzliche Profile beider Gene. Während der frühen embryonalen Entwicklung zeigen

beide Gene kaum Regulation, in späteren Stadien dagegen ist, zunächst in CG7973 und ab den larvalen Stadien auch in HDC10292, eine sehr dynamische Regulation vorhanden. Obwohl es nicht ausgeschlossen werden kann, daß es sich bei HDC10292 um eine alternative Form von CG7973 handelt, geben die stark unterschiedlichen Expressionsmuster keinen Hinweis hierauf. Potentielle Spleißformen würden demnach unter unterschiedlicher Kontrolle liegen. Eine Analyse der Kozak-Sequenz ergab weiterhin, daß HDC10292 über eine mögliche Kozak-Sequenz verfügt, mit Adenin an der Position -3 und Cytosin an der Position +4.

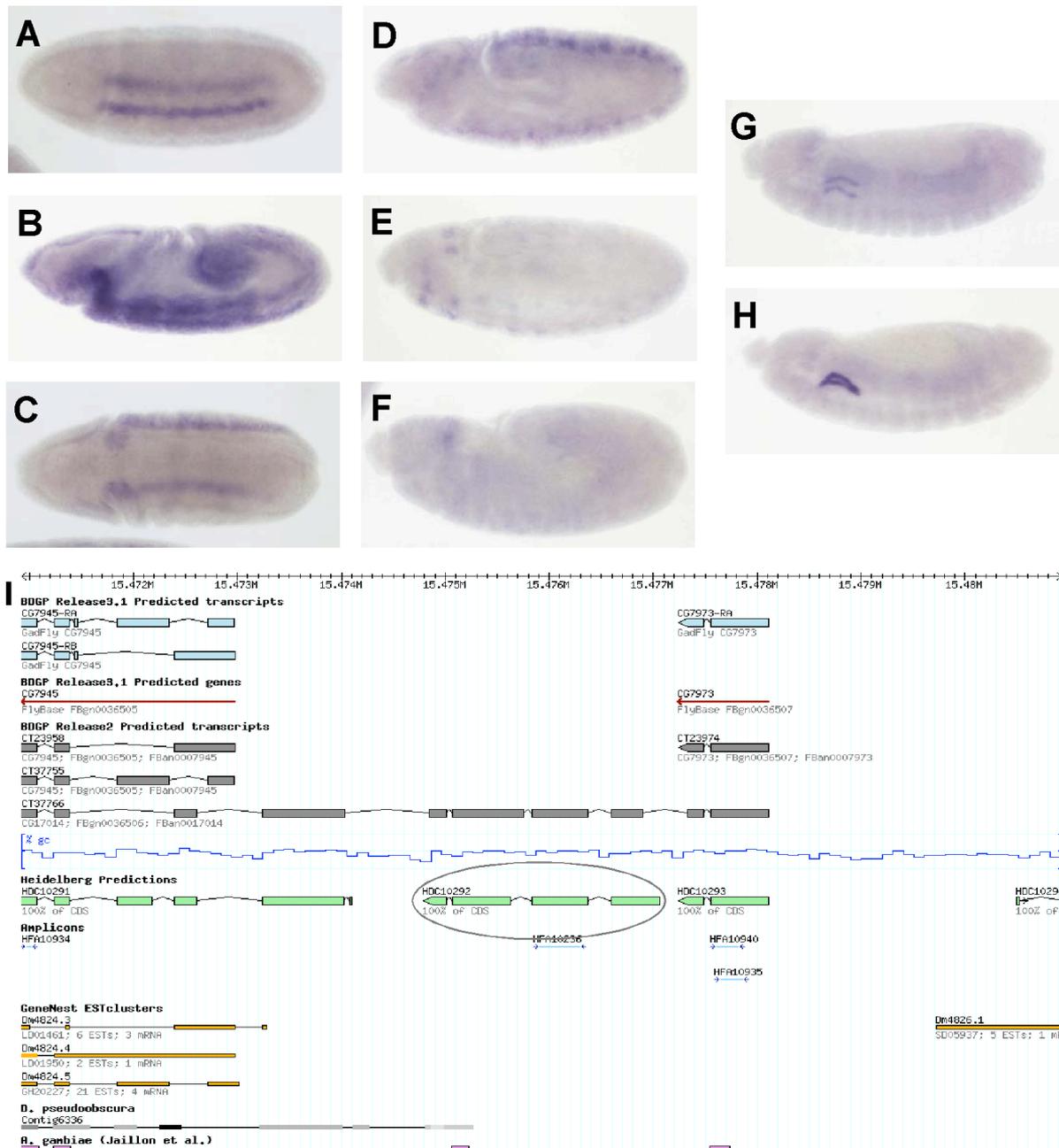
Bei HDC10292 handelt es sich um eines der wenigen neuen Gene, für die eine Proteindomäne gefunden wurde. Eine Chitin-bindende Domäne (CBP14) liegt in der Region, die sowohl in *Drosophila pseudoobscura* als auch in *Anopheles gambiae* konserviert ist.



**Abb. 13 Die Expressionsprofile von HDC10292 und HDC10293.** Die fast gegenläufigen Entwicklungsprofile beider Gene sprechen dafür, daß es sich bei HDC10292 mit hoher Wahrscheinlichkeit um ein eigenständiges Gen handelt.

Die Expression von HDC10292 beginnt bereits vor der Gastrulation im zellulären Blastoderm. In diesem Stadium beginnt die Neurogenese von *Drosophila*. Das zentrale Nervensystem entwickelt sich von einem bilateralen Neuroektoderm, welches zu beiden Seiten der Mittellinie liegt. Einzelne neuroektodermale Zellen delaminieren vom Oberflächenepithelium und wandern in das Innere des Embryos, um dort die Neuroblasten zu bilden. Diese frühen Neuroblasten bilden ein orthogonales Gitter von vier Reihen entlang der ventrolateralen Achse und drei Wellen entlang der anteroposterioren Achse. Jeder der Neuroblasten exprimiert eine charakteristische Kombination von Genen, und wird so zu einer stereotypen Familie von Neuronen und Gliazellen des ZNS. In der Vergangenheit konnte die Funktion der Gene *ventral nervous system defective (vnd)* für die anteriore Welle der Neuroblasten (Jimenez et al. 1995), *intermediate nervous system defective (ind)* für die mittlere (Weiss et al. 1998), und *muscle segment homeobox (msh)* für die posteriore Welle von Neuroblasten gezeigt werden (D'Alessio und Frasch 1996). HDC10292 zeigt auffällige Ähnlichkeit in der Expression zu *ind*. Im zellulären Blastoderm im Stadium 5 kann man die Expression in bilateralen Streifen beobachten. In diesem Stadium ist jeder Streifen etwa 5 Zellen breit und reicht von der Region des Prozephalon bis in die posterioren Bereich des Embryos (Abb. 14, A). Während der Gastrulation verdünnen sich die bilateralen Streifen, bis sie schließlich im Stadium 10 nur noch eine Breite von drei Zellen aufweisen (Abb. 14, B -D). Ab Stadium 10 kann man Expression in einzelnen Neuronen erkennen (Abb. 14, E). Nach dem Stadium 11 ist die Expression von HDC10292 nicht mehr erkennbar. Eine vermeintliche Expression in den Speicheldrüsen ist auf Hintergrundfärbung zurückzuführen, die in diesen Geweben häufig zu finden ist (Abb. 14, F-H).

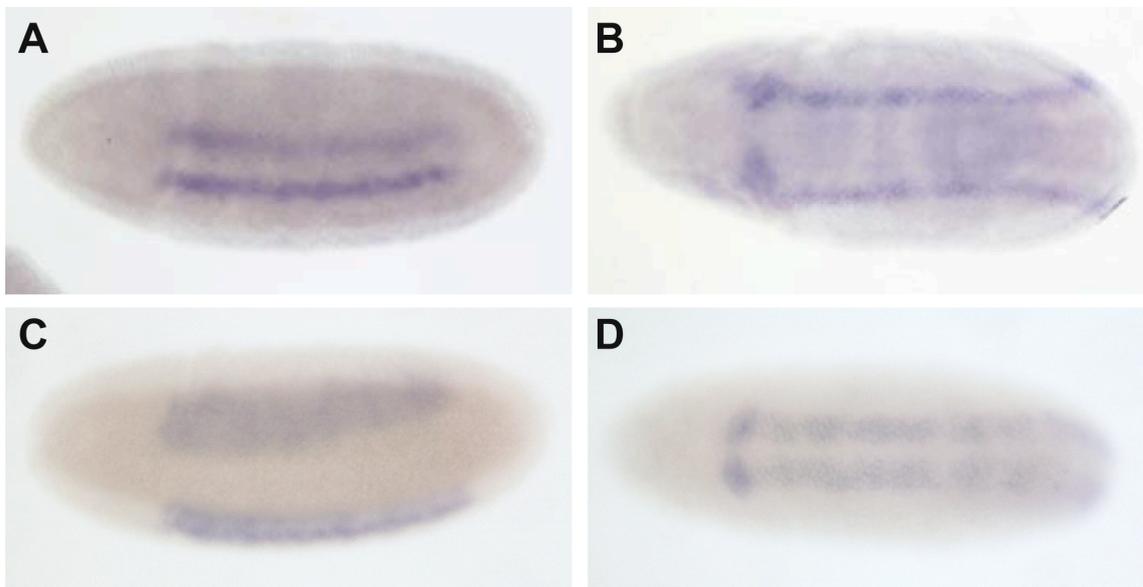
*Ind* liegt auf dem dritten Chromosom und wurde zytologisch auf den Bereich 71A kartiert. Um auszuschließen, daß es sich bei der Vorhersage von HDC10292 um eine homologe Genregion oder um eine Duplikation aufgrund schlechter Sequenzierungsergebnisse in dieser Region handelt, wurden die Amplikons beider Gene, die für die Herstellung der *antisense*-Proben für die *in-situ* Hybridisierung verwendet wurden, miteinander verglichen. Es konnte kein Alignment für die beiden Sequenzen gefunden werden. Des Weiteren ergab ein Sequenzvergleich der kompletten Genvorhersagen beider Gene ebenfalls keine Ähnlichkeit, so daß es sich bei *ind* und HDC10292 eindeutig um zwei unterschiedliche Gene handelt.



**Abb. 14 Die Expression von HDC10292.** Während des zellulären Blastoderms im Stadium 5 ist die Expression von HDC10292 in bilateralen Streifen mit einer Breite von 5 Zellen zu beobachten (A). Während der Gastrulation verdünnen sich die Streifen (B und C), bis sie im Stadium 9 nur noch etwa 3 Zellen breit sind (D). Im späten Stadium 10 kann man Expression in einigen Neuronen erkennen (E), während in späteren Stadien keine Expression mehr erkennbar ist (F-H). Die Embryonen wurden mit einer Antisense Probe gegen HDC10292 hybridisiert und von lateral (A, B, D-H) oder ventral (C) aufgenommen, anterior ist links. Die Darstellung der 20 kb flankierenden Region des Gens mit den BDGP Transkript Vorhersagen der Version 2 (grau) und 3 (blau) sowie der entsprechenden Genvorhersage (rot), der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplifikons (blaue Linien), vorhandenen ESTs und cDNAs sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) und *Anopheles gambiae* (pink) sind der HDFlyArray Webseite entnommen (hdflyarray.zmbh.uni-heidelberg.de).

Die Expression von *ind* in der mittleren Welle von Neuroblasten wird durch das Gen *vnd* reguliert. In *vnd* Nullmutanten ist die Expression von *ind* nicht auf die mittlere Welle beschränkt, sondern dehnt sich auf die ventrale Welle aus.

Interessanterweise zeigt die *in-situ* Hybridisierung von HDC10292 auf *vnd*-mutanten Embryonen ebenfalls eine Ausweitung der Expression in die ventrale Reihe von Neuroblasten. Die zu beobachtenden Streifen sind während des Stadiums 5 etwa doppelt so breit wie die im Wildtyp erkennbaren. Während der Gastrulation verengen sich die Streifen, aber auch im Stadium 10 sind die bilateralen Streifen deutlich breiter als im Wildtyp (Abb. 15). Ähnlich wie für *ind*, scheint die Expression von HDC10292 von *vnd* reguliert zu werden, eventuell durch die Fehlexpression von *ind*.



**Abb. 15 Die Expression von HDC10292 wird von *vnd* reguliert.** Während die Expression von HDC10292 im Wildtyp auf einen bilateralen Streifen in der Region der zweiten Reihe von Neuroblasten limitiert ist (A und B), kann man in *vnd* Mutanten eine Ausdehnung des Expressionsmusters von HDC10292 auf die erste Neuroblastenreihe beobachten (C und D). Alle Embryonen wurden mit der gleichen *antisense*-RNA gegen HDC10292 hybridisiert. A und C zeigen Embryonen im zellulären Blastoderm im Stadium 5, C und D zeigen die Expression im Stadium 10. Die Embryonen sind lateral (A), ventrolateral (C) und ventral (B und D) dargestellt, anterior ist links.

## 2.8 Neue Gene in isolierten Lokalisationen

Wie bereits oben beschrieben gehören Gene dieser Gruppe an, wenn innerhalb von 10 kb kein bekanntes Gen in der gleichen Transkriptionsrichtung vorhergesagt wurde. Damit ist die Wahrscheinlichkeit, daß es sich bei diesen Vorhersagen um bisher nicht vorhergesagte, neue Gene handelt, in dieser Gruppe am höchsten. Eine Liste

mit allen Genen dieser Gruppe ist in Tabelle 6 dargestellt, und drei Mitglieder dieser Gruppe sollen im Folgenden besprochen werden.

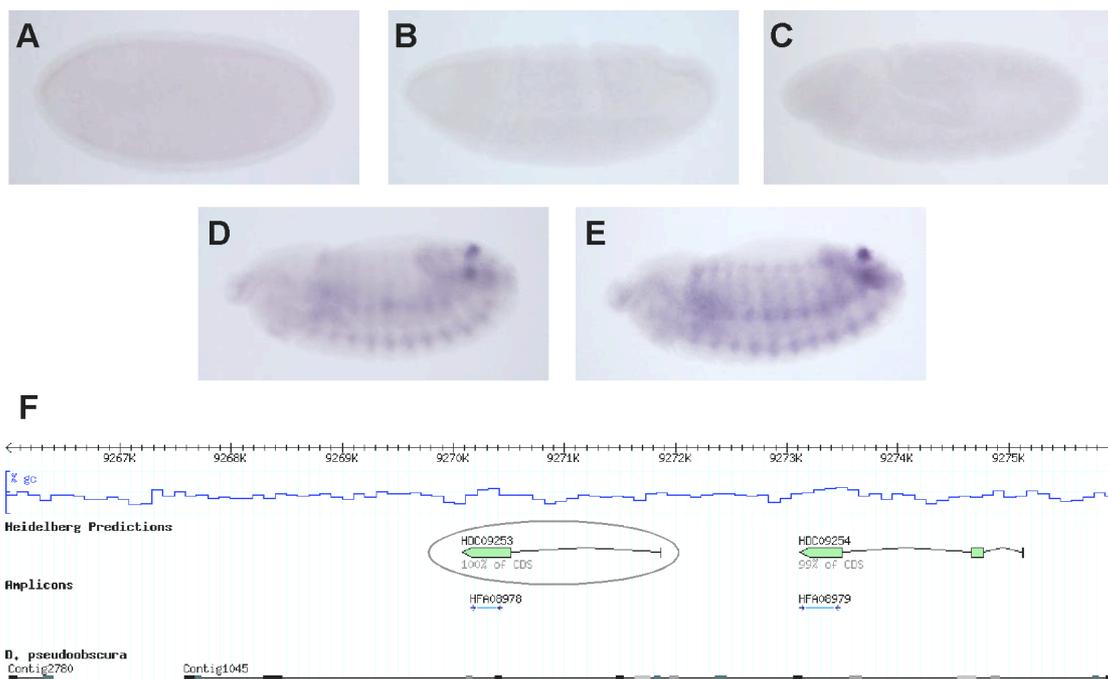
Genname	Genbank-ID	Chrom.	Expression	Evidenz
HDC00627	BK003299	2L	Ektoderm	-
HDC00658	BK003302	2L	zelluläres Blastoderm, Speicheldrüsen	-
HDC00888	BK003318	2L	Streifenmuster Stadium 10	-
HDC00979	BK003327	2L	Dotterkerne	-
HDC02005	BK003369	2L	maternal, Darm	<i>Dp</i>
HDC02009	BK003370	2L	Gehirn, Mesodermprimordium	<i>Dp</i>
HDC02141	BK003388	2L	Darm, Zellen im Kopf, (Stadium 10 / 11)	-
HDC02262	BK003403	2L	Schwach	<i>Dp</i>
HDC02272	BK003405	2L	Schwach	<i>Dp</i>
HDC02279	-	2L	Blutgefäße/Herz?	-
HDC02528	BK003430	2L	Gehirn, Mitteldarmprimordium	<i>Dp</i>
HDC02600	-	U	Somatische Muskulatur	-
HDC02634	BK003455	2L	Zelluläres Blastoderm	<i>Dp</i>
HDC02817	-	2L	Mesodermanlage	<i>Ag, Dp, EST</i>
HDC02849	BK003505	2L	maternal	-
HDC04256	BK003630	2R	Mesoderm, Gonaden	<i>Dp</i>
HDC05090	BK003664	2R	Ektodermprimordium, Mitteldarm	-
HDC05183	BK003670	2R	Ubiquitär	<i>Dp</i>
HDC06636	BK003845	2R	maternal	<i>Dp, EST</i>
HDC07387	BK003934	2R	maternal, einzelne Zellen bis Stadium 12	-
HDC07791	BK001850	3L	schwach, ubiquitär, 4-8h	-
HDC08749	BK001956	3L	schwach, ubiquitär, 4-8h	<i>Dp</i>
HDC09253	BK002020	3L	posteriore Spirakel, Ektoderm	<i>Dp</i>
HDC09513	BK002067	3R	schwach, ubiquitär 4-8h	<i>Dp</i>
HDC10646	BK002195	3L	Polplasma, Mitteldarm	<i>Dp</i>
HDC10913	BK002212	3L	Mitteldarmprimordium	<i>Dp</i>
HDC11249	BK002252	3L	Malpighi, Gonaden	<i>Dp</i>
HDC11876	BK002318	3R	schwach, ubiquitär 4-8 h	<i>Dp, EST</i>
HDC12497	BK002400	3R	schwach, ubiquitär 4-8h	-
HDC13248	BK002490	3R	schwach, ubiquitär 4-8h	-
HDC13350	BK002511	3R	Ektoderm	-
HDC13470	BK002532	3R	Streifenmuster im zellulären Blastoderm, Ektoderm, Darm	<i>Dp, EST</i>
HDC13644	BK002551	3R	Mitteldarm, Analplatten	-
HDC13905	BK002563	3R	Mesodermanlage, Mitteldarm	<i>Dp</i>
HDC14221	BK002623	3R	ventrale Ektodermanlage, posteriore Endodermanlage	<i>Dp</i>
HDC15090	BK002773	3R	Maternal	<i>Dp</i>
HDC15728	BK002837	3R	Maternal	<i>Dp</i>
HDC16092	BK002888	3R	schwach, ubiquitär 4-8h	<i>Dp</i>
HDC16243	BK002914	3R	Anteriore Endodermanlage, Mitteldarmprimordium,	<i>Dp</i>
HDC17001	-	X	Speicheldrüsenprimordium, Speicheldrüsen	<i>Dp</i>
HDC19530	BK003190	X	Schwach, ubiquitär 4-8h	-
HDC19643	BK003204	X	Mitteldarm	<i>Ag</i>
HDC19645	BK003205	X	Becherzellen	-
HDC20140	-	2h		-

**Tabelle 6 Neu vorhergesagte Gene, die nicht in der Nähe bekannter Gene lokalisiert sind..**

Gennamen, die zugehörigen Genbank IDs und das Chromosom, auf dem sie kartieren, sind dargestellt. Die beobachteten Expressionsmuster, Homologie zu *Drosophila pseudoobscura* (*Dp*) und *Anopheles gambiae* (*Ag*), sowie vorhandene cDNAs und ESTs geben weitere Hinweise auf die Existenz der potentiellen Gene.

### 2.8.1 HDC09253, ein ektodermal exprimiertes Gen

Auf dem Chromosomenarm 3L, in einer Region, in der innerhalb der nächsten 35 kb in 5'- oder in 3'-Richtung kein bekanntes Gen in der gleichen Transkriptionsrichtung vorhergesagt wurde, liegt HDC09253 (Abb. 16). Allerdings liegt etwa 1 kb oberhalb ein weiteres von fgenesh vorhergesagtes Gen, HDC09254. Für beide Gene sind keine ESTs oder cDNAs bekannt, kein P-Element ist in dieser Region inseriert, und die Ähnlichkeit zu *Drosophila pseudoobscura* beträgt weniger als 40%.



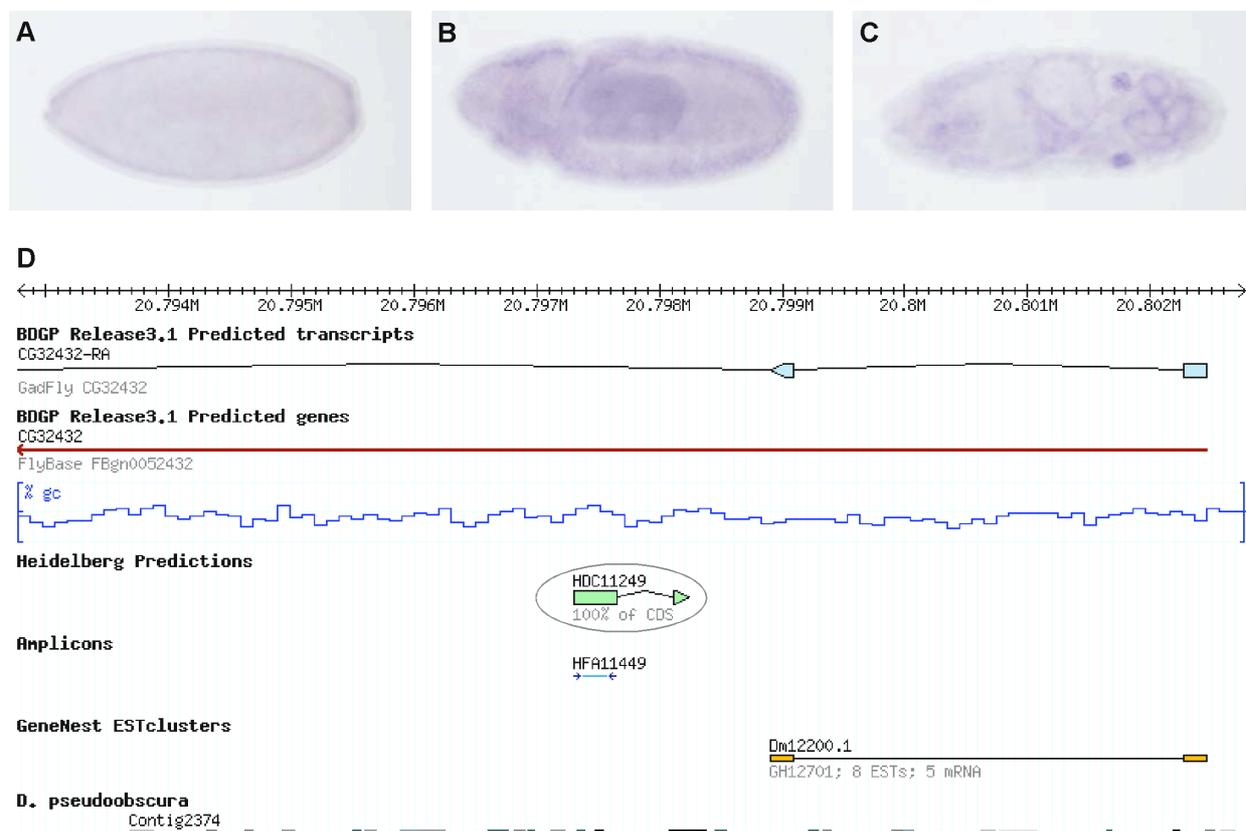
**Abb. 16 Expressionsmuster und Lokalisation von HDC09253.** In jungen Embryonen bis zum Stadium 12 kann keine Expression beobachtet werden (A-C). Ab dem Stadium 12 ist das RNA Transkript deutlich im Ektoderm und in den Analplatten lokalisiert und ist bis zum Ende der Embryonalentwicklung beobachtbar (D & E). HDC09253 liegt in einer Region, in der innerhalb von 20 kb kein Berkeley Gen lokalisiert ist. Keine ESTs oder cDNAs geben Hinweise auf das Vorhandensein eines Gens in dieser Region. Auch existiert nur eine schwache Homologie zu *Drosophila pseudoobscura* (I). Alle Embryonen wurden mit einer *antisense*-RNA-Probe gegen HDC09253 hybridisiert. Die Ansicht der Embryonen ist von lateral, anterior ist links. Die Darstellung der 20 kb flankierenden Region des Gens mit der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplikons (blaue Linien) sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) sind der HDFlyArray Webseite entnommen ([hdflyarray.zmbh.uni-heidelberg.de](http://hdflyarray.zmbh.uni-heidelberg.de)).

Bis zum Stadium 10 ist für HDC09253 keine Expression zu beobachten (Abb. 16, A – C). Starke Expression im Ektoderm und in den Analplatten kann jedoch ab dem Stadium 12 bis in die späten Embryonalstadien beobachtet werden (Abb. 16, D & E). HDC09253 ist ein Beispiel für ein bisher nicht vorhergesagtes Gen, für dessen

Existenz aufgrund des spezifischen Expressionsmusters keine Zweifel bestehen. Inwieweit das Genmodell mit HDC09254 assoziiert ist, wurde nicht untersucht.

## 2.8.2 HDC11249 ist in den malpighischen Gefäßen und in den Gonaden exprimiert

Als zweites Beispiel für ein Gen dieser Gruppe dient HDC11249. Auch dieses Gen liegt innerhalb einer Region, in der innerhalb von 10 kb kein weiteres Gen vorhergesagt wurde. In entgegengesetzter Transkriptionsrichtung ist das Gen CG32432 vorhergesagt.



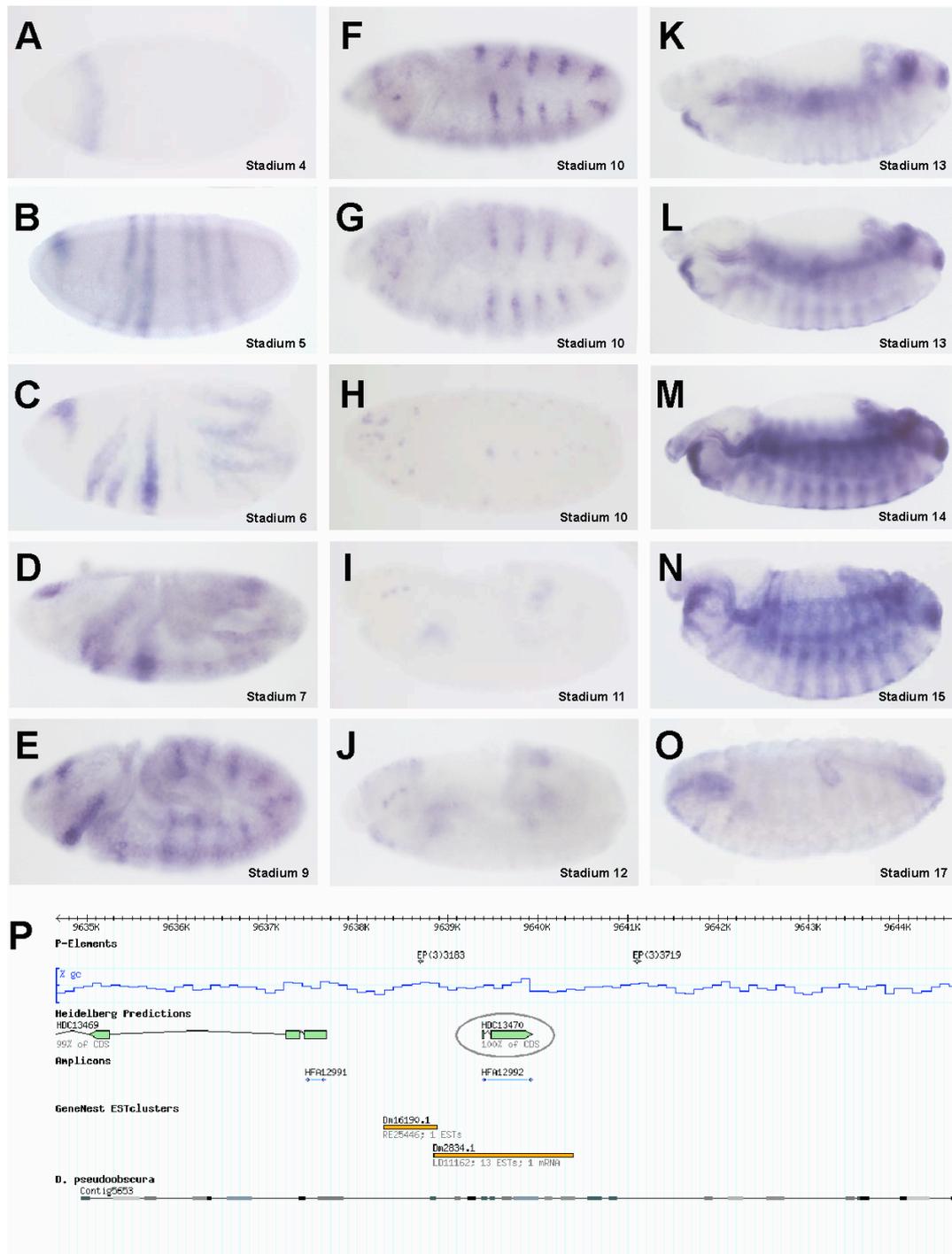
**Abb. 17 Die Expression von HDC11249** ist zuerst im Stadium 8 in der Einstülpung des entstehenden Darmes zu erkennen (B). Im Stadium 15 erkennt man deutlich die Expression in den Gonaden und den malpighischen Gefäßen (C). HDC11249 liegt auf dem 3. Chromosom, in derselben Region, in der CG32432 vorhergesagt wurde, allerdings in entgegengesetzter Transkriptionsrichtung (D). Die Embryonen sind lateral (A) und (C) oder ventrolateral (B) aufgenommen, anterior ist links. Die Darstellung der 20 kb flankierenden Region des Gens mit der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplikons (blau) sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) sind der HDFlyArray Webseite entnommen (hdflyarray.zmbh.uni-heidelberg.de) (I).

Die RNA von HDC11249 ist ab dem Stadium 10 im sich entwickelnden Mitteldarm lokalisiert (Abb. 17, B). Spät in der Embryogenese, ab dem Stadium 15, findet man die Expression in den Gonaden und den malpighischen Gefäßen (Abb. 17, C). Große Teile der exonischen Region sind in *Drosophila pseudoobscura* konserviert, was auf eine Funktion des Gens in beiden Organismen schließen lässt.

Wie schon die meisten anderen besprochenen Gene weist HDC11249 keine bekannten Domänenstrukturen auf, mit deren Hilfe man auf eine Funktion schließen könnte. Dennoch kann aufgrund der spezifischen Expression und der isolierten Lage die Existenz von HDC11249 als gesichert angesehen werden.

### **2.8.3 HDC13470 zeigt eine sehr dynamische Expression und scheint mehrere identische Peptide zu kodieren**

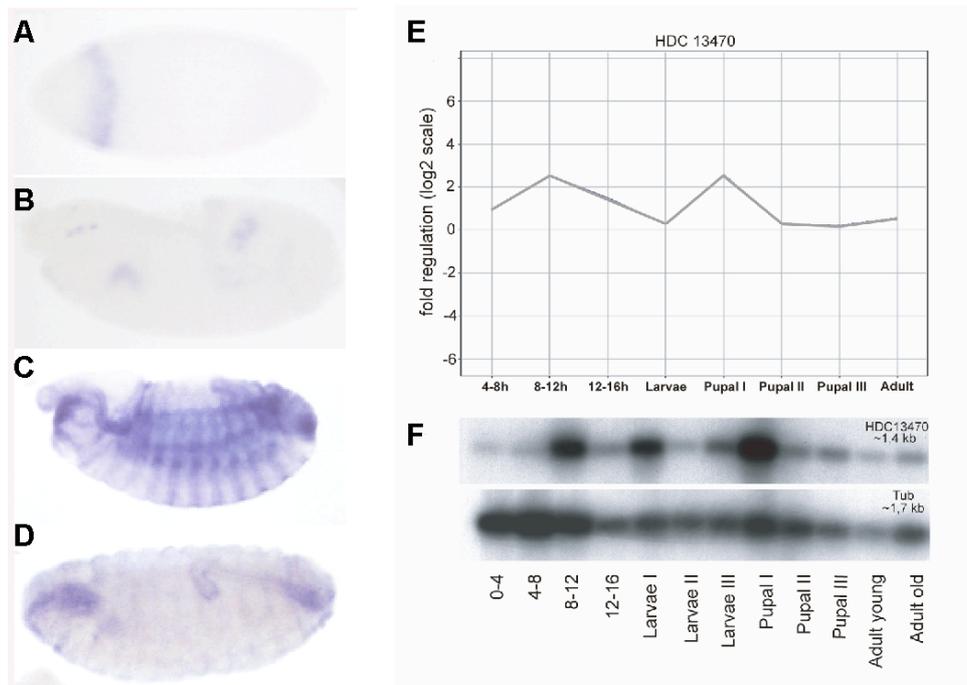
HDC13470 war eines der ersten Gene, die wegen ihrer hochvariablen Expression in der *in-situ* Hybridisierung auffielen. Bereits im synzytialen Blastoderm kann man einen einzelnen anterioren Streifen entlang der Dorsoventralachse beobachten (Abb. 18, A). Wenig später, im zellulären Blastoderm, scheint dieser Streifen etwas nach anterior verschoben zu sein, und sieben weitere Streifen erscheinen in einem segmentalen Muster (Abb. 18, B). Der vierte dieser Streifen erscheint deutlich schwächer als die anderen. Während der Gastrulation bleiben die Streifen erhalten und verlagern sich entsprechend der Invaginationsbewegung des Embryos. (Abb. 18, C - E). Im Stadium 10 kann man den anterioren Streifen weiterhin gut vor der Kopffurche erkennen, zehn weitere Streifen sind ab der Mitte des Embryos erkennbar, wobei weiterhin der vierte Streifen etwas schwächer exprimiert ist als die übrigen (Abb. 18, F). Im Stadium 10, nur wenig später, sind einzelne Zellen oder Zellgruppen in der Kopfregion beobachtbar, während die Expression in den Streifen verblasst (Abb. 18, G). Kurz darauf sind nur noch einzelne Zellen in den Regionen zu erkennen, wo vorher das Streifenmuster zu sehen war. Drei Zellgruppen im Kopf sind weiterhin deutlich sichtbar (Abb. 18, H – J).



**Abb. 18 Expressionsmuster von HDC13470 während der embryonalen Entwicklung.** Während im zellulären Blastoderm zunächst ein Streifen gefolgt von sieben Streifen sowie einige Zellen in der Kopfregion exprimiert sind, die durch die Gastrulation erhalten bleiben, verliert sich die Expression im Stadium 10. Nur einige Zellen im Kopfbereich bleiben sichtbar exprimiert. Ab Stadium 12 kann man ein Signal in den Analplatten erkennen, sowie Expression im Darm und im Ektoderm, die bis zum Stadium 15 stärker wird. Obwohl für HDC13470 bis jetzt kein Gen vorhergesagt wurde, existieren eine cDNA sowie 13 ESTs für diese Region. Die Darstellung der 20 kb flankierenden Region des Gens mit der HeidelbergCollection (grün), den auf dem Array vorhandenen Amplikons (blau), existierenden P-Elementen (blaue Raute), vorhandenen ESTs und cDNAs sowie den Ähnlichkeiten zu *Drosophila pseudoobscura* (grau schattiert) sind der HDFlyArray Webseite entnommen (hdflyarray.zmbh.uni-heidelberg.de).

Die beobachteten Expressionsprofile der *in-situ* Hybridisierung und des Northern Blots wurden mit denen der Microarrayversuche verglichen. Obwohl *in-situ* Hybridisierungen wegen der unterschiedlichen Expression in verschiedenen Zellen oftmals schlecht mit Expressionsprofilen aus Microarrayexperimenten vergleichbar sind, kann man in diesem Fall eine gute Korrelation der *in-situ* Hybridisierung mit den beiden anderen Ansätzen feststellen (Abb. 19). Weiterhin ist in den Microarrayexperimenten und im Northern Blot erkennbar, daß die dynamische Expression von HDC13470 in den späteren Lebensstadien erhalten bleibt.

**Abb. 19** Der Vergleich unterschiedlicher Expressionsdaten von HDC13470. *In-situ* Hybridisierungen (A-F), das Entwicklungsprofil der Microarrayanalysen (G) sowie ein Northern Blot aller Entwicklungsstadien (H) zeigen gut vergleichbare Ergebnisse. Die Microarrayanalysen und der Northern Blot zeigen die Fortsetzung der hochdynamischen Expression von HDC13470 während späterer Stadien des Lebenszyklus.



### 2.8.3.1 Überprüfung des Genmodells

HDC13470 liegt auf dem dritten Chromosom in der Region 83F. Obwohl bisher kein Gen in dieser Region vorhergesagt wurde, existiert eine cDNA (LD11162). Auffällig ist, daß die Sequenz der cDNA nicht vollständig mit der von HDC13470 übereinstimmt. Während für LD11162 aus einem einzelnen Exon besteht, ist HDC13470 mit zwei Exons vorhergesagt. Zunächst wurde überprüft, ob die

Vorhersage des Genmodells korrekt ist. Über eine Northern Blot Analyse wurde die Länge der mRNA sowie die relative Expression in den einzelnen Entwicklungsstadien untersucht (Abb. 19). Der Northern Blot ergab eine mRNA Länge von etwa 1,500 Basenpaaren, was der Länge von LD11162 entspricht. Dieses Ergebnis schließt aus, daß HDC13470 mit einem weiteren Gen assoziiert ist.

Als nächstes wurde untersucht, ob das für HDC13470 vorhergesagte Intron existiert. Die Primer für das auf dem Microarray vorhandene 500 bp Fragment wurden parallel in einer RT-PCR und in einer PCR mit genomischer DNA als Template verwendet. Die entstehenden Fragmente sind von ihrer Länge identisch, was gegen das Vorhandensein eines Introns spricht. Zusätzlich wurde das in der RT-PCR entstandene Fragment und LD11162 sequenziert. Da beide Sequenzen identisch sind und mit der genomischen Sequenz übereinstimmen, kann eindeutig nachgewiesen werden, daß HDC13470 aus nur einem Exon besteht.

### **2.8.3.2 Analyse der Aminosäuresequenz**

Während das von fgenesh vorhergesagte Genmodell von HDC13470 ein potentielles Peptid von 162 Aminosäuren kodierte, existiert für die nachgewiesene cDNA kein längeres ORF. Handelt es sich bei HDC13470 um eine nicht protein-kodierende RNA, die nur aufgrund eines falsch vorhergesagten Introns in der ORF Vorhersage vorkam, oder gibt es Hinweise auf eine Proteinexpression?

Um dies zu analysieren, wurden die möglichen Proteinsequenzen in allen Leserastern genauer untersucht. Während auf der DNA-Sequenz keine Auffälligkeiten gefunden wurden, konnten dagegen Sequenzen für vier potentielle Peptide identifiziert werden (Abb. 20), in denen fünfmal ein fast identisches Motiv auftaucht. Die ersten beiden Peptide sind über die gesamte Länge dieses Motivs MAAYLDPTGQY\* identisch (Abb. 20). Das dritte Peptid unterscheidet sich in der Anfangssequenz, die acht letzten Aminosäuren sind jedoch bis auf den Austausch des Glutamins durch Threonin identisch. Das vierte Mal ist die Sequenz identisch zum dritten Peptid, mit dem Unterschied, daß sie kein Stoppkodon aufweist, sondern den Start eines längeren ORFs bildet. Die fünfte Wiederholung der Sequenz ist ebenfalls innerhalb dieses ORFs lokalisiert, und ist in der Sequenz identisch zu den ersten beiden. Die kurzen ORFs tauchen in allen drei Leserastern auf, und die Nukleotidsequenz ist für keines der Peptide identisch.

```

1/1
31/11
61/21
91/31
CRA GAT TCG ACG AGT GCA GAT CAC CAG CTA AAA GAA AAC CAG CTG AGA CAT CAG AAA AGT CCG CAG ATA TTC ACG TAA CGC CTT AAA GAT TTT CCG TCG GGT TCC CGA ACA AAC TAA ACA
Q H S T S A D H Q L E N Q L R R H Q K S P Q T F T * R L K D F P C G S R T N * T
N I R R V Q I T S * K K T S * D I R Y S R N A L K I F R A V P E P E T K H I
T F D E C R S P A K R K P A E T S E K S A D I H V T P * R F S V R F P N K L N H
121/41
151/51
181/61
211/71
TTA TTA ACA AAC ATT AAA CGA ATT TGT AGT GTC AGT GAC TTT TGA ACG CAC GAA AAA ATT CCC AAA CAC ACA ACC AAA CGT GAC TGT ATA TTC AGC CCC AAG AAA CCC AAC AAC TGG TGG
L L T N N K R I C S V S D F * T H E K I P K H T T K R D C I F S P K K P N N W W
Y * Q T I N E F V V S V T F E R T K K F P N T Q P N V T V Y S A P R N P T T G G
I N K Q * T N L * C Q * L L N A R K N S Q T H N Q T * L Y I Q P Q E T Q Q L V V
241/81
271/91
301/101
331/111
TGA TAA TAA AAG AAC TTA CAA CAA CAG CGC CGA GAA CCA GTA TAA AGT TCA ATA CCG CGC TGA TTC AAA TTA AAC AAA GGA GAA TCG ACA GCA GCA GCA GCA GCA GAA ACA AAA AGC CAG
* * * K N L Q Q Q R R E P V * S S I P R * * F K L N K G E S T A A A A A E T K S Q
D N K R T Y N N S A E N Q Y I K V Q Y R A D S N * T K E N R Q Q Q Q Q K Q K A A
I I K E L T T T A P R T S Y I K F N T A L I Q I K Q R R I D S S S S R N K K P A
361/121
391/131
421/141
451/151
CTC GGT TTT GTC ATT CAA GFA TTT TTG GTC AAT ACA CGG CAT ACG AAA TGG CAG CCT ACT TGG ATC CCA CTC GCG AGT ACT AAA GAA GCT ACA CGA CGA CAA GAC ATC GFA ATC GTA
L G F V I Q V F L V N T R H T K W G Q P T L G I P L A S T K E A T R R Q Q D I V I V
S V L S F K Y F W S I H G I R R A L L L G S H W P V L K K L H D R D S K T S *
R F C H S S I F G Q Y T A Y E M A A Y L D P T G Q Y * R S Y T T A R H R N R R
481/161
511/171
541/181
571/191
GAC CTC TTT TAG AAA ATC CAA TAA ATC ACA GAT CTT CGC CAT GGC CCG CTA TCT GGA TCC CCG CGG TCA GTA CTG AAG TTG GAG CAA GCA AGC AGA AGC AAT ATT TTG AGT TCC ARG
D L F * K I Q * I T D L R H G R L S G S H W S V L K L E Q A S R S S N I L F * V P S
T S F R K S N K S Q I F A M A A Y L D P T G Q Y * S W S K Q A E A I F * V P S
P L L E N P I N H R S S P W P P I W I P L V S T E V G A S K Q K Q Q Y F E F Q A
601/201
631/211
661/221
691/231
CCG AAA GAT ATT TAA AAC AGT ATC AAA ATG TCG CAC GAT TTG GAC CCC ACT GCG ACC TAC TAA GGT TCT ATC ACA AGA ACT CCA CAT AGC CAA GCA TTC TAA GGC TGA ATA CTA TAC CCA
P K V I * N S I K M S H D L D P T G T Y * G S I A R T P H S G A F * G * I L Y P
R K L F K T V S K C R T I W T P L A P T K V Y L S Q E L H I A K H S K A E Y Y T H
E S Y L K Q Y S K N V A R F G P H W H L L R F Y R K N S T * P S I L R L N T * P T
721/241
751/251
781/261
811/271
CIT CAA AAG CTC CAC AAA TAC AAT CCT TAA AAA TGC TGG ATC CCA CTG GAA CAT ACC GGC CAG CAC GCG ACA GCG AGG ACT CCC GCC AAA AGC GGC GAC GGC ACT GCC TGG ATC CAA CCG
L Q K L H K Y N P * K C W I P L E H T G D H A T R R R T P A K R G D R T A W I Q P
F K S S T N T I L K N A G S H W N I P A T T R H A G L P P K E A T G L P G S N R
S K A P Q I Q S L K M L D P T G T Y R R R D T Q D S R Q K R R Q D C L D P T G
841/281
871/291
901/301
931/311
GGC AGT ACT AGA CGC TGA YAT CCC AAC AAC AGT GCC CCA TAA CGC CCG TGC CTT ATC CAC AAA CTC TGG GCA TGA TTG GGG GAG CCG GCT GGT TGC GCG TTC GTG GCC GCG AGG AGA CTT
G S T R R * * P N H S A P * R R P C L I H K L W A * L G E R A G C A F V A A R R L
A V L D A I P T T V P H N R A L L S Q N T L G M I G G A R W L R V R G R E T D F
Q Y * T L I S Q Q C P I T P V Y P Q T L G M I G G A R W L R V R G R E T D F
961/321
991/331
1021/341
1051/351
CCA GCT GCG GCG GGA GAA AAG TGG GGA TCG GGG CTT CCC CAA GCG ATC TTG GCG AGC CCT GCG ATG GAG ACT TTT GTA TTT ATG TAG TTT TTG GGT AGC CTA TCA ATA CCT ATT ATA
P A A G G E E S W G S G L P Q A I L G S P A M E T F V F M * F L R S P L R S I P I I
Q L P A E K K A G D R G F P K R S W G A L R W R L L Y L C S F C V A P Q X Y L L Y
S C R R R R K L G I G A S P S D L G E P C D G D F C I Y V V F A * P I N T Y L Y I
1081/361
1111/371
1141/381
1171/391
TTA ATT ATT TAT TAT TAT CAT ACT ATT TAA ATT AAA TAC TGT TCA TTG CTG TTC ACA AAA CAC CGA TAC GCA CAC ATA CAT ATC CTA TAT CCT ATA TTG TAT CAC ACA TAC ACC ATA TAT
L I I Y Y I H T I * I K Y C S L C L F T K H R Y A H I H I Y L P I L Y I H T Y T I Y
* L F I I I I L F K L N T V L H C S Q N T D T H T Y I S Y I L Y C I T H T Y P Y M
N Y L L L S Y Y L N * I L F I A V H K T P I R T H T Y P I S Y I V S H I H H I C
1201/401
1231/411
1261/421
1291/431
GTT ATA TAT ATA TAC TAT ACA CTT GTA ATT GCT TCT TCA AAT TGG AAA AGA TTA CCG AAA GAG TAT TAT GTT TAG TGT CAT ATT TCC CGA GCA AAT CAT CGT TTG TTT AAA TTA TCA TTT
V I Y I Y Y T L V I A S S N W K R L R K I V * C H I S R A N H R L F K L S F
L Y I Y T I H L * L L L Q I G K D Y A K S I M F S V I F P E Q I I S V C L * I Y I Y
Y I Y I L Y T C N C F F K L E K I T Q R V L C L V S Y F P S K I S F * I I Y I Y
1321/441
1351/451
1381/461
1411/471
ATT TAT TGC CAA GAA TTT GFA ATG TTC TTT TCT CTC TCT CGC TGA GAG CAA GGA AAA CCA TTC GGA GAG CAA GAA ATT TTG TTT AGA TCA TAA GCG TTT TTA AGC TAT TTA TAT GTC
I Y C Q R R F V M F F F S L S R * E Q G K P F G E R E I L F R S * A F L S Y V
F I A N D L * C S F F L S L A E S K E N H S E S E K F C L D H K R F * A I Y M S
L L P T I C N V L F F S L S L R A R K T I R R A R N F V * I I S V F K L F I C L
1441/481
1471/491
1501/501
1531/511
TAC ACC TTC GCA CAT CCA GAG AAC CCC AAC ACA CAC CTC TCA CAC CCA TTT AAA TAT AAT ATT AAA AGA AAA CCA TAT TTA AAA CTG AAA AAA AAA AAA AAA A
Y T P S D R H P E R P N T H L S H P F K Y H I K R K P Y L K L K K K K K
T P S T D I Q R N P T H T L S H T H L N I I L K E N H I * N * K K K K K K
H L R P T S R E P Q H T P L T P I * I * Y * K K T I F K T E K K K K

```

**Abb. 20 Die cDNA Sequenz und die entsprechenden Aminosäuresequenzen der einzelnen Leseraster von HDC13470.** Kurze, identische Aminosäuresequenzen kommen in allen drei Leserastern innerhalb von HDC13470 vor (rot). Die Tatsache, daß nur diese kurzen Motive in *Drosophila pseudoobscura* konserviert sind, wohingegen auf DNA-Ebene keine Konserviertheit beobachtet werden kann, spricht für eine Funktion der kurzen Proteine. Innerhalb der cDNA Sequenz ist die für HDC13470 vorhergesagte Sequenz in blau dargestellt. Zur besseren Übersicht sind die Anzahl der Aminosäuren sowie der entsprechenden Basenpaare alle 10 Aminosäuren angegeben.

Das einzige längere ORF dieser Vorhersage mit einer Gesamtlänge von 50 Aminosäuren befindet sich hinter diesen kurzen Peptiden. Vergleicht man die Region mit der von *Drosophila pseudoobscura*, so kann man feststellen, daß die Region auf DNA Ebene nicht konserviert ist, daß aber auf der Proteinebene eben jene kurzen Aminosäuresequenzen in genau demselben Abstand wie in *Drosophila melanogaster* zu 100% konserviert sind. Auch der N-Terminus des längeren ORFs ist in *Drosophila pseudoobscura* konserviert, wogegen die restliche Aminosäuresequenz keine Ähnlichkeit aufweist. Insgesamt deuten die bisherigen Ergebnisse darauf hin, daß die kurzen ORFs reell sind. Allerdings müssen weitere Analysen gemacht werden, um

eindeutig zu klären, ob die kurzen ORFs tatsächlich existieren und ob alle konservierten Sequenzen translatiert werden.

### **2.8.3.3 Genetische Hinweise auf eine Funktion von HDC13470**

Bisher konnten Hinweise auf die Funktion nur über die dynamische Expression und das Vorhandensein der mRNA über die gesamte Entwicklung vermutet werden. Ein Hinweis auf eine Funktion von HDC13470 ergab der P-Element Screen von Deak *et al.* (Deak *et al.* 1997). P-Elemente sind mobile Elemente, die bevorzugt in die 5'-Region von Genen integrieren und so zum Funktionsverlust des Gens führen können. In der Region von HDC13470 existiert eine P-Element Insertion, die zu Letalität in embryonalen Stadien führt. Die Linie I(3)S099049 wurde vom Szeged Stock Centre bezogen und die Insertionsstelle über einen Plasmid Rescue untersucht. Nach Sequenzierung konnte die Insertionsstelle in der 5'-UTR von HDC135470 bestimmt werden. Die Insertion des P-Elements fällt 5' vom vorhergesagten Transkriptionsstart von HDC13470, so daß die Promotersequenz eventuell zerstört oder vom Gen getrennt wird.

Suche nach kleinen Defizienzen führte zur Detektion von vier Defizienzen, welche die Region von HDC13470 deletieren. Alle diese Defizienzen führen zur Letalität in frühen larvalen Stadien. Damit sprechen alle bisherigen Hinweise für eine wichtige Funktion von HDC13470 während der Entwicklung von *Drosophila*, möglicherweise über einen in diesem Organismus bisher nicht beobachteten Regulationsmechanismus. Zusammenfassend bietet sich HDC13470 als ein hochinteressanter Kandidat für weitere molekulare und genetische Funktionsanalysen an.

## **2.9 Die HeidelbergFlyArray Web-Seite**

Zur besseren Zugänglichkeit und visuellen Darstellung aller Gene wurde eine Webseite mit allen generierten Daten und weiteren links eingerichtet, die im Internet unter <http://heidelbergflyarray.zmbh.uni-heidelberg.de> abrufbar ist. Als Basis dazu diente der Generic Genome Browser (Stein *et al.* 2002). Die Genomsequenz, die in den Berkeley Versionen 2 und 3.1 sowie die in der Heidelberg Collection vorhergesagten Gensequenzen und die auf dem Microarray gespotteten Amplikonsequenzen wurden als gff (generic file format) formatiert und in eine mysql Datenbank

geladen. Auch alle zugänglichen EST und cDNA Informationen sowie P-Element Insertionen wurden gespeichert. Zusätzlich wurde ein Blast mit den Genomsequenzen von *Drosophila pseudoobscura* und *Anopheles gambiae* gegen die *Drosophila melanogaster* Genomsequenz durchgeführt und homologe Sequenzen ebenfalls der Datenbank hinzugefügt.

The screenshot displays the HeidelbergFlyArray website interface, which is a genomic browser for *Drosophila melanogaster*. The central panel shows a genomic track for a 20 kbp region on chromosome 2R, from positions 18274472.5 to 18294471.5. This track includes predicted transcripts, genes, and P-elements. The gene *Dm12242* (CG32443 PC) is highlighted. To the left, a table titled 'Heidelberg Amplicon Set' lists various PCR amplicons with their names, coordinates, and sequences. To the right, a grid of images shows *in-situ* hybridization results for different amplicons. Below the main track, there are several panels: 'spliceNest' showing alternative splicing, 'Gene Sources' listing various databases, and 'Annotations' showing restriction sites and other features. The interface is user-friendly, with search bars and navigation options.

**Abb. 21 Die HeidelbergFlyArray Webseite und ihre Weblinks.** Über den GenericGenome Browser der HeidelbergFlyarray Webseite können alle vorhergesagten Gene der HeidelbergCollection, der Berkeley Versionen 2 und 3.1, die jeweiligen Fragmente, die auf dem Microarray vertreten sind, cDNAs und ESTs, P-Element Insertionen sowie Ähnlichkeiten zu *Drosophila pseudoobscura* und *Anopheles gambiae* dargestellt werden (Mitte). Die Informationen über die jeweiligen Amplikons, wie z.B. die Sequenz der verwendeten Primer, die Amplikonsequenz und Informationen über den Erfolg in den PCRs ist über das dargestellte Amplikon vernetzt (links oben). Die Bilder der *in-situ* Hybridisierungen sind ebenfalls über den Amplikon-link zu erreichen (rechts oben). Die cDNAs und ESTs sind mit der Datenbank GeneNest verknüpft (rechts unten), die wiederum einen link zu der Proteindatenbank Systems liefert (links unten).

Mit Hilfe dieser Webseite können die unterschiedlichen Genvorhersagen für jedes beliebige Gen verglichen werden. Des Weiteren können Informationen zu den Genen, die Sequenzinformationen und Herkunft der ESTs sowie die genauen Insertionsstellen der P-Elemente über links zu den jeweiligen Webseiten erreicht werden.

Viele Informationen zu den Amplikons, wie z.B. die Amplikonsequenz, die Primersequenzen, Informationen über die Amplifikation in der ersten und zweiten PCR und über weitere Validierungen sind ebenfalls zu erreichen. Auch die Bilder aller durchgeführten *in-situ* Hybridisierungen sind auf der Hauptseite sowie einzelne Bilder über den Amplikonlink zu erreichen.

## **2.10 Zielgene für Polycomb/Trithorax Gruppen Proteine**

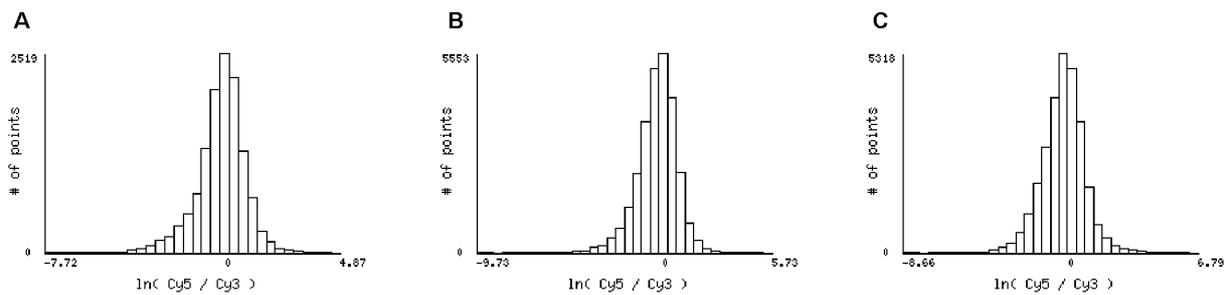
Bisher konnten die Funktionalität des HeidelbergFlyArrays gezeigt und etwa 2,000 neu vorhergesagte Gene bestätigt werden. Der zweite Teil dieser Arbeit bestand darin, ein System zu etablieren, welches es ermöglicht, Zielgene der Polycomb- und Trithoraxgruppen Proteine zu finden.

Mittels Microarrayanalysen soll über den Vergleich von wildtypischen Proben mit denen, die in einzelnen Mitgliedern der PcG/trxG mutiert sind, analysiert werden, welche Gene von Polycomb oder Trithorax reguliert werden. Damit sollen die unterschiedlichen Zielgene in verschiedenen Geweben und zu unterschiedlichen Entwicklungsstadien aufgedeckt werden.

## **2.11 Untersuchung Polycomb-mutanter Embryonen und Wildtyp-embryonen mittels Microarrayanalysen**

Für *Polycomb* sind viele verschiedene Mutanten beschrieben, von denen die meisten nicht molekular untersucht wurden. Bei  $Pc^{XL5}$ , einer oft verwendeten und gut untersuchten Mutante, sind die Aminosäuren 69 und 70 innerhalb der Chromodomäne deletiert, was zu einem Verlust der Proteinfunktion führt (Franke et al. 1995). Daher wurde diese Mutation gewählt, um die Zielgene von *Pc* auf dem Transkriptomarray zu untersuchen. Um homozygote Embryonen für das Microarrayexperiment zu sammeln, wurde die  $Pc^{XL5}$  Linie mit einem Balancer gekreuzt, der das grün fluoreszierende Protein (GFP) unter der Kontrolle des *Twist* Promoters exprimiert. Ab Stadium 9 können homozygote Embryonen am Fehlen dieses grün leuchtenden Markers unter dem Fluoreszenz-Mikroskop erkannt und isoliert werden. Etwa 800 Embryonen der homozygoten Mutante im Stadium 10 sowie die gleiche Menge an wildtypischen Embryonen des gleichen Alters wurden gesammelt, die RNA isoliert und über das Eberwine Protokoll amplifiziert.

Vier Microarrays wurden mit dem aus diesen Amplifikationen gewonnenen Material hybridisiert, wobei für zwei Arrays die Farbstoffe ausgetauscht wurden. Drei dieser Arrays waren von ihrer Qualität für eine Auswertung geeignet, so daß für jedes Gen sechs Datenpunkte zur Verfügung stehen. Die Normalisierung und Analyse der Expressionsstärke aller Gene wurde mit Hilfe der Software „express yourself“ durchgeführt.



**Abb. 22 Histogramm der ausgewerteten Microarrays.** Über das Histogramm kann die Qualität der Hybridisierungen überprüft werden. Für den Großteil aller Gene wird keine Regulation erwartet, die Farbintensität sollte daher für beide Farbstoffe gleich sein. Für alle Arrays kann dieses Qualitätsmerkmal erfüllt werden. Die Anzahl der Spots, die genau um den Nullpunkt liegen, ist für den in A dargestellten Array jedoch geringer als in B und C.

Für die hier durchgeführten Hybridisierungen dienten vergleichende Histogramme als Qualitätskontrolle. In einem Microarrayexperiment wird erwartet, daß die Mehrheit aller Gene nicht reguliert wird, und daher in einem Histogramm, auf dem die logarithmischen Intensitäten der Farbstoffe dividiert werden, um den Nullpunkt verteilt sind. (Abb. 22). Für den in A dargestellten Array liegt dieses Verhältnis für deutlich weniger Gene direkt auf dem Nullpunkt als im Vergleich zu B und C, die in der Anzahl der nicht-regulierten Gene sehr ähnlich sind. Alle Arrays sind jedoch in der Mehrheit der Gene nicht deutlich reguliert; ein Hinweis auf eine gute Hybridisierungsqualität.

308 Gene sind in allen Datenpunkten mindestens zweifach hochreguliert. Darunter sind auch erwartete Zielgene der PcG, z.B. die Hoxgene *Abd-A* und *Ubx* sowie *Homothorax* und *l(2)giant larvae*, und CG5070 für die ebenfalls eine Regulation durch Polycomb angenommen wird (Ringrose et al. 2003).

Die anderen sechs Hoxgene bzw. weitere Gene, die unter der Kontrolle von Polycomb sind, konnten in diesem Datensatz nicht gefunden werden. Ebenso wurde Polycomb in diesem Datensatz nicht als reguliert eingestuft. Dies liegt daran, daß die Mutante durch einen Aminosäureaustausch nicht funktionsfähig ist, während die hybridisierte cDNA aufgrund der geringen Veränderung immer noch stringent bindet und kein Unterschied zwischen der Mutante und dem Wildtyp erwartet wird.

Die meisten der hochregulierten Gene (180/308) können keiner bestimmten Proteinfamilie zugeordnet werden, so daß über ihre Funktion keine Aussage gemacht werden kann. Von den 139 Genen, die Domänen aufweisen, konnten 120 einer Annotation in der Gene Ontology Datenbank (GO-Annotation) zugeordnet werden.

Bereits auf der obersten Ebene der GO Annotation, auf der die generellen molekularen Funktionen von Genen in Gruppen geordnet werden, wurden die Gene in vielfältige Funktionsgruppen eingeteilt (Abb. 23 und Tabelle 1 im Anhang). Insgesamt sind fast alle Gruppen der GO-Annotation in der Liste der regulierten Gene vertreten, wobei die prozentuale Anzahl der Gene in den einzelnen Gruppen zum Teil variiert. So sind in der Liste der hochregulierten Gene solche der Gruppe „structural molecule activity“ und „enzymatic regulator activity“ deutlich angereichert, wohingegen die Gruppen „obsolete function“ und „transcription regulator activity“ in dieser Liste etwas unterrepräsentiert sind.

Da bekannt ist, daß viele Transkriptionsfaktoren durch Polycomb reguliert werden, und diese Gruppe in den Genen, für die eine Regulation durch ein PRE vorhergesagt wurde, deutlich angereichert ist (41,24%), ist dieses Ergebnis verwunderlich. Eine Erklärung für diese Beobachtung kann sein, daß die Hochregulation dieser Transkriptionsfaktoren durch sekundäre Effekte beeinflusst wird und in der Auswertung der Microarrayexperimente nicht mehr erkennbar ist.

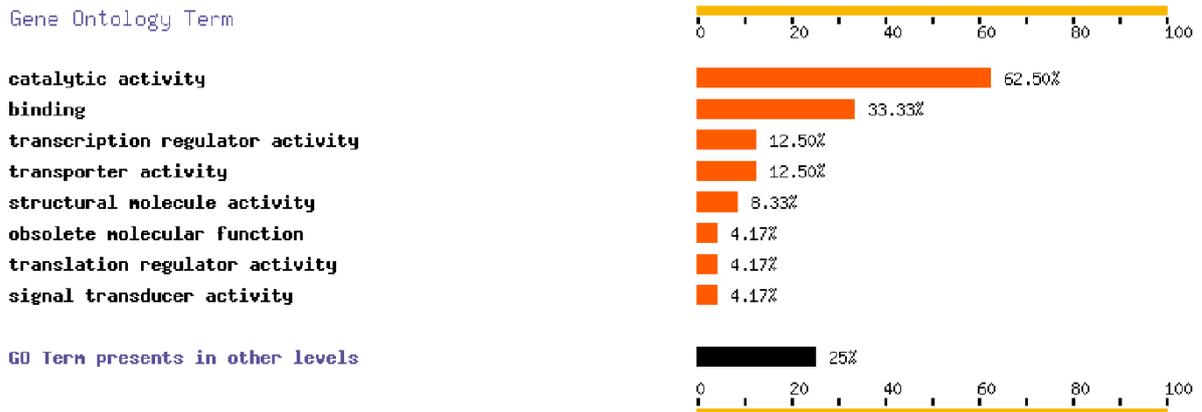
Insgesamt erfüllen die Gene, für die eine Regulation durch Pc gezeigt werden konnte, viele verschiedene molekulare Funktionen, so daß die Anzahl der verschiedenen GO-Annotations-Gruppen nicht verwunderlich ist. Obwohl eine endgültige Auswertung dieser Daten zu diesem Zeitpunkt nicht möglich ist, enthalten sie doch viel versprechende Kandidaten für weitere Experimente.



**Abb. 23 Die Einordnung der hochregulierten Gene in ihre potentielle molekulare Funktion.** Für 5,936 der 13,861 der von Berkeley in der Version 3.1 vorhergesagten Gene (A) sowie für 120 der auf dem Microarray hochregulierten Gene (B) konnte ein Eintrag in der GO-Datenbank fatigo gefunden werden. Obwohl nur auf der Ebene der Oberbegriffe der molekularen Funktionen gesucht wurde, ist bereits hier eine große Vielfalt an Funktionen für die regulierten Gene erkennbar. Die prozentuale Verteilung der Gene in den verschiedenen Gruppen unterscheidet sich zu der aller Gene. Dargestellt sind die Ergebnisseiten der Suchabfragen aus fatigo (<http://fatigo.bioinfo.cnio.es/>).

74 Gene sind in den Microarrayexperimenten herunterreguliert. Dies ist auf den ersten Blick erstaunlich, da die Funktion von Pc die Aufrechterhaltung des reprimierten Zustandes von Genen ist, und keine direkten Zielgene erwartet werden, deren Expressionsstatus in der Pc-Mutante verringert ist. Daher handelt es sich bei diesen Genen wahrscheinlich um sekundäre Zielgene. 49 der herunterregulierten Gene weisen eine Domäne auf, und 33 können einer GO-Annotation zugeordnet werden (Abb. 24). Auch für diese Gene ist eine ähnlich große Vielfalt der grundlegenden molekularen Funktionen zu finden, wobei die Gruppe „catalytic ativity“

bei diesen Genen prozentual überrepräsentiert ist. Obwohl einige Gene mögliche Transkriptionsfaktoren sind oder Chromatin-bindende Funktion haben (Tabelle Anhang), so trifft auch hier zu, daß man ohne die Analyse weiterer Daten keine definitive Aussage über diese Gene machen kann.



**Abb. 24 Die Zuordnung von GO-Annotationen zu den herunterregulierten Genen.** 33 Gene, die auf dem Microarray herunterreguliert waren, konnten mit Hilfe der Webseite fatigo in eine grundlegende molekulare Klasse eingeordnet werden. Dargestellt ist die Ergebnisseite der Suchabfrage aus fatigo (<http://fatigo.bioinfo.cnio.es/>).

## 2.12 Vergleiche mit weiteren Daten über Polycomb-Bindungsstellen

In der Vergangenheit wurden viele Experimente durchgeführt, um Informationen über die Zielgene von Pc und trx zu bekommen. Dazu gehörten Antikörperfärbungen auf polytären Speicheldrüsenchromosomen von *Drosophila* (Zink et al. 1991; Rastelli 1993) und *in-silico* Vorhersagen von PREs (Ringrose et al. 2003). Über diese Daten konnten etwa 100 Bindungsstellen für Polycomb gefunden werden. Trithorax und Ash1 sind Mitglieder der trxG, die spezifisch an der Regulation der PREs beteiligt sind. Für beide Proteine konnte ebenfalls eine Bindung an den Banden gefunden werden, an denen auch Pc bindet, sowie an 80 zusätzlichen Banden (Zink et al. 1991; Rastelli 1993). In Speicheldrüsen werden also insgesamt etwa 180 Regionen von diesen drei Proteinen gebunden und möglicherweise durch die PcG und die trxG reguliert. Die genomweite bioinformatische Analyse von PREs ergab Hinweise auf 167 PREs. Obwohl ein Vergleich dieser Daten mit den Microarraydaten

aufgrund der unterschiedlichen Auflösung der einzelnen Methoden schwierig ist, soll hier die allgemeine Korrelation der Daten überprüft werden.

128 der 308 hochregulierten Gene fallen in insgesamt 86 zytologische Regionen, an denen eine Bindung von Pc bzw. trx/Ash1 gezeigt werden konnte (Tab. 7). Dabei fallen 38 dieser Regionen auf eine Bande, an der eine Bindung für Polycomb vorliegt, wogegen für 48 Regionen eine Bindung von trx/Ash1 bekannt ist. Die hochregulierten Gene decken damit 38% der von Pc gebundenen Banden ab, sowie 47% aller mit PREs assoziierten Banden auf den Speicheldrüsenchromosomen. Im Gegensatz dazu konnte nur 14 Genen direkt ein vorhergesagtes PRE zugeordnet werden (Tab. 7). Dabei wurden jedoch nur Gene in der direkten Umgebung der PREs in die Untersuchung mit einbezogen. Da es unklar ist, welche Gene der Regulation des jeweiligen PREs unterliegen, ist es schwierig, hier eine Korrelation zwischen den PRE Vorhersagen und den potentiellen Zielgenen von Polycomb zu machen. Des Weiteren werden mit dem bioinformatischen Vorhersagen alle PREs im Genom abgedeckt, wohingegen es unklar ist, in welchen Geweben und zu welchen Zeitpunkten diese PREs eine regulatorische Funktion ausüben.

Für 180 Gene, die auf dem Microarray hochreguliert waren, konnte keine Assoziation zu Pc oder trx/Ash1 Bindungsstellen oder vorhergesagten PREs gemacht werden. Da aber nicht bekannt ist, wie viele Bindungsstellen für Pc für eine immun-histochemische detektierbare Bande notwendig sind, und die zytologische Kartierung nur größere Bereiche des Genoms darstellt, ist es möglich, daß einzelne PREs unentdeckt blieben. Darüberhinaus ist nicht bekannt, wie groß die Korrelation der Bindungsstellen in einem differenzierten Gewebe im Vergleich zu einem früheren Entwicklungsstadium ist. So konnte beispielsweise für das Gen *caudal* eine Regulation durch die PcG gezeigt werden (Ringrose et al. 2003). In der zytologischen Region 38E, in der *caudal* lokalisiert ist, kann aber auf Speicheldrüsenchromosomen keine Pc-Bande detektiert werden.

Eine komplette Überschneidung der existierenden Daten kann aus den angegebenen Gründen nicht erwartet werden. Dennoch ist, statistisch gesehen, eine Anreicherung von Banden, die mit einem PRE assoziiert werden, zu erkennen. Ausgehend von der Annahme, daß alle Gene gleichmäßig über die zytologischen Banden verteilt sind, wird erwartet, dass von 308 zufällig ausgewählten Genen 93 (30%) in eine Region fallen, in der eine Bande für Pc, trx oder Ash1 gezeigt werden konnte. Tatsächlich fallen 128 (41%) der 308 hochregulierten Gene in eine solche Region.

Obwohl die Daten noch weiter prozessiert und validiert werden müssen, und weitere Arrays für eine statistisch aussagekräftige Analyse der Zielgene von Polycomb hybridisiert werden müssen, geben diese Versuche erste viel versprechende Hinweise auf potentielle Zielgene und zeigen, daß mit der verwendeten Methodik eine genomweite Suche nach Pc-Zielgenen möglich ist.

NAME	InterPro	Zytologie	Pc- Bande	trx/Ash1-Bande	PRE
CG8974	Znf_ring	13E		A	
CG9914	3HCDH_N	14B	+	A	
CG9673	Chymotrypsin	15A		A	
RSG7	Regl_Gprotein	15A		A	
CG5070		15E	15F -	15F A	+
HDC19530		17A	+		
HDC19530		17A	+		
HDC19530		17A	+		
MgstI	MAPEG	19E	19 D +		
CG11686		1B		A	
CG13362	PRO_rich	1B		A	
EG:8D8.7	Transmem_4	1F		A	
I(2)gl	Lethal2_giant	21A	+	A	+
CG13678		21C	-	A	
CG33127	Ser_protease_Try	21C	-	A	
CG15361		22B	+	T	
CG17468	ANK	25D	25 E +		+
Vm26Ab		26A		AT	
CG13986	FA58_C	26B		AT	
CG5177	Trehalose_PPase	27F	+		
CG13102		29F		T	
CG9496	Transmem_4	29F		T	
CG9568		29F		T	
Try29F	Chymotrypsin	29F		T	
BcDNA:LD21503		2B	-	A	
Grip75	Spc97_Spc98	31E		31F A	
CG17127		32A		A	
Adh	Insect_adh_fam	35B	+	AT	+
ldgf2	2S_Globulin	36A	+	A	
Aac11		36C		A	
CG10621	S_methyl_trans	37B	+	T	
CG10746		38B		T	
CG9331	2-Hacid_DH	38F	+		
CG9338		38F	+		
CG9253	DEAD	39B		T	
CG32797		3A	-	A	
CG12129	KH_dom	46D	46C +	T	
CG12052	Znf_C2H2	47A		A	
CG3845	eIF5C	49D		A	
CG4679		49F	+	AT	+
Mp20	SM22_calponin	49F	+	AT	+
CG2934	ATPsynt_AC39sub	4A	-	A	
EG:EG0007.11	Thymosin_b4	4A	-	A	
CG3665	FnIII_subd	4B		A	
CG8092	Znf_C2H2	51E	-	A	
CG8157	Eggshell	51F	-	A	
CG11797	PBP_GOBP	56E		T	
Rad23	UBA_domain	56E		T	

Cdk9	Tyr_kinase	58F		A	
Blw	ATPase_a/bN	59A	+		
CG11716	Znf_CCHC	59B	+		
I(2)efl	Crystallin_alpha	59F	+		
CG2803	UPF0131	60E	+	A	
CG9358	A10_OS-D	60E	+	A	
CG7047	Arrestin	61A	+	A	
miple2	PTN_MK	61A/B	+	A	
HDC07758		61C	+	AT	+
HDC07759		61C	+	AT	+
Emc	HLH_basic	61D	+		
sp5	Serpin	61F	+	AT	
CG13805	Chitin_bind_PerA	62D	-	A	
CG13806	Chitin_bind_PerA	62D	-	AT	
CG16745	Chitin_bind_PerA	63B		A	
CG32280		63B		A	
CG18773	Insect_cuticle	64A	64B -	64B A	
CG1135	FHA	64B	-	A	
CG1259	Insect_cuticle	64B	-	A	
CG12607	NLS_BP	64B	-	A	
CG15022	PRO_rich	64B	-	A	
CG32248	Gyr	64B	-	A	
Chd64	SM22_calponin	64B	-	A	
CG18779		65A	-		+
CG8634	Insect_cuticle	65E	-	T	
CG8640	Insect_cuticle	65E	-	T	
CG8600		65F	-	T	
CG7892	Prot_kinase	66A	-	T	
CG18178		67B		A	
scylla		68B		A	
charybde		68C		A	
HDC09623		68C		A	
HDC09626		68C		A	
CG10116		70A	+	AT	
HDC09973		70A	+	AT	
Mlp84B	LIM	70A	+	AT	
CG8100	hemocyanin_N	70C		A	
CG9238	CBM_21	70E	+	AT	
Mpcp	Mitoch_carrier	70E/F	+	AT	
CG7724		73E			+
CG5567	Hydrolase	75A	-	A	
Rpr		75C	74D -	A	
CG8748	LDL_receptor_A	76C	+		
CG8756	LDL_receptor_A	76C	+		
CG14573		78F	+		
CG14569		78F/79A	+		
CG14656		82D	+		
Ccp84Ac	Insect_cuticle	84A	+	AT	
CG14598		84D	+	T	
CG7878	KH_dom	84E	+	A	
CG14568		84F	+	A	
CG12418		85E	+	A	
HDC12717		85E	+	A	
Hth	Homeobox	86B	86C +	T	+
CG4511	Thioredox_dom2	86C	+	T	
CG6584	DUF25	86C	+	T	
Fau		86C	+	T	
Fer2LCH	Ferritin	86C	+	T	

CG17738		87B	+		
CG4115	Lectin_C	87B	+		
Vha55	ATPase_a/bcentre	87C	+		
CG8031	DUF52	87D	-	A	
CG15887		87E	-	A	
HDC13470		87F	+	A	
Rpl140	RNA_pol_Rpb2_6	88A	+	A	+
CG6803		88E	-	A	
Hsc70Cb	Hsp70	88E	-	A	
Abd-A	Homeobox	89E	+	AT	+
Ubx	Homeobox	89E	+	AT	+
CG7267		8C	-	AT	
Edg91	Eggshell	90B		A	
CG12349	PDZ	90D		A	
CG14327	Eggshell	90D		A	
CG31243	RNA_rec_mot	90D		A	
CG14290	UPF0041	91D		A	
HDC14405		91E		AT	
Nrv1	Na/K_ATPase_beta	92C/27B	-	A	
Mod	RNA_rec_mot	93D	-	T	
CG11786	P_rich_extensn	96B	+		
CG11938		96B	+		
CG13638	Chymotrypsin	96B	+		
OstStt3	Oligo_trans_STT3	96B	+		

**Tabelle 7 Vergleich der potentiellen Pc-Zielgene mit den Bandenmustern von Pc, trx und Ash1 auf Polytänschrososomen sowie mit vorhergesagten PREs.** Die Gennamen und die vorhandenen Interpro-Einträge der regulierten Gene, die in einer zytologischen Region liegen, in der eine Bindung von Pc bzw. von trx (T) oder Ash1 (A) auf polytänen Speicheldrüsenchromosomen gezeigt werden konnte, sind dargestellt. Zusätzlich sind die Gene aufgelistet, für die eine Assoziation mit einem PRE vorhergesagt wurde.

### 3 Diskussion

#### **3.1 Beiträge der einzelnen Mitarbeiter an der Herstellung und Validierung des Transkriptomarrays**

Der Vergleich verschiedener Genvorhersagen und die Herstellung und Validierung eines neuen DNA-Microarrays erfordert die Zusammenarbeit verschiedener Gruppen mit unterschiedlichen Expertisen. Innerhalb dieser Arbeit waren drei Gruppen an dem Projekt beteiligt. Die unterschiedlichen Genvorhersagen wurden in der Gruppe von Prof. Dr. Martin Vingron am DKFZ durch Dr. Stefan Haas verglichen und zu einer neuen Annotation zusammengefügt. Die Überprüfung des Algorithmus' wurde in den Gruppen von Prof. Dr. Martin Vingron und Prof. Dr. Renato Paro durch Dr. Stefan Haas und Dr. Marc Hild durchgeführt. Das Primerdesign wurde ebenfalls von Dr. Stefan Haas durchgeführt. Die berechneten Primer wurden stichprobenweise von Dr. Marc Hild und mir überprüft. Die synthetisierten Primer wurden in der Gruppe von Prof. Dr. Renato Paro zu jeweiligen Primermixen zusammengefügt. Ein PCR Mastermix für alle PCRs wurde ebenfalls von Dr. Marc Hild und mir angesetzt. Die dazu benötigten Mengen an genomischer DNA als Template wurden im Vorfeld von mir isoliert. Der vorbereitete Mastermix wurde am DKFZ in der Gruppe von Dr. Jörg Hoheisel mit Taq-Polymerase versetzt, die PCR durchgeführt und auf Gelen auf die richtige Produktlänge überprüft. Die zweite PCR wurde sowohl in der Gruppe von Dr. Hoheisel durch Boris Beckmann als auch in der Gruppe von Prof. Dr. Paro durch Dr. Marc Hild und mir zusammen mit Sylvia Hagar durchgeführt. Die Microarrays für den ersten Teil dieser Arbeit wurden in der Gruppe von Dr. Jörg Hoheisel gespottet, während die Arrays für den zweiten Teil in der Gruppe von Prof. Dr. Paro hergestellt wurden. Die Kontrollen für die jeweiligen Arrays wurden dabei parallel in beiden Laboren kloniert.

Die verschiedenen Stadien für das Entwicklungsprofil wurden von mir auf das richtige Alter überprüft und gesammelt, und die RNA zusammen mit Sylvia Hagar isoliert und auf ihre Qualität überprüft. Die RNA wurde in der Gruppe von Dr. Hoheisel für die Probensynthese verwendet und auf die Microarrays hybridisiert. Die Auswertung der Microarrays erfolgte ebenfalls in der Gruppe von Dr. Hoheisel durch Boris Beckmann, unterstützt durch Dr. Marc Hild. Die biologischen Validierungen mittels RT-PCR wurden von Dr. Marc Hild durchgeführt. Die weiteren Untersuchungen über *in-situ* Hybridisierungen und nachfolgende Analysen wurden von mir im Labor von

Prof. Paro durchgeführt. An der Herstellung der Webseiten waren Dr. Stefan Haas, Dr. Marc Hild und ich beteiligt.

### **3.2 Genvorhersagen für *Drosophila melanogaster***

Im Jahr 2000 wurden die euchromatischen Regionen von *Drosophila melanogaster* in einem gemeinsamen Projekt des BDGP/EDGP und Celera Genomics Inc. vollständig sequenziert (Adams et al. 2000). Gleichzeitig wurden die Transkript- und Proteinsequenzen in einer ersten Annotation vorhergesagt. Für diesen Ansatz wurden zwei verschiedene Genvorhersageprogramme verwendet: Genscan (Burge und Karlin 1997) und eine Version von Genie, die zusätzlich EST-, cDNA- und Proteindaten für die Vorhersage einschloß (Reese et al. 2000). Genscan sagte in diesem Versuch 17,464 Gene vorher, während Genie auf eine Zahl von 13,189 Genen kam. Nach Studien an der gut untersuchten *Adh* Region mit beiden Programmen nahmen die Autoren die Anzahl der Gene von Genie als die akkuratere an (Reese et al. 2000). Die Tatsache, daß Genie für *Drosophila* optimiert wurde, während es für Genscan keine spezifischen Parameter gab, unterstützte diese Annahme.

Unter Berücksichtigung beider Programme wurden in dieser Annotation 13,601 Gene vorhergesagt, die aufgrund von alternativem Spleißen durch 14,113 Transkripte repräsentiert werden. Die Autoren diskutierten gleichzeitig, daß die Anzahl der alternativen Spleißformen um ein wesentliches größer sein würde, nur fehlten aufgrund der geringen Menge an ESTs die notwendigen Grundlagen für eine computergestützte Vorhersage.

Die Qualität dieser ersten Vorhersage wurde bald in Frage gestellt. So ergab ein Vergleich der veröffentlichten Daten mit denen der publizierten Proteinsequenzen für etwa 45% der Gene eine deutliche Diskrepanz in der Sequenz, verursacht durch Mißmatches, Insertionen und Deletionen (Karlin et al. 2001).

In den nachfolgenden Annotationen änderte sich die Anzahl der Gene kaum. Wurden in der ersten Version 13,601 Gene vorhergesagt, so waren es in der zweiten 13,474 und in der dritten Version 13,676 Gene, von denen 13,379 Gene in der euchromatischen Region und 297 innerhalb des Heterochromatins liegen (Misra et al. 2002). Was sich jedoch änderte, war die Genstruktur. In der ersten Version wurden durchschnittlich 4,1 Exons pro Gen vorhergesagt, in der dritten Version

waren es 4,6 Exons, was der durchschnittlichen Anzahl von Exons in *A. thaliana* (Lin et al. 1999) und *C. elegans* (Consortium 1998) entspricht. Dieser Anstieg der exonischen Elemente war jedoch nicht durch veränderte oder bessere Programme begründet, sondern durch die Analyse von EST's und Vollängen-cDNAs. Diese Analysen sind derzeit die einzigen Möglichkeiten, um nicht-translatierte-Regionen (UTRs) zu annotieren.

Für jedes Gen mußten zusätzlich verschiedene Kriterien zutreffen, damit es in die Vorhersage aufgenommen wurde. Dazu gehörten das Vorhandensein von cDNAs und/oder ESTs, Proteindomänenvergleiche, Sequenzähnlichkeiten zu verwandten Arten und eine bestimmte Mindestlänge der ORFs. Während durch diese Stringenz bekannte Gene mit einer hohen Sicherheit richtig vorhergesagt werden können, werden zum Teil Gene, für die weniger Daten generiert wurden, aus der Vorhersage ausgeschlossen.

Im Gegensatz zu diesen stringenten Kriterien basiert die Heidelberg Annotation auf einer *a priori* Vorhersage von offenen Leserastern bei der keine derart stringente Mindestlänge für Proteine gesetzt, und keine weiteren Informationen über Domänen oder cDNAs einbezogen wurden. Dadurch wurden auch Gene vorhergesagt, die keine bekannten Domänen aufweisen oder von denen keine cDNAs und ESTs existieren. Auch Gene, die nur aus einem kurzen Exon bestehen, und damit unter die von der Berkeley Annotation vorausgesetzte Mindestlänge fallen, konnten mit diesem Ansatz erfasst werden. Die erwartete große Anzahl der durch diese offenen Kriterien falsch positiv vorhergesagten Gene führt zu der Notwendigkeit, die Gene in biologischen Versuchen zu validieren, um aus der Kombination beider Methoden eine sinnvolle Annotation zu generieren.

### **3.2.1 Eine Abschätzung der Qualität der Genmodelle**

Die Gene, die vom BDGP in der Version 3.1 vorhergesagt wurden, wurden mit Hilfe von weiteren Informationen auf ihr Genmodell überprüft. Daher sind die Genmodelle in den meisten Fällen sehr genau. In den Fällen, in denen die Gene neu vorhergesagt wurden, und wenige Informationen zur Verfügung standen, sind die Genmodelle allerdings unsicherer. Häufige Fehler sind zum Beispiel die falsche Assoziation von Exons, falsches spleißen von Genen sowie das Fehlen von Mini-

oder Microexons (Misra et al. 2002). Auch gab es Schwierigkeiten, Gene in Introns anderer Genen vorherzusagen.

Da die Genvorhersagen der Heidelberg Collection keine EST- oder cDNA Daten berücksichtigten, konnten die nicht-translatierten 5'- und 3'-Regionen nicht vorhergesagt werden, so daß die durchschnittliche Anzahl der Exons denen der Berkeley Version 1 entspricht, die ebenfalls diese Daten nicht in ihre Vorhersagen einbezogen hatte. Auch wurden lange Introns oft als Grenze zwischen Genen bewertet, so daß bekannte Gene häufig als zwei getrennte Gene vorhergesagt wurden. Dieser Fehler wird auch bei den neu vorhergesagten Genen aufgetreten sein. Im Gegensatz dazu werden Gene innerhalb von Introns häufiger erkannt, und obwohl auch in dieser Vorhersage Exongrenzen oft falsch vorhergesagt wurden, scheinen mehr alternative Exons erkannt worden zu sein. Auch Gene, die nur aus einem einzelnen Exon bestehen, wurden mit diesem Programm häufiger vorhergesagt als mit den anderen Methoden.

Insgesamt sind die bisherigen Genmodelle noch unsicher, und nur biologische Analysen werden nach und nach das Bild der Gene und deren alternativen Varianten verbessern und komplettieren. So wurde zum Beispiel von Reboul et. al, für *C. elegans* gezeigt, daß für mehr als 50% der vorhergesagten Gene eine Berichtigung der Intron-Exon Strukturen vorgenommen werden muß (Reboul et al. 2001; Reboul et al. 2003).

### **3.3 Der Heidelberg Transkriptom-Microarray**

Innerhalb dieser Arbeit wurde der HeidelbergFlyArray entworfen, der mit insgesamt 21,396 verschiedenen Genfragmenten 92,7% der BDGP/Celera Genomannotation 3.1 abdeckt. Zusätzlich wurden Amplikons für 6,224 weitere potentielle Gene gespottet. Damit ist der HeidelbergFlyArray derzeit einer der vollständigsten Transkriptomarrays für *Drosophila melanogaster*.

#### **3.3.1 Die Fragmente für den HeidelbergFlyArray**

Alle auf dem Microarray repräsentierten Fragmente wurden durch PCR amplifiziert. Dabei lieferte das verwendete Tag-System die Möglichkeit, den gesamten Satz mit

wenigen Primern zu reamplifizieren. Dieses System erlaubt zum einen, durch eine Auswahl der Primer verschiedene weitere Tags für unterschiedliche Anwendungen anzuhängen (siehe 5.6.3). Zum anderen ermöglicht es die Produktion großer Mengen der einzelnen Fragmente mit wenig Pipettieraufwand und weniger Verbrauch an Material, so daß viele nachfolgende PCRs zur Generierung von Fragmenten für den Microarray zeit- und kostengünstiger durchgeführt werden können.

Insgesamt konnten in der PCR über 98% der Fragmente erfolgreich amplifiziert werden. Diese hohe Erfolgsrate ist auf die Homogenität der Primer und der resultierenden Fragmentlängen zurückzuführen. Für PCR-Resultate vergleichbarer High-Throughput Projekte, wie zum Beispiel des Incyte Arrays, wurden Erfolgsraten zwischen 92% - 95% berichtet (M. Pankratz, B. Oliver, persönliche Mitteilungen).

Alle PCR Fragmente wurden ohne weitere Aufreinigungsschritte gespottet. Vergleiche zwischen aufgereinigten und nicht aufgereinigten PCR Produkten hatten im Vorfeld ergeben, daß unspezifische Hybridisierungen oder Verunreinigungseffekte durch nicht entfernte Oligonukleotide ausgeschlossen werden konnten (Diehl et al. 2002).

Auch die Produktmenge, die in den einzelnen Wells vorhanden war, wurde nicht weiter überprüft. Die DNA wurde in einer durchschnittlichen Konzentration von 150 – 350 ng/µl gespottet. Studien mit Microarrays ergaben, daß in diesem Konzentrationsbereich die exakte DNA-Menge in einem Spot nicht von Relevanz ist, wenn der Spottinglösung Betain hinzugefügt wurde (Diehl et al. 2001). Die Zugabe von Betain zu der Spottinglösung führt zu besseren Bindungseigenschaften der Probe auf dem Glas und einer größeren Spothomogenität. Dies ist vermutlich auf die physikalischen Eigenschaften von Betain zurückzuführen, die eine erhöhte Viskosität und eine verminderte Verdunstungsrate bewirken.

Hybridisiert man den Microarray zur Kontrolle mit genomischer DNA, die mit beiden Farbstoffen markiert wurde, so kann man bei Betrachtung der visuellen Ausgabe der Spotintensitäten deutliche Unterschiede in der Gesamtintensität der Spots feststellen. Da aber in Zwei-Farben Microarrayexperimenten nur die beiden Farben innerhalb eines Spots miteinander verglichen werden, ist die beobachtete Divergenz nicht von Relevanz für das eigentliche Experiment.

### 3.3.2 Der Versuchsaufbau für das Entwicklungsprofil

Die meisten cDNA Microarrays werden heute durch den Vergleich der zu untersuchenden Probe mit einer gemeinsamen Kontrolle analysiert, indem das Verhältnis der Intensitäten des Farbeinbaus der zu untersuchenden Probe zu dem der Kontrolle für jedes Gen berechnet wird (Eisen und Brown 1999). Dabei ist es erforderlich, in der Kontrolle so wenig Varianz wie möglich zu haben.

Auch der hier verwendete Versuchsaufbau unterlag diesem Schema. Für das Entwicklungsprofil wurde der Lebenszyklus von *Drosophila* in neun Stadien aufgeteilt. Um die Durchführbarkeit des Experimentes zu gewährleisten, wurde dabei die RNA aus sechzehn Stadien zu diesen neun Stadien vereinigt. Acht dieser Stadien wurden jeweils gegen die Kontrolle hybridisiert. Um den gesetzten Anforderungen an ein Microarrayexperiment gemäß MIAMI (Minimal Information About a Microarray Experiment) zu genügen (Brazma et al. 2001), wurden die Hybridisierungen für jedes Stadium viermal wiederholt und dabei in mindestens einer dieser Wiederholungen einem Farbstofftausch zwischen Probe und Kontrolle unterzogen. Hybridisierungen von minderwertiger Qualität wurden dabei nicht in die Auswertung einbezogen. Insgesamt wurden daher mehr als 40 Microarrays hybridisiert. Gene wurden dann als exprimiert angesehen, wenn in mindestens einem Stadium eine Expressionsveränderung zur Kontrolle bzw. zu anderen Stadien nachweisbar war. Gene, die ubiquitär und in der gleichen Stärke durch den Lebenszyklus hindurch vorhanden exprimiert sind, konnten in diesem Versuchsaufbau nicht berücksichtigt werden.

Die Zusammenfassung verschiedener Stadien kann dazu geführt haben, daß Gene, die in nur einem dieser Stadien exprimiert sind, unter die Detektionsgrenze gefallen sind und daher in den weiteren Berechnungen nicht berücksichtigt wurden. Dennoch war die Aufteilung des Lebenszyklus in mehr Stadien aus technischen Gründen nicht durchführbar.

Als Kontrolle für die Hybridisierungen diente RNA aus den ersten vier Stunden der Embryonalentwicklung, die gegen alle anderen Zeitpunkte hybridisiert wurde. In einem anderen Microarrayexperiment, in dem ebenfalls die Expression der Gene während des Lebenszyklus von *Drosophila* analysiert wurde, diente eine Mischung aller Stadien als Kontrolle (Arbeitman et al. 2002). Allerdings ist die Reproduzierbarkeit eines gleichen Gemisches aus 16 Proben immer mit Fehlern behaftet. Diese Unsicherheit wurde durch die Wahl einer Kontrolle aus nur einem Stadium vermieden.

### 3.3.3 Die Ergebnisse des Expressionsprofils

Alle Microarrays wurden mit Hilfe der ScanArray Software und anschließender Analyse mittels MChIPS (Fellenberg et al. 2002), ausgewertet. Insgesamt konnte die Expression für 9,908 (78,3%) der in der BDGP Version 3.1 vorhergesagten Gene gezeigt werden. Von den neu vorhergesagten Genen der Heidelberg Collection konnte für 2,646 (42,9%) eine Expression nachgewiesen werden. Zusätzlich enthält der Array 286 Vorhersagen, die nur in der BDGP Version 2 vorhergesagt wurden, und die später nicht mehr in die Vorhersagen einbezogen wurden. Von diesen Genen konnten 160 (55%) über die Microarrayergebnisse bestätigt werden.

Die in diesen Experimenten erhaltenen Ergebnisse stimmen sehr gut mit Studien überein, die für einem Teil der vorhergesagten menschlichen Gene durchgeführt wurden (Shoemaker et al. 2001). Die Autoren zeigen dabei, daß etwa 80-85% der bekannten Gene, aber nur 58% der vorhergesagten Gene durch die Microarrayexperimente validiert werden konnten. Es ist daher durchaus wahrscheinlich, daß die Expressionsstärke der etwa 22% der BDGP Gene, die über den Microarray nicht validiert wurden, unter der Detektionsgrenze von Microarrayanalysen liegt.

### 3.4 Wie viele Gene gibt es in *Drosophila melanogaster*?

Viele Ansätze geben Hinweise darauf, daß die Anzahl der *Drosophila* Gene größer ist als bisher angenommen. Gopal et. al fanden in einer Homologie-basierten Annotation Hinweise auf 1,042 neue Gene (Gopal et al. 2001). Etwa 300 dieser Gene sind auch in der neuesten Annotation von BDGP nicht berücksichtigt (Misra et al. 2002). In einem weiteren Ansatz wurde eine große Anzahl an Testis-spezifischen ESTs untersucht. Einige dieser ESTs liegen in bisher nicht berücksichtigten 5'-Regionen von bekannten Genen, und liefern alternative, Testis-spezifische Genprodukte. Viele dieser ESTs liegen jedoch in Bereichen, in denen bisher kein Gen vorhergesagt wurde (Andrews et al. 2000).

Innerhalb der letzten Jahre wurden vermehrt Vergleiche zu anderen *Drosophila* Arten und weiter entfernten Organismen durchgeführt. Erste Ergebnisse aus diesen vergleichenden Studien deuten auf die Existenz von mindestens 500 weiteren Genen hin (Richards 2004).

Im Verlauf dieser Arbeit wurden Genvorhersagen miteinander verglichen, überprüft und analysiert. Dabei konnten über Microarrayanalysen, RT-PCR und *in-situ*

Hybridisierungen Hinweise auf etwa 2,000 neue Gene gefunden werden. Aufgrund der unsicheren Genmodelle war es aber fraglich, ob alle diese Gene wirklich eigenständig oder mit anderen Genen assoziiert sind. Weitere Hochdurchsatz-Ansätze mit den Amplikons der Heidelberg Collection gaben hilfreiche Hinweise zur Beantwortung dieser Frage.

### **3.4.1 Ein genomweiter RNAi Screen stützt die Ergebnisse des HeidelbergFlyArrays**

Wie bereits dargestellt, kann das Design des HeidelbergFlyArrays für weitere Anwendungen nützlich sein. So wurde der Satz an PCR-Fragmenten in Kollaboration mit Michael Boutros im Labor von Norbert Perrimon (Harvard Medical School, Boston) ebenfalls für die Herstellung eines genomweiten RNAi Ansatzes verwendet (Boutros et al. 2004). Für jedes Fragment wurde über *in-vitro* Transkription doppelsträngige RNA produziert, und in Zellkulturexperimenten getestet. Als erste Serie von Experimenten wurde getestet, wie viele Gene zum Tod der Zellen führen. Diese Experimente bildeten die Basis für weitere, verfeinerte Screens.

Die Experimente zeigen 437 dsRNAs, die zu Letalität führen. 68 dieser dsRNA stammen von Fragmenten neuer Gene. Wenn man annimmt, daß die prozentuale Anzahl der Gene, die zu Letalität führen, in beiden Genvorhersagen gleich bleibt, und die Gene in der Berkeley Vorhersage als gesichert annimmt, so kann man über diese Zahlen eine Voraussage über die Anzahl neuer Gene machen. 369 (2,9%) der 12,655 Berkeley Gene und 68 (1,1%) der ursprünglich vorhergesagten 6,143 HDC-Gene zeigen Letalität. Zusammengefasst sprechen diese Daten dafür, daß 2,330 (38%) der neuen Gene exprimiert sind. Diese Zahl stimmt gut mit den aus den anderen Experimenten gewonnenen Daten überein und stützt daher die bisherigen Ergebnisse.

### **3.4.2 Sind die neuen Gene zusätzliche Exons von bereits bekannten Genen?**

Die Gene, die in dieser Arbeit näher untersucht wurden und innerhalb von Introns anderer Gene liegen, sind in einigen Fällen nachweislich Spleißvarianten bekannter

Gene. Um aber eine Abschätzung zu machen, wie viele dieser neuen Genvorhersagen mögliche Exons bekannter Gene sind, wurden erneut die Ergebnisse der genomweiten RNAi Experimente verwendet. Dazu wurde untersucht, wie viele der Berkeley Gene, die Letalität zeigten, ein benachbartes HDC-Gen haben, welches ebenfalls Letalität zeigte. Innerhalb von 50 kb auf- oder abwärts von diesen Genen wurden 14,7% der neuen Gene gefunden, die ebenfalls Letalität zeigten, und damit ein neues Exon des jeweiligen Nachbargens sein könnten. Bei etwa 5% der untersuchten Gene zeigten zwei benachbarte HDC-Gene Letalität, es könnte sich hierbei also auch um dasselbe Gen handeln. Vergleicht man diese Daten nun mit denen von Berkeley vorhergesagten Genen, so kann man keine Unterschiede feststellen. Hier zeigten 16% der benachbarten Gene ebenfalls Letalität, und könnten so als mögliche weitere Exons interpretiert werden. Insgesamt geben diese Daten keinen Hinweis darauf, daß es sich bei den neuen Genen mehrheitlich um zusätzliche Exons anstatt um eigenständige Gene handelt.

### 3.4.3 Wurden dennoch Gene übersehen?

Alle bisherigen Daten sprechen dafür, daß nur eine Minderheit der 2,000 neu vorhergesagten Gene Teile von bereits vorhergesagten Genen sind. Eine große Anzahl der Gene fehlt also in der Berkeley Genannotation 3.1.

Gleichzeitig beruhen die durchgeführten Validierungen auf der Annahme, daß jedes Gen während des Lebenszyklus von *Drosophila* mindestens einmal exprimiert ist. Nur die Gene, die in den Microarrayexperimenten beobachtet worden waren, wurden in die weiteren Berechnungen und Validierungen mit aufgenommen. Dabei wurde nicht berücksichtigt, daß viele Gene nur in besonderen Situationen, wie zum Beispiel Stress durch Hitze, Kälte oder Hunger exprimiert werden.

Wie bereits beschrieben, kann der Aufbau des Entwicklungsprofils ebenfalls dazu geführt haben, daß Gene übersehen wurden. Da der gesamte Lebenszyklus in nur neun Stadien zusammengefasst wurde, ist es sehr wahrscheinlich, daß Gene, die nur während einer kurzen Zeitspanne exprimiert werden, durch die Mischung vieler verschiedener Stadien nicht mehr detektierbar sind. Auch Gene, deren Expressionsniveau sich über den gesamten Lebenszyklus nicht verändert, können mit dem hier

gewählten Ansatz nicht nachgewiesen werden. Eine Abschätzung der aus diesen Gründen nicht mit einbezogenen Genen ist derzeit allerdings unmöglich.

Ein weiterer wichtiger Aspekt ist, daß alle verwendeten Genvorhersageprogramme auf ORF-Vorhersagen basieren. Gene, die nicht für Proteine kodieren, sondern zum Beispiel für regulatorische RNAs, wurden bisher nicht vorhergesagt und nur wenige Ausnahmen sind der Berkeley Annotation manuell hinzugefügt worden (Misra et al. 2002).

Insgesamt kann aufgrund der in dieser Arbeit durchgeführten Analysen zusammen mit den Beobachtungen in anderen Laboren davon ausgegangen werden, daß eine große Anzahl an Genen in der Berkeley Genannotation fehlt. Über die genaue Anzahl der fehlenden Gene kann derzeit nur spekuliert werden. Experimente mit Microarrays, welche die gesamte genomische Sequenz abdecken (White, *Drosophila* Conference), werden in naher Zukunft viele weitere Experimente ermöglichen, die Hinweise auf die wirkliche Anzahl an Genen sowie deren Spleißformen geben werden.

### **3.5 *In-situ* Hybridisierungen**

Eine Auswahl an Genen wurde über *in-situ* Hybridisierungen validiert. Dabei konnte für viele der neuen Gene eine spezifische Expression in einzelnen Geweben beobachtet werden. Eine große Anzahl an Genen (62%) zeigte jedoch keine detektierbare Expression. Daß viele dieser Gene dennoch exprimiert wurden, konnte beispielhaft über RT-PCRs bestätigt werden. Zwei Gründe können für diese Diskrepanz angeführt werden: zum einen spielt die Länge der Sequenz, die als Probe für die *in-situ* Hybridisierung verwendet wird, eine Rolle für die Intensität der Färbung. Da die durchschnittliche Länge der verwendeten Fragmente nur 456 Basenpaare beträgt, kann die kritische Länge für ein detektierbares Signal bei einigen Genen unterschritten worden sein. Zum anderen kann über *in-situ* Hybridisierungen zwar mit hoher Spezifität die Genexpression in bestimmten Geweben oder Zellen nachgewiesen werden. Handelt es sich jedoch um eine schwache, ubiquitäre Expression, so kann diese nicht von einer Hintergrundfärbung unterschieden werden. Da die Hybridisierungen mit Kontrollgenen eine ähnliche Erfolgsrate aufwiesen, und für eine Kontrolle die Mindestlänge der Probe mit 350

Basenpaaren festgelegt werden konnte, kann man davon ausgehen, daß die beiden angeführten Gründe zu der geringen Erfolgsrate der Hybridisierungen geführt haben.

### 3.5.1 Neue Spleißformen für bekannte Gene

Für einige der neu vorhergesagten Gene, die innerhalb eines Introns vorhergesagt wurden, konnte gezeigt werden, daß es sich um Spleißformen anderer Gene handelt, und nicht um eigenständige Gene. Zwei dieser Beispiele sind HDC06000 und HDC03057. HDC03057 liegt im Intron des Gens *Fasciclin 3*, einem Gen, das in der Vergangenheit molekular und genetisch intensiv untersucht wurde (Snow et al. 1989). Das beobachtete Expressionsmuster von HDC03057 weicht stark von der Expression, die für *Fas3* gezeigt wurde, ab. Diese Diskrepanz in der Expression kann darauf hindeuten, daß die Spleißform, die HDC03057 einschließt, eine andere Funktion ausübt als die bisher beschriebenen Formen. Da die *in-situ* Hybridisierung für *Fas3* aber mit einer cDNA gemacht wurde, die auch das gemeinsame Exon mit einschließt, ist die unterschiedliche Expression nicht direkt erklärbar. Letztendlich wird nur durch weitere molekulare Analysen eindeutig erwiesen werden können, wie das Genmodell von HDC03057 und *Fas3* tatsächlich aussieht, und welche molekularen Funktionen die alternativen Formen erfüllen.

HDC06000 liegt im Intron des ebenfalls molekular untersuchten Gens *Lachesin* (Karlstrom et al. 1993; Llimargas et al. 2004), und auch hier gibt es Hinweise darauf, daß es sich bei HDC06000 um eine Variante dieses Gens handelt. Im Gegensatz zu HDC03057 konnten die Exon-Exon Grenzen für diese eventuelle Spleißform noch nicht eindeutig durch Sequenzierung bestimmt werden.

Das beobachtete Expressionsmuster zeigt, daß HDC06000 während frühen Stadien in den gleichen Bereichen exprimiert wie die bereits bekannten Varianten von *Lac*. Daß diese mögliche alternative Form bis jetzt nicht beobachtet wurde, kann eventuell daran liegen, daß sie schwächer exprimiert wird als die übrigen, und daher nur mit Hilfe einer normalisierten cDNA Bibliothek hätte gefunden werden können.

Schaut man sich die Gruppe der neu vorhergesagten Gene an, die innerhalb von Introns anderer Gene liegen, so findet man eine große Anzahl an Genen, die nur aus einem Exon bestehen. Die Wahrscheinlichkeit, daß diese Gene mit anderen Genen assoziiert sind, ist dementsprechend hoch. Um einen Eindruck zu bekommen, wie

viele der neu vorhergesagten Gene innerhalb eines Introns liegen und damit eventuell Spleißformen anderer Gene sind, wurden die exprimierten neu vorhergesagten Gene auf ihre Lokalisation untersucht. Insgesamt liegen 350 der neu vorhergesagten Gene, die mit den durchgeführten Experimenten validiert werden konnten, innerhalb Introns anderer Gene. Wie viele von diesen tatsächlich Varianten von anderen Genen sind, wird nur durch genaue molekulare Analysen erforscht werden können. Auf die Gesamtzahl der Gene in *Drosophila* würden diese potentiell zu fusionierenden Gene keinen großen Einfluß nehmen.

### **3.5.2 Spielt HDC10292 eine Rolle in der frühen Entwicklung des zentralen Nervensystems?**

HDC10292 zeigt eine Expression, die der des Gens *ind* (Weiss et al. 1998) auffällig ähnlich ist. Im Verlauf dieser Arbeit konnte gezeigt werden, daß HDC10292 unter der Kontrolle von *vnd* liegt, einem Gen, daß ebenfalls während der Entwicklung des zentralen Nervensystems (ZNS) eine Rolle spielt. Ob diese Kontrolle aber direkt oder indirekt, eventuell über *ind* ausgeübt wird, ist bisher nicht geklärt.

Auch die Funktion von HDC10292 konnte bisher nicht gezeigt werden. Aufschluß darüber wird eventuell die Analyse der Expression des Gens *msh* (muscle segment homeobox) in HDC10292 mutanten Embryonen geben können. Ähnlich wie Gene der ventralen Reihe von Neuroblasten die Expression von Genen der mittleren Reihe von Neuroblasten regulieren, wurde *ind* als erstes Gen der mittleren Neuroblastenreihe beschrieben, welches die Expression von *msh* (muscle segment homeobox) auf die dorsalen Neuroblasten beschränkt (Cornell und Ohlen 2000; Koizumi et al. 2003). In *ind* Mutanten dehnt sich die Expression von *msh* auf die mittlere Neuroblastenreihe aus (Weiss et al. 1998). Ob die fehlende Expression von HDC10292 ausreichend ist, um zu einer vergleichbaren Fehlexpression von *msh* zu führen, kann über die oben angesprochene Analyse mit Hilfe von Defizienzen, die HDC10292 deletieren, sowie auf solchen, die sowohl HDC10292 als auch *ind* deletieren, untersucht werden. Färbungen von Embryonen dieser Defizienzen mit einem Antikörper gegen das Gen *evenskipped*, mit dem alle Neurone erkannt werden, werden darüber hinaus klären, in welchen Neuronen HDC10292 exprimiert ist. Die verfügbaren Defizienzen aus dieser Region sind sehr klein und führen alle zur

Letalität in embryonalen Stadien. Aufgrund der Tatsache, daß immer noch andere Gene in diesen Linien deletiert wurden, kann aber nicht auf eine lebensnotwendige Funktion von HDC10292 geschlossen werden.

Daß HDC10292 eine Funktion in den Neuronen ausübt, scheint zu diesem Zeitpunkt bestätigt. Inwieweit dieses Gen allerdings für die korrekte Entwicklung des ZNS notwendig ist, kann nur durch weiterführende Experimente geklärt werden.

### 3.5.3 Mögliche Funktionen von HDC13470

Eines der interessantesten Gene, die über die *in-situ* Hybridisierungen validiert wurden, ist das Gen HDC13470. cDNAs und ESTs sind für diese Region bereits bekannt, und über einen Northern Blot konnte die Länge der cDNA bestätigt werden; ein eindeutiges ORF existiert in dieser Region jedoch nicht. Aus diesem Grund wurde zunächst vermutet, daß es sich bei HDC13470 um eine nicht-translatierte RNA handelt. Untersuchungen der DNA-Sequenz zeigten, daß die Region in anderen *Drosophila* Spezies nur in kleinen Bereichen konserviert ist. Bei genauerer Analyse der potentiellen Proteinsequenz wurde jedoch festgestellt, daß vier kurze ORFs in unterschiedlichen Leserastern in der Sequenz vorkommen, in denen insgesamt fünfmal eine nahezu identische Sequenz auftaucht. Gefolgt werden diese Sequenzen von einem 50 Aminosäure enthaltenden ORF. Die Tatsache, daß alle fünf Peptidsequenzen zu 100% mit konstantem Abstand voneinander und in den gleichen Leserastern in *Drosophila pseudoobscura* zu finden sind, wohingegen die Nukleotidsequenz in beiden Organismen weniger konserviert ist, spricht dafür, daß die kurzen ORFs translatiert werden. Auch das längere ORF weist Homologie zu *Drosophila pseudoobscura* auf. Über die Funktion von HDC13470 kann derzeit jedoch nur spekuliert werden.

Eine Möglichkeit ist, es sich bei den kurzen ORFs um sogenannte upstream ORFs (uORFs) handelt, die eine regulatorische Funktion auf das nachfolgende, längere ORF ausüben (Abb. 25). uORFs sind in der 5'-UTR Sequenz anderer Proteinen angesiedelt, und kommen häufig in mRNAs vor, die für regulatorische Proteine kodieren (Meijer und Thomas 2003). Oft übernehmen uORFs eine wichtige Funktion in der Kontrolle der Translationseffizienz, indem ihre Translation die eines in 3'-Richtung liegenden ORFs durch eine Ablenkung des Initiationskomplexes blockiert

(Geballe et al. 2000). Der Einfluß von uORFs auf die Initiation der Translation durch eine Modulation des Cap-abhängigen Scanningmechanismus ist bereits bekannt, und das Vorkommen von vorangehenden AUGs (uAUGs) und uORFs in etwa 50% aller humanen mRNAs hat das Interesse an diesem Regulationsmechanismus in letzter Zeit verstärkt (Kochetov et al. 1998; Davuluri et al. 2000; Pesole et al. 2000; Suzuki et al. 2000; Rogozin et al. 2001). Normalerweise wird die Translation des eigentlichen ORFs durch die Translation der uORFs gehemmt, indem die Anzahl der Ribosomen, die das eigentliche ATG erreichen, verringert wird (Gray 1998), (Morris und Geballe 2000).

Ein Beispiel für eine erweiterte Regulation der Translationsinitiation durch uORFs ist für die  $\beta$ -Sekretase (BACE1) beschrieben worden (Rogers et al. 2004). In den Gehirnen an Alzheimer erkrankter Patienten ist die Menge des synthetisierten  $\beta$ -Sekretase Proteins 2.7-mal höher als in Gehirnen gesunder Personen (Fukumoto et al. 2002; Holsinger et al. 2002; Yang et al. 2003), während die BACE1 mRNA in den betroffenen Patienten nicht höher ist als in den Kontrollen (Yasojima et al. 2001; Gatta et al. 2002; Preece und Cairns 2003). Die Überexpression der  $\beta$ -Sekretase muß demzufolge durch eine erhöhte Translationsrate zustande kommen. Rogers et al. diskutieren, daß die drei uORFs in der 5'-UTR Sequenz des BACE1 Proteins aus der Maus im gesunden Organismus zur Hemmung der Translationsinitiation führen, während es in kranken Tieren möglicherweise zu einer erhöhten Translationsinitiation kommt, indem die uORFs durch einen so genannten Shunting Mechanismus umgangen werden. Durch den Shunting Mechanismus umgehen Ribosomen eine stabile sekundäre RNA Struktur und vorangehende AUGs bevor sie die Translation initiieren (Hohn et al. 2001).

Der Unterschied zwischen den beschriebenen uORFs und den beobachteten potentiellen Peptiden von HDC13470 liegt darin, daß uORFs normalerweise nicht in ihrer Sequenz, sondern nur in ihrer Funktion konserviert sind. Auch wurden bisher keine uORFs beschrieben, die in ihrer Sequenz identisch sind. Die hohe Konservierung der kurzen ORFs spricht daher für eine weitere Funktion für die Peptide. Zunächst muß allerdings geklärt werden, ob die gefundenen ORFs tatsächlich translatiert werden. RT-PCRs auf isolierten Polysomen können Antworten auf diese Fragen geben, obwohl die zu erwartende Größe der kurzen ORFs ein Problem bei der PCR darstellt.



**Abb. 25 Mögliche regulatorische Funktion von HDC13470 über uORFs.** Die in der 5'-Region von HDC13470 vorkommenden kurzen ORFs sind in blau und die darin vorkommenden hochkonservierten Regionen in gelb dargestellt. Das potentiell über diese uORFs regulierte 3' liegende ORF (weißer Kasten) ist ebenfalls in der vorhergesagten Region von HDC13470 enthalten.

Genetische Untersuchungen zeigen, daß Mutationen im genomischen Lokus von HDC13470 letal sind. Ob diese Letalität durch das Fehlen der kurzen Peptide oder des längeren ORFs zustande kommt, muß in weiteren Analysen, zum Beispiel durch P-Element Exzisionen oder Punktmutationen in den einzelnen ORFs gezeigt werden. Die Mutanten sterben in frühen larvalen Stadien. Aufgrund des regelmäßigen Musters der Expression im zellulären Blastoderm kann eine Funktion in der Segmentierung des Embryos nicht ausgeschlossen werden. Mit Hilfe der Zähnenreihen in Kutikula Präparationen kann man die korrekte Segmentierung untersuchen, und so eventuell einen weiteren Einblick in die Funktionen von HDC13470 bekommen.

### **3.6 Zielgene von Polycomb**

Zielgene von Polycomb sind bislang in nur wenigen genomischen Regionen untersucht worden. Darunter befinden sich die Bithorax (BX-C) und Antennapedia (ANT-C) Komplexe, welche die Hoxgene beinhalten (Zink et al. 1991; Gindhart und Kaufman 1995; Orlando und Paro 1995; Ringrose et al. 2003); der *polyhomeotic* locus (Bloyer et al. 2003), der *engrailed* locus (Kassis 1994) und der *hedgehog* locus (Maurange und Paro 2002).

Eine große Zahl an Zielgenen ist jedoch bisher unbekannt, was die etwa 100 Bindungsstellen für PcG-Mitglieder auf polytären Speicheldrüsenchromosomen (Zink et al. 1991; Rastelli 1993) belegen. Microarrays bieten eine gute Möglichkeit, um genomweit nach Zielgenen für Polycomb zu suchen.

#### **3.6.1 Der Versuchsaufbau der Microarrayanalyse**

In dem hier durchgeführten Versuch wurde RNA einer molekular untersuchten *Pc*-Mutante gegen eine Kontroll-RNA hybridisiert. Nach vier Stunden übernehmen die PcG/trxG Proteine die Funktion, den Expressionszustand der Hoxgene zu übernehmen und stabil zu halten. In den Mutanten Embryonen kann das Pc Protein seine Funktion nicht ausüben. Dies sollte zur Expression der bis dahin reprimierten, unter der Kontrolle von PC liegenden Hoxgene führen. Daher wurden zunächst homozygot mutante Embryonen des Stadiums 10, eine Stunde nach der beschriebenen Regulationsübernahme durch die PcG/trxG Mitglieder, verwendet.

Der Zeitpunkt wurde so gewählt, daß die Funktion von Pc bereits von Bedeutung im Embryo ist, aber so früh, daß wenige sekundäre Zielgene betroffen sein sollten. des Weiteren sollte die Wahl des Zeitpunktes sicherstellen, daß durch die von der Mutation verursachte Letalität die Expression weiterer Gene, wie z.B. der Apoptosegene, nicht verändert werden. Ob der gewählte Zeitpunkt allerdings schon ausreicht, um eine Fehlregulation der meisten Zielgene zu beobachten, ist noch nicht erwiesen.

### 3.6.2 Probleme bei der Auswertung der Microarrays

Die Auswertung von Microarrays wird von verschiedenen Faktoren beeinflusst, die je nach gewählter Normalisierung und Analyse voneinander abweichen. Die meisten Analysen ermitteln regulierte Gene anhand des Verhältnisses von zu untersuchender Probe zur Kontrolle. Je mehr Arrays für ein Experiment hybridisiert werden, desto sicherer sind die Aussagen, die über die Regulation der einzelnen Gene gemacht werden. Kleine Staubkörner, Hintergrundsignale und ähnliches können dazu führen, daß Spots bei der Auswertung nicht berücksichtigt werden und damit aus der Berechnung der regulierten Gene herausfallen. Für den hier vorliegenden Versuch sind bisher nur sechs Datenpunkte pro Gen vorhanden, von denen jeweils zwei auf einem Array liegen. Obwohl die Gesamtqualität der Hybridisierungen für eine Auswertung ausreichend ist, fehlen doch noch weitere Datenpunkte für eine gesicherte Analyse der Pc-Zielgene. Auch fehlt im Experiment zum jetzigen Zeitpunkt eine biologische Wiederholung, d.h. eine Wiederholung des Experimentes mit einer unabhängig isolierten und amplifizierten RNA-Probe. Die generierte Liste der Zielgene ist daher nur als vorläufiges Ergebnis anzusehen, welches über weitere Analysen bestätigt werden muß.

### 3.6.3 Bekannte Zielgene von Polycomb

Einige der Gene, für die eine Regulation durch die PcG bekannt ist, wurden in der vorläufigen Auswertung gefunden. Die Funktion der Hoxgene wird ab Stadium 9 der Embryonalentwicklung durch die Mitglieder der PcG und trxG aufrecht gehalten. Die Hox-Gene *Abd-B* und *Ubx*, sowie einige weitere Gene, für die eine Regulation durch Pc nur aufgrund bioinformatischer Vorhersagen angenommen wird, wurden in der Liste der Zielgene gefunden. Andere hingegen, wie die anterior exprimierte Hoxgene *lab* und *Pb* konnten jedoch in der hier vorliegenden Analyse nicht gefunden werden. Ein Grund für diese Beobachtung ist, daß ganze Embryonen für das vorliegende Microarrayexperiment verwendet wurden. Die Expression einiger Hoxgene ist nur auf wenige Segmente beschränkt (siehe Abb. 2). Beispielsweise sind *lab*, *Pb* und *Dfd* jeweils nur in einem anterioren Segment exprimiert; diese Gene können aufgrund ihrer, in Bezug zum Gesamtembryo, geringen Expression in der Analyse der Kontrolle durch die Normalisierung verloren gegangen sein. Im Gege-

satz dazu sind die Gene *Abd-B* und *Antp* in mehreren Geweben exprimiert. Eine Hochregulation der Expression dieser Gene sollte auf einem Microarray detektierbar sein. Tatsächlich konnte *Abd-B* auf dem Microarray als reguliert detektiert werden, allerdings sind sie durch die stringenten Auswertungsbedingungen unter den gesetzten Grenzwert gefallen. *Antp* konnte in dieser Analyse nicht gefunden werden. Eine mögliche Ursache hierfür ist, daß *Antp* durch die Regulation der weiter posterior exprimierten Hox-Gene negativ reguliert wird, ein Phänomen, das in Drosophila als „posterior prevalence“ bezeichnet wird (Macias et al. 1990). Eventuell ist der gewählte Zeitpunkt für dieses Experiment zu spät, um die primäre Auswirkung des Funktionsverlustes von Pc für *Antp* zu beobachten.

Andere Gene, die unter der Kontrolle von Polycomb liegen, konnten auf dem Array ebenfalls nicht als reguliert nachgewiesen werden. *Engrailed* ist eines dieser Gene. Allerdings ist dabei zu beachten, daß die Funktion von Polycomb auf *engrailed* bisher nur in Flügelscheiben gezeigt werden konnte. Ob eine solche Funktion bereits in frühen embryonalen Stadien besteht, ist nicht bekannt (Moazed und O'Farrell 1992). Insgesamt wurden durch Antikörperfärbungen auf polytären Speicheldrüsenchromosomen mehr als 100 Bindungsstellen von Polycomb beschrieben (Zink et al. 1991; Rastelli et al. 1993); eine *in-silico* Vorhersage ergibt mindestens 167 PREs im Genom (Ringrose et al. 2003). 38% der Pc-Bindungsstellen und 14 der vorhergesagten PREs werden durch die Microarraydaten ebenfalls gefunden. Eine größere Korrelation kann aufgrund der unterschiedlichen Methoden nicht erwartet werden. Das liegt zum Einen daran, dass über Antikörperfärbungen an polytären Speicheldrüsenchromosomen nur ein differenziertes larvales Gewebe untersucht wird, und in einem anderen Gewebe, wie es für die Microarrayexperimente verwendet wurde, zum Teil andere Bindungsstellen erwartet werden.

Ebenfalls ist unbekannt, wie viele Bindungsstellen notwendig sind, um eine zytologische Bande histochemisch zu detektieren. So kann es sein, daß nicht alle Pc-Bindungsstellen auf den Chromosomen erkennbar sind. Daß dennoch die Korrelation der beiden Ansätze relativ hoch ist, kann ein Anhaltspunkt darauf sein, daß viele Gene während der gesamten Entwicklung und in verschiedenen Geweben durch Pc reguliert bleiben.

Der Vergleich mit vorhergesagten PREs ist dahingehend schwierig, da man nicht auf das regulierte Gen schaut, sondern auf die Region, über die ein Gen reguliert wird. Dabei ist nicht immer eindeutig, welches Gen dieser Regulation unterliegt. So konnte

beispielsweise für das Gen *Ubx* gezeigt werden, daß es der Regulation des bxd-PREs unterliegt (Simon et al., 1993). In der Region, in der das bxd-PRE lokalisiert ist, liegen jedoch zwei weitere Gene näher an dem PRE als *Ubx* selber. Dies zeigt, daß eine einfache Zuordnung von Genen zu vorhergesagten PREs nur bedingt möglich ist.

Ein weiterer Unterschied zwischen den Daten, der einen direkten Vergleich erschwert, ist, daß in dem bioinformatischen Ansatz alle PREs, die in bestimmten Geweben und in unterschiedlichen Zeitpunkten eine Rolle spielen, vorhergesagt werden. Wie viele dieser PREs allerdings während der embryonalen Entwicklung von Bedeutung sind und welche Gene unter der Kontrolle dieser PREs liegen, ist zur Zeit unklar und wird nur durch zukünftige Experimente geklärt werden können.

### **3.7 Ausblick**

Mit Hilfe des in dieser Arbeit entwickelten HeidelbergFlyArrays wird es möglich sein, weitere, bisher nicht identifizierte Zielgene der PcG/trxG Gene zu ermitteln. Die Bindungsstellen dieser Multiproteinkomplexe an die DNA können damit aber nicht genau analysiert werden.

Parallel zu dem in dieser Arbeit gewählten Ansatz wurden über einen Algorithmus 167 Pc-Bindungsstellen vorhergesagt (Ringrose et al. 2003). Die Gene, die über diese Bindung des Komplexes an die DNA reguliert werden, können jedoch aus diesen Daten nicht geschlussfolgert werden. Liegt die Bindung direkt an den Promotoren von Genen, so ist die Wahrscheinlichkeit, daß diese Gene auch reguliert werden, sehr hoch. Oftmals ist die Bindungsstelle jedoch weiter entfernt von dem nächsten Gen, so daß kein direkter Bezug hergestellt werden kann.

Genomische Microarrays sind von großer Bedeutung für die Analyse der PcG Bindungsstellen im Genom. Für diejenigen Gene, für die mit Hilfe der Microarrayanalysen eine Regulation durch Pc gezeigt werden konnte, kann durch genomische Arrays deren jeweilige Bindungsstelle auf der DNA nachgewiesen werden. Dabei bedient man sich der Verbindung aus Chromatin-Immunpräzipitationen und Microarrays (ChIP-chip) (Ng et al. 2002).

Für die genomischen Regionen des Antennapedia-Komplexes sowie des Bithorax-Komplexes sind genomische Microarrays bereits vorhanden. In vorangegangenen Experimenten wurden die PREs für die Gene des Bithorax-Komplexes kartiert

(Strutt et al. 1999), so dass diese bekannten Regionen auf dem Microarray als Kontrolle für die Qualität der Hybridisierungen dienen können. Erste Hybridisierungen mit Material aus Chromatin-IPs unter Verwendung des in dieser Arbeit hergestellten Pc-Antikörpers wurden bereits durchgeführt; die Auswertung dieser Arrays ist in Arbeit (C. Beisel, persönliche Mitteilung). Weitere Microarrays, die zusätzlich flankierende genomische Regionen potentieller Zielgene der PcG enthalten, sind ebenfalls in Vorbereitung (C. Popp und C. Beisel, persönliche Mitteilung).

Eine große Hilfe bei dieser Methode der Feinkartierung werden in Zukunft genomische Microarrays sein, die das gesamte Genom von *Drosophila* in 1 kb Fragmenten abdecken, und die voraussichtlich ab 2005 zur Verfügung stehen werden (White 2004).

Die flankierende genomische Region dieser Gene wird dabei in Fragmente von 1 kb Größe aufgeteilt und auf Microarrays gespottet.

Die Verbindung der auf Transkriptomarrays beruhenden Ergebnisse mit denen von genomischen Microarrays und *in-silico* Vorhersagen der Pc-Bindungsstellen, gefolgt von validierenden genetischen sowie molekularbiologischen Versuchen werden es in Zukunft erlauben, alle durch Pc regulierte Gene und deren jeweilige Bindungsstellen zu identifizieren.

## 4 Material

### 4.1 Laborausstattung

Agilent Bioanalyzer	Agilent, Böblingen
Analytische Waage	Sartorius, Göttingen
Autoradiographiekassetten Suprema	Dr. Goos, Bautzen
Binokular Stemi SV6	Zeiss, Wetzlar
Biogrid II Microarray Spotter	Geneworks, Oberhachingen
DNA-Gellaufkammern	ZMBH-Werkstatt
Filmentwickler	Amersham Pharmacia, München
Fliegeninkubator	RUMED, Laatzen
Flugkäfige	ZMBH, Werkstatt
Heizblock Thermomixer 5436	Eppendorf, Hamburg
Mikro-Pipetten P10, P20, P200, P1000	Gilson Pipetman, Den Haag
Mikroskope BX60	Olympus
DMXRA	Leica, Wetzlar
MS5	Leica, Wetzlar
MilliQ Wasser System	Millipore® Molsheim, France
Multipette plus	Eppendorf, Hamburg
Multikanalpipetten	Capp, Odense
pH-Meter MP 220	Mettler Toledo
Photometer Ultrospect 3000	Amersham Pharmacia
Pipettierhilfe, Pipetboy	Gilson
Rotoren SS34, GS3	GSA DuPont
Schüttelinkubator Innova 4230	New Brunswick Scientific
Thermocycler	Perkin Elmer
Über-Kopf Schüttler	NeoLab, Heidelberg
Ultrazentrifuge L8-70M	Beckmann, Krefeld
UV Stratalinker	Stratagene
Wasserbad SUB 14	Grant

## 4.2 Software, Hardware und Internetadressen

### 4.2.1 Software für Genvorhersagen

BDGP Genvorhersage

Version 2-4 [www.bdgp.org](http://www.bdgp.org) (Celniker et al. 2002; Stapleton et al. 2002)

Fgenesh software [www.softberry.com](http://www.softberry.com) (Salamov und Solovyev 2000)

GenomePride <http://pride.molgen.mpg.de/genomepride.html> (Haas et al. 2003)

perl [www.perl.org](http://www.perl.org)

### 4.2.2 Software für Microarrayauswertung

Agilent Lab on a Chip [www.chem.agilent.com](http://www.chem.agilent.com)

express yourself <http://array.mbb.yale.edu/analysis/>

GenePix Pro™ 4.0 [www.axon.com](http://www.axon.com)

MCHiPS [www.mchips.org](http://www.mchips.org)

R statistical language [www.r-project.org](http://www.r-project.org)

S-Plus Software [www.insightful.com](http://www.insightful.com)

### 4.2.3 Verwendete Datenbanken

Blast [www.ncbi.nih.gov](http://www.ncbi.nih.gov)

Ensembl <http://www.ensembl.org/Multi/martview>

Fatigo <http://fatigo.bioinfo.cnio.es/>

Flybase [www.flybase.org](http://www.flybase.org)

Flymove [www.flymove.uni-muenster.de](http://www.flymove.uni-muenster.de)

Gene Expression patterns of *Drosophila melanogaster* [www.fruitfly.org/cgi-bin/ex/insitu.pl](http://www.fruitfly.org/cgi-bin/ex/insitu.pl) (Tomancak et al. 2002)

Gene Expression Omnibus [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)

Gene Ontology [www.geneontology.org](http://www.geneontology.org)

Genbank [www.ncbi.nih.gov](http://www.ncbi.nih.gov)

Interpro <http://www.ebi.ac.uk/interpro/>

### 4.3 Verbrauchsmaterial

Edelstahlprüfsiebe (1 mm, 25 µm)	Neolab, Heidelberg
Einmalhandschuhe Powder-Free Pehasoft	Neolab, Heidelberg
Filter 0,2 und 0,45 mm, Cellulose Ester	Schleicher & Schuell
Filterpapier (3 MM)	Whatman, Cambridge
Küvetten 10 mm x 4 mm x 45 mm	Sarstedt, Nümbrecht
Pasteur Pipetten	WU Mainz
PCR-Reaktionsgefäße	Greiner, Frickenhausen
PCR 96-well Platten	Applied Biosystems, Halbrand
396-well Platten, Polysterol, V-bottom	Genomix
Pipettiergefäße	Brand
Pipettenspitzen	Molecular Bioproducts
Abdeckfolie für 96-well Platten	Peqlab, Erlangen
Polypropylenröhrchen (15/50 ml)	Greiner, Frickenhausen
Reaktionsgefäße 0,5, 1,5 u. 2 ml, Sarstedt	Eppendorf, Hamburg

#### 4.3.1 Chemikalien, Kits, Enzyme, Proteine, Molekulargewichtsmarker

Nicht aufgeführte Chemikalien und wurden in der Qualität p.A. von den Firmen Sigma (München), Merck AG (Darmstadt), Roth (Karlsruhe) und Fluka AG (Buchs, CH) bezogen.

Acrylamid, linear	Ambion
Agarose ultra pure	Gibco BRL
<i>Anti</i> -DIG AP-Konjugat	Roche
Bicin	Sigma, Deisenhofen
Bis-Tris	Sigma, Deisenhofen
Bromphenolblau	Serva, Heidelberg
BSA	Boehringer, Mannheim
DNA-Standards 1 kb DNA Ladder	MBI Fermentas
DNA-Polymerase-I	NEB, Frankfurt
dNTPs	Peqlab, Krefeld
EDTA	Merck, Darmstadt

---

<i>E. coli</i> DNA Ligase	NEB, Frankfurt
Ethanol, 99%	Zentrallager INF
Glykogen	Roche, Mannheim
Glutaraldehyd	Merck, Darmstadt
Hefeextrakt	Gibco BRL
Heptan	Applichem, Darmstadt
Indirect Labelling Kit	Amersham Bioscience
Lineares Acrylamid	Ambion, Huntington
Megaskript T7	Ambion, Huntington
Methanol	Merck, Darmstadt
Natriumhydroxid	J.T. Baker
Natriumhypochlorit, 12% Chlor	Zentrallager INF
Nipagin	Merck, Darmstadt
Paraformaldehyd	Sigma, Deisenhofen
Phenol	Applichem, Darmstadt
Plasmid Midi/Maxi Kit	Qiagen, Hilden
Proteinase K	Roche, Mannheim
Reverse Transkriptase	Promega
RNase A	Roche, Mannheim
RNase H	Gibco
RNasin, Ribonuklease-Inhibitor	Promega
Saccharose	J.T. Baker
Superscript II	Gibco
Taq-Polymerase	Qiagen, Hilden
T7-Polymerase	Roche, Mannheim
T3-Polymerase	Roche, Mannheim
Trehalose	Sigma, Deisenhofen
Tris	Applichem, Darmstadt
Triton X-100	Merck, Darmstad
Trizol	Invitrogen

Restriktionsenzyme wurden von New England Biolabs oder Roche bezogen.

## 4.4 DNA

### 4.4.1 Oligonukleotide

Alle Oligonukleotide für die Herstellung der Fragmente für den Microarray sind der Webseite <http://hdflyarray.zmbh.uni-heidelberg.de> oder der beiliegenden CD zu entnehmen.

TAG-1	5'CGACGCCCCGCTGATA
TAG-2	5'TAGGTCTAGCCCCGC
TAG-3	5'GGCGTCTGACGCTAG
TAG-4	5'CGCATGTAGCCTGCC
TAG-5	5'TAGCCTCCCTAGCGC
TAG-6	5'AGCCCCTAACGTGCG
TAG-7	5'AGCTAGCCCCGGGTGA
TAG-8	5'TGACCCGGGTAGCCT
TAG-9	5'TGAGGTACGCGTGGG
UNI-TAG	5'TGGCGCCCCTAGATG
T7-Tag	5'TAATACGACTCACTATAGG
Oligo-dT/T	5'AAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGCGCT15

### 4.4.2 Plasmide

pBluescript II KS	Stratagene
pCRII-Topo	Invitrogen
pSP64	Promega

## 4.5 Bakterienstämme (*Escherichia coli*)

XL1-blue (Bullock und Wright 1987)

*recA* - (*recA1 lac - endA1 gyrA96 thi hsdR17 supE44 relA1 {F' proAB lacIq lacZDM15Tn10}*)

## 4.6 Fliegenstämme

### 4.6.1 Balancerchromosomen

Balancer	Chromosom	Referenz
$y^1 w^*$ ; $D^* gl^3/TM3$ , $P\{w^{+mC}=GAL4Kr.C\}DC2$ , $P\{w^{+mC}=UAS-GFP.S65T\}DC10$ , $Sb^1$	3	(Casso et al. 2000)
$w^{1118}$ ; $In(2LR)Gla$ , $wg^{Gla-1} Bc^1/CyO$	2	Cook, 1990
$Sp/CyO^{bb}$ ; $TM2/TM6^{bb}$	2, 3	Klämbt, unveröffentlicht
$w^{1118}$ ; $Dr^{Mio}/TM3$ , $P\{w^{+mC}=GAL4-twi.G\}2.3$ , $P\{AUS-2xEGFP\}AH2.3$ , $Sb^1 Ser^1$	3	(Halfon et al. 2002)
$w^{1118}/Dp(1;Y)y^+$ ; $TM2/TM6C$ , $Sb^1$	3	

### 4.6.2 Defizienzen

Defizienz	Bruchpunkte/ Insertion	Donor
$w^{1118}$ ; $Df(3R)Exel6288$ , $P\{w^{+mC}=XPU\}Exel6288/TM6B$ , $Tb^1$	87F14;88A4	Exelixis, Inc.
$w^{1118}$ ; $Df(3R)Exel6170$ , $P\{w^{+mC}=XPU\}Exel6170/TM6B$ , $Tb^1$	87F10;87F14	Exelixis, Inc.
$Df(3L)XG10/TM3$ , $P\{w^{+mW.hs}=Thb8lacZ\}WD1$ , $Sb^1 Ser^1$	71C3-D1;71F2-5	J. Grosshans
$w^{1118}$ ; $Df(3L)Exel6126$ , $P\{w^{+mC}=XPU\}Exel6126/TM6B$ , $Tb^1$ $w^{1118}$ ; $Df(3R)Exel6171$ , $P\{w^{+mC}=XPU\}Exel6171/TM6B$ , $Tb^1$	71A3;71B3	Exelixis, Inc.
$Df(3L)XG1/TM3$ , $P\{w^{+mW.hs}=Thb8lacZ\}WD1$ , $Sb^1 Ser^1$	71C3-D1;71F2-5	J.Grosshans
$w^{1118}$ ; $Df(3L)ED218$ , $P\{w^{+mW.ScerFRT.hs3=3'.RS5+3.3'}ED218/$ $TM6C$ , $Sb^1$	71B1;71E1	DrosDel proj.
$Df(3L)Brd15$ , $p^D/TM6B$ , $Tb^1$	71A1-2;71C1-2	Jim Posakony
$Df(3L)Ly$ , $Df(3L)BK10$ , $ru^1 sens^{Ly-1} red^1 cv-c^1 Sb^{sbd-1} sr$ , $e^1/$ $TM3$ , $Sb^1$	70A2-3;70A5-6, 71C;71F	T.Kaufmann
$w^{1118}$ ; $Df(3L)ED219$ , $P\{w^{+mW.ScerFRT.hs3=3'.RS5+3.3'}ED219/$ $TM6C$ , $Sb^1$	71B1;71E1	DrosDel proj.
$w^{1118}$ ; $Df(3R)Exel6171$ , $P\{w^{+mC}=XPU\}Exel6171/TM6B$ , $Tb^1$	87F14;88A4	Exelixis, Inc.

### 4.6.3 Einzelmutationen

Name	Allele	Referenz
<i>ventral nervous system defective vnd</i>	vnd <sup>6</sup>	(White et al. 1983)
<i>Polycomb</i>	Pc <sup>XL5</sup> /Tm3, Ser Pc <sup>3</sup> /Tm3, Ser	
I(3)S011041 <sup>S011041</sup>		(Deak et al. 1997)

## 4.7 Medien, Puffer und Stammlösungen

### 4.7.1 Antibiotika

Ampicillin (Na-Salz) 100 mg/ml

Die Stammlösung wurde in der Verdünnung 1:1000 eingesetzt

### 4.7.2 Medien und Lösungen zur Herstellung kompetenter *E. coli*

#### LB-Medium

1% Pepton aus Casein

0,5% Hefeextrakt

0,5% NaCl

#### SOB-Medium

0,5% Hefeextrakt

10 mM NaCl

2,5 mM KCl

10 mM MgCl<sub>2</sub>

10 mM MgSO<sub>4</sub>

pH 7, steril filtrieren

### **4.7.3 Lösungen für Plasmid Rescue**

#### **Puffer A**

100 mM Tris-HCl, pH 7,5

100 mM EDTA

100 mM NaCl

0,5% SDS

### **4.7.4 Lösungen für Northern Blot**

#### **Transferpuffer**

2 x SSC

#### **Prähybridisierungspuffer**

5 x Denhardt's

5 x SSC

0,5% SDS

50% Formamid

#### **Hybridisierungspuffer**

5 x Denhardt's

5 x SSC

0,5% SDS

50% Formamid

0,1 µg/ml Heringssperma DNA (10mg/ml, Roche)

#### **Waschpuffer 1**

2 x SSC

0,1% SDS

#### **Waschpuffer 2**

0,1 x SSX

0,1% SDS

### **Strippingpuffer**

0,1% SSC, kochend

## **4.7.5 Lösungen für Microarray Spotting**

### **Spottinglösung**

3 x SSC

1,5 M Betain (AG Paro)

## **4.7.6 Lösungen für Microarray Hybridisierung**

### **Prähybridisierungslösung**

6 x SSC

0,2% SDS

2% BSA

### **Hybridisierungslösung**

50% Formamid

25% H<sub>2</sub>O

25% 4 x Hybridisierungspuffer (Amersham Bioscience)

### **Wash A**

2 x SSC

2% SDS

### **Wash B**

2 x SSC

0,2% SDS

### **Wash C**

2 x SSC

### **Prewash A**

0,2% SDS

### **4.7.7 Lösungen für In-situ Hybridisierung**

#### **Prähybridisierung**

50% Formamid

5 x SSC

0,1% Tween20

#### **Hybridisierung**

50% Formamid

5 x SSC

50 µg /ml Heparin (50mg/ml)

100 µg/ml Heringssperma DNA (10 µg/µl)

0,1% Tween20

#### **AP-Puffer**

20 mM Tris-HCl pH 9,5

100 mM NaCl

50 mM MgCl<sub>2</sub>

### **4.7.8 Lösungen für die lineare Amplifikation von RNA nach Eberwine**

1,7 M Trehalose

#### **10x Zweitstrangpuffer**

200 mM Tris pH6,9

900 mM KCl

46 mM MgCl<sub>2</sub>

1,5 mM Nicotin Adenin Dinukleotid (Calbiochem)

100 mM (NH<sub>4</sub>)<sub>3</sub>SO<sub>4</sub>

#### **4.7.9 Häufig verwendete Stammlösungen/Puffer**

##### **50 x TAE-Puffer**

242 g Tris base

57,1 ml Eisessig

37,2 g Na<sub>2</sub>EDTA·2H<sub>2</sub>O

H<sub>2</sub>O auf 1l

##### **10 x PBS**

80 g NaCl

2 g KCl

11,5 g Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O

2 g KH<sub>2</sub>PO<sub>4</sub>

## 5 Methoden

### 5.1 Molekularbiologische Methoden

#### 5.1.1 Isolierung genomischer DNA

Ca. 200-500 Fliegen wurden in einem 15 ml Plastikgefäß gesammelt, in flüssigem Stickstoff eingefroren und anschließend in 2 ml 1 x Homogenisierungspuffer aufgenommen. In einem gekühlten Glashomogenisator wurden die Fliegen homogenisiert. Nach einer Zentrifugation bei 1,000 Upm für 1 min wurde der Überstand abgenommen und 5 weitere min bei 8,000 Upm zentrifugiert. Das Pellet wurde in 500  $\mu$ l Homogenisierungspuffer aufgenommen und 50  $\mu$ l 10% SDS und Proteinase K in einer Konzentration von 100  $\mu$ g/ml hinzugefügt. Der Ansatz wurde bei 37°C für eine Stunde inkubiert; nach 30 min wurde 1  $\mu$ l RNase dazugegeben. Durch anschließende Phenol-Chloroform Extraktion wurde die DNA gereinigt, durch Zugabe von 1 Volumen Isopropanol präzipitiert, mit 80% Ethanol gewaschen und nach kurzem Trocknen in ddH<sub>2</sub>O aufgenommen.

#### 5.1.2 Isolierung von Plasmid-DNA aus *E. coli*

Miniprep Plasmidpräparation: 1,5 ml einer ÜNK wurden in einem Eppendorfgefäß zentrifugiert (Tischzentrifuge, 13,000 Upm, 45 sek). Das Bakterienpellet wurde in 200  $\mu$ l eiskaltem Puffer A (Qiagen) durch vortexen resuspendiert und für 5 min bei RT inkubiert. 200  $\mu$ l Puffer B (Qiagen) wurden hinzugegeben, und nach 5 min Inkubation wurden 200  $\mu$ l eiskaltem Puffers C (Qiagen) hinzugefügt, die Suspension auf Eis gekühlt und anschließend zentrifugiert (Tischzentrifuge, 13,000 Upm, 8 min, 4°C). Der Überstand wurde in ein neues Eppendorfgefäß überführt, und die DNA durch Zugabe von 0,7 Volumen 2-Propanol für 5 -10 min bei 4°C gefällt. Nach Zentrifugation (Tischzentrifuge, 13,000 Upm, 5 min) wurde die DNA durch Zugabe von 1 ml 70% Ethanol gewaschen, zentrifugiert und getrocknet. Anschließend wurde das Pellet in 50  $\mu$ l H<sub>2</sub>O aufgenommen und bei -20°C aufbewahrt. Midiprep/Maxiprep Plasmidpräparationen wurden mit den entsprechenden Kits von Qiagen nach den Anweisungen des Herstellers durchgeführt.

### 5.1.3 Phenol/Chloroform-Extraktion

Um aus einem Protein/Nukleinsäuregemisch die DNA zu reinigen, wurde die Phenol/Chloroform-Extraktion angewendet. Die Lösung wurde mit dem gleichen Volumen Phenol (Appllichem) versetzt und 2 min gevortext. Anschließend wurde ein Volumen Chloroform zugegeben, 2 min gevortext und die Phasen durch Zentrifugation (Tischzentrifuge, 13,000 Upm, 3 min) voneinander getrennt. Die wässrige Phase, welche die gelöste DNA enthält, wurde abgenommen, ohne die weiße proteinhaltige Interphase aufzuwirbeln. Sofern in der Interphase noch viel Protein vorhanden war, wurde die DNA-Lösung durch die Zugabe von 1 Volumen Chloroform, vortexen, Zentrifugation und Abnahme der oberen Phase weiter gereinigt.

### 5.1.4 Isopropanol-Präzipitation

Um die DNA nach der Phenol/Chloroform-Extraktion zu konzentrieren wurde die Isopropanol-Präzipitation angewendet. Zu der Nukleinsäurelösung wurde 1/10 Volumen 3 M NaOAc, pH 5,5 und 1 Volumen Isopropanol zugegeben. Die Nukleinsäure wurde bei  $-20\text{ }^{\circ}\text{C}$   $\infty$ N oder bei  $-80\text{ }^{\circ}\text{C}$  für 30 min gefällt und zentrifugiert (Tischzentrifuge, 13,000 Upm, 20 min). Das Präzipitat wurde mit 70% Ethanol gewaschen, erneut zentrifugiert (13,000 Upm, 5 min) und getrocknet. Anschließend wurde das Pellet in 50 ml  $\text{H}_2\text{O}$  aufgenommen und bei  $-20\text{ }^{\circ}\text{C}$  aufbewahrt.

### 5.1.5 Konzentrationsbestimmung von DNA-Lösungen

Die DNA-Konzentration wurde spektrophotometrisch bestimmt. Hierzu wurden zu 80  $\mu\text{l}$   $\text{H}_2\text{O}$  1-5  $\mu\text{l}$  DNA gegeben und die Lösung bei einer Wellenlänge von 260 nm gemessen. Nach dem Lambert-Beerschen Gesetz ist dieser Wert proportional der DNA-Konzentration, wobei eine  $\text{OD}_{260}$  von 1 einer Konzentration von 50  $\mu\text{g}/\mu\text{l}$  entspricht. Die Reinheit der DNA kann über das Verhältnis  $\text{OD}_{260}/\text{OD}_{280}$  bestimmt werden, welches über einem Wert von 1,8 liegen sollte.

### **5.1.6 Restriktionsverdau von DNA**

Der Restriktionsverdau von DNA wurde in Volumina von 20  $\mu$ l (analytisch)-100  $\mu$ l (präparativ) durchgeführt. Es wurden die vom Hersteller (New England Biolabs, Roche) empfohlenen Reaktionspuffer verwendet. Je nach Ansatz wurden 1-10  $\mu$ g DNA für eine Stunde bis  $\infty$ N bei der für die verwendeten Enzyme empfohlenen Temperatur verdaut.

### **5.1.7 Dephosphorylierung linearisierter Plasmide**

Zur Vermeidung intramolekularer Ligation wurden die 5'-Enden der linearisierten Plasmide durch Behandlung mit alkalischer Phosphatase dephosphoryliert. Nach dem Restriktionsverdau wurde zu dem Ansatz 1/10 Volumen 10 x Dephosphorylierungspuffer (Roche) und 1 Unit Shrimp alkalischer Phosphatase (Roche) zugegeben. Die Reaktion wurde für 1 h bei 37°C inkubiert und anschließend auf einem Agarosegel aufgetrennt.

### **5.1.8 Agarosegelelektrophorese**

Je nach Größe der zu bearbeitenden DNA-Fragmente wurden 0,8-2% Agarosegele verwendet. Die entsprechende Menge Agarose wurde eingewogen und in 1 x TAE bzw. 1/2 x TAE bei Verwendung des Mupidgelsystems durch Erwärmen gelöst. Danach wurde Ethidiumbromid (10  $\mu$ g/ml Endkonzentration) zugegeben und die Lösung in einen Gelschlitten gegossen. Die DNA-Proben wurden mit 1/10 Volumen 10 x Probenpuffer (1 x TBE, 50% Glycerin, 0,25% Orange G) versetzt, auf das Gel aufgetragen und elektrophoretisch aufgetrennt. Die DNA wurde mit Hilfe eines UV-Illuminators sichtbar gemacht und zur Dokumentation fotografiert.

### **5.1.9 Isolierung von DNA-Fragmenten aus dem Agarosegel**

DNA-Fragmente wurden mit Hilfe des QIA Gel Extraction-Kits (QIAGEN) nach Angaben des Herstellers aus einem Agarosegel isoliert.

### 5.1.10 Ligation von DNA-Fragmenten

Die Ligation von DNA-Fragmenten mit linearisierten dephosphorylierten Plasmiden erfolgte in einem Volumen von 20 µl für 4 h bei RT oder bei 16°C üN. Es wurden ungefähr 100 ng Plasmid und die 3fache molare Menge des zu inserierenden DNA-Fragmentes verwendet. Der Ansatz wurde durch Zugabe von 2 µl 10 x Ligasepuffer (Roche) auf die optimalen Bedingungen für die T4 DNA Ligase (1 U, Roche) eingestellt. Die Ligationsansätze wurden nach abgelaufener Inkubation entweder bei –20°C eingefroren oder direkt in kompetente *E. coli*- Zellen transformiert.

### 5.1.11 Herstellung kompetenter *E. coli*-Zellen nach Hanahan

Eine ÜNK wurde 1:100 in 25 ml SOB-Medium verdünnt und bei guter Durchlüftung aufgezogen. Nach Erreichen einer OD<sub>600</sub> von 0,3 wurden 2 x 10 ml der Kultur in zwei eisgekühlte Corex-Röhrchen überführt und zentrifugiert (HB-4, 2,500 Upm, 10 min, 4 °C). Die Zellpellets wurden in je 3,5 ml eiskaltem FSB-Puffer resuspendiert, für 10 min auf Eis inkubiert und ein weiteres Mal zentrifugiert (HB-4, 2,500 Upm, 10 min, 4 °C). Die Pellets wurden in insgesamt 2 ml eiskaltem FSB-Puffer resuspendiert. Im Abstand von 5 min wurden zweimal 70 µl DMSO zugegeben. Anschließend wurden die Zellen in Eppendorfgefäße aliquotiert und bei –80°C aufbewahrt (Hanahan 1983).

### 5.1.12 Transformation kompetenter *E. coli*-Zellen

10-20 µl eines Ligationsansatzes oder eines Plasmids wurden vorsichtig mit 100 µl einer eiskalten Suspension kompetenter *E. coli*-Zellen gemischt. Der Ansatz wurde 30 min auf Eis inkubiert, anschließend für 90 sek auf 42°C erwärmt und sofort wieder auf Eis gekühlt. Danach wurden 800 µl LB-Medium zugegeben und für weitere 45 min bei 37°C inkubiert. Die Zellen wurden zentrifugiert (Tischzentrifuge, 4,000 Upm, 5 min) und der Überstand abgeschüttet. Anschließend wurden die Zellen in dem im Eppendorfgefäß verbliebenen Rest (50-100 µl) resuspendiert, auf LB-Agarplatten mit entsprechenden Antibiotika ausplattiert und bei 37 °C inkubiert. Als Kontrollansätze dienten kompetente Zellen ohne DNA (Negativkontrolle).

### 5.1.13 Isolierung von total-RNA aus *Drosophila melanogaster*

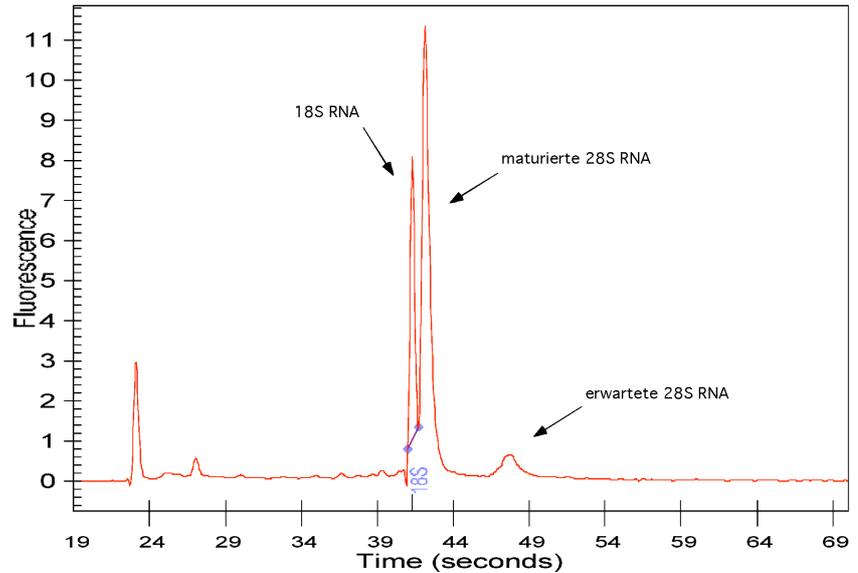
0,4 g des entsprechenden Entwicklungsstadiums wurden in einem 2 ml Eppendorfgefäß in 1 ml Trizol aufgenommen und in einem Ultraturrax homogenisiert. Das Homogenisat wurde für 5 min bei Raumtemperatur inkubiert, bevor 0,2 ml Chloroform dazugegeben wurden, und für 15 sek kräftig geschüttelt wurde. Nach weiterer zweiminütiger Inkubation bei RT wurde das Gemisch bei 4°C für 15 min bei 12,000 g zentrifugiert. Die obere, wässrige Phase wurde vorsichtig abgenommen und in ein neues Eppendorfgefäß überführt, 0,5 ml Isopropanol wurden zugesetzt und das Gemisch für 15 min bei RT inkubiert, bevor es bei 4°C bei 12,000 Upm für 10 min zentrifugiert wurde. Die Flüssigkeit wurde vorsichtig entnommen und das Präzipitat wurde mit 75% Ethanol gewaschen, anschließend leicht getrocknet und mit ddH<sub>2</sub>O aufgenommen.

Für Larval- und Pupalstadien wurde ein Reinigungsschritt durchgeführt, um überschüssige Polisaccharide und Fette zu entfernen: anstelle von 0,5 ml wurden hier nur 0,25 ml Isopropanol hinzugegeben, zusätzlich wurden 0,25 ml einer Hochsalzpräzipitationslösung (0,8 M Natriumzitat und 1,2 M NaCl) zugesetzt, um Fette und Polisaccharide in der Lösung zu halten, während die RNA präzipitiert wurde.

### 5.1.14 Qualitätskontrolle von total RNA

Die Qualität der RNA wurde über das Agilent-System ermittelt. 50-150 ng total RNA wurden entsprechend den Herstelleranweisungen mit Marker und Puffer versehen und auf dem Agilent RNAChip analysiert. Anders als bei totaler RNA von höheren Wirbeltieren kann man bei *Drosophila* nicht von dem Verhältnis von 28S zu 18S RNA auf die Qualität schliessen. Die 28S RNA scheint während der Entwicklung zu maturieren (Jordan et al. 1976), und ein weiterer Peak bildet sich zwischen den 18S und 28S RNA Peaks. Das Verhältnis zwischen dem Peak der prozessierten 28S RNA und dem 18S Peak gibt Aufschluß über die Qualität der RNA und sollte in einem Verhältnis von 2:1 vorliegen. Weitere Peaks können Aufschluß über DNA Kontamination (Fragmente, die größer als die 28S RNA sind) oder Degradation (kleinere Peaks unterhalb der 18S RNA) geben.

**Abb. 26 Analyse von *Drosophila* totaler RNA auf dem Agilent-Chip.** Während die 18S RNA gut zu erkennen ist, ist die erwartete 28S RNA-Bande klein. Eine prozessierte Form dieser 28S RNA bildet sich zwischen der 18S- und der 28S RNA. Das Verhältnis der maturierten 28S RNA und der 18S RNA sollte in einem Verhältnis von 2:1 vorliegen. Das Fehlen weiterer Banden, die DNA Kontamination und Degradation kennzeichnen, zeigt die hohe Qualität der RNA.



### 5.1.15 RT-PCR

Für die RT-PCR wurden zunächst 20  $\mu$ g RNA mit 10 Units DNase für 30 min bei 37°C inkubiert. Anschließend wurde die RNA über Phenol-Chloroform Extraktion aufgereinigt, mit Isopropanol präzipitiert und in H<sub>2</sub>O aufgenommen. Für ein Experiment wurden 1  $\mu$ g RNA mit Random Nonamers in einem Volumen von 9  $\mu$ l bei 72°C für 5 min inkubiert, und anschließend für 5 min bei Raumtemperatur stehen gelassen, damit die Primer gut an die RNA binden konnten. Anschließend wurden 4  $\mu$ l dNTPs (je 10 mM), 4  $\mu$ l 5x Reverse Transkriptionspuffer, 1  $\mu$ l RNAsin (40 U/ $\mu$ l, Promega) und 1  $\mu$ l Reverse Transkriptase (200U/ $\mu$ l, Promega) zugegeben und die Reaktion in der PCR Maschine für 10 min bei 30°C, gefolgt von 60 min bei 42°C und 15 min bei 55°C inkubiert. Ein zweiter Ansatz ohne Reverse Transkriptase diente jeweils als Negativkontrolle.

Von der synthetisierten cDNA wurde jeweils 1  $\mu$ l als Template für die PCR eingesetzt. 20  $\mu$ mol Primer, 0,2 mM dNTPs (10 mM), 10x PCR Puffer, 1  $\mu$ l Taq-Polymerase (5 Units/ $\mu$ l) und H<sub>2</sub>O wurden in einem Gesamtvolumen von 25  $\mu$ l gemischt und auf das Template gegeben. Je nach Anwendung wurden zwischen 25 und 40 Zyklen gefahren.

## 5.1.16 Plasmid Rescue

### 5.1.16.1 „Quick Fly“ genomische DNA Präparation

30 Fliegen wurden gesammelt und in flüssigem Stickstoff eingefroren. Mit einem Plastikmörser (Kontes) wurden die Fliegen in einem 1,5 ml Eppendorfgefäß in 100  $\mu$ l Puffer A zerkleinert. 300  $\mu$ l Puffer A wurden in 100  $\mu$ l Schritten hinzugefügt, und die Fliegen weiter zerkleinert. Die homogenisierten Fliegen wurden für 30 min bei 65°C inkubiert. 800  $\mu$ l LiCl/KAc Lösung wurde dazugegeben und das Ganze für 10 min auf Eis inkubiert. Danach wurde das Homogenisat für 25 min bei 12,000 Upm abzentrifugiert. Der Überstand wurde in ein neues Eppendorfgefäß übertragen, 600  $\mu$ l Isopropanol wurden hinzugefügt, und nochmals für 15 min zentrifugiert. Der Überstand wurde abgenommen, das Pellet mit 70% Ethanol gewaschen und in 150  $\mu$ l H<sub>2</sub>O aufgenommen, bei 65°C für 15 min inkubiert, um eine komplette Lösung des Pellets zu erreichen und die DNA bei –20°C gelagert.

### 5.1.16.2 Verdau

Für den Verdau wurden 25  $\mu$ l genomischer DNA mit 4  $\mu$ l 10 x Puffer, 100  $\mu$ g/ml RNase, 2  $\mu$ l H<sub>2</sub>O und 40 Units EcoRI gemischt, und der Verdau für 3 h bei 37°C inkubiert. Für jede Fliegenlinie wurde zusätzlich ein Kontrollansatz ohne Enzym angesetzt. Anschließend wurde das Enzym bei 70°C inaktiviert. 10  $\mu$ l des Verdau wurden für ein Testgel aufgespart, und der Rest über Phenol/Chloroform Extraktion und anschließender Ethanol Präzipitation aufgereinigt und in 40  $\mu$ l H<sub>2</sub>O aufgenommen.

### 5.1.16.3 Ligation, Transformation und Analyse

Die verdaute genomische DNA wurde mit T4-Ligase für 20 h bei RT ligiert und anschließend über erneute Phenol/Chloroform Extraktion und Ethanol Präzipitation aufgereinigt und in 10  $\mu$ l H<sub>2</sub>O aufgenommen. Die gereinigte DNA wurde anschließend in elektrokompetente XL1blue Zellen transformiert. Die Zellen wurden 1 h bei 37°C inkubiert und dann auf einer LB-Platte mit 25  $\mu$ g Ampicillin ausplattiert. Am nächsten Morgen wurden die Kolonien gepickt, die DNA über

eine Minipräparation gereinigt und in 30  $\mu$ l H<sub>2</sub>O aufgenommen. Die P-Insertionsstelle wurde durch Sequenzierung (Seqlab) ermittelt. Als Primer diente der P-Element spezifische Primer 5'-CAAGCATACGTTAAGTGGATG-3', dessen 3'-Nukleotid an der Position +10.650 des 10.691-bp PlacW Konstrukts liegt.

## 5.1.17 Northern Blot

### 5.1.17.1 Gelelektrophorese

Für einen Northern Blot wurde für alle Entwicklungsstadien aus 10  $\mu$ g total-RNA mit Hilfe des mRNAeasy Kits von Qiagen entsprechend den Anweisungen des Herstellers mRNA isoliert. Die mRNA wurde mit Ethanol in Anwesenheit von 1  $\mu$ l Glycogen präzipitiert, mit 70% Ethanol gewaschen, leicht angetrocknet und mit 5,4  $\mu$ l H<sub>2</sub>O aufgenommen. Anschließend wurden 5,6  $\mu$ l deionisiertes Glyoxal, 16  $\mu$ l DMSO und 3  $\mu$ l 100 mM Phosphatpuffer pH 7.0 hinzugegeben, 1 h bei 50°C inkubiert und dann auf Eis gegeben. Nach Zugabe von 4  $\mu$ l Ladepuffer wurden die Proben auf ein 1% Agarosegel / 10 mM Phosphatpuffer aufgetragen. Die Elektrophorese erfolgte bei 100 V für 3 Stunden, wobei der Puffer mit einer Pumpe vom – Pol zum + Pol zirkuliert wurde. Die erste Spur, die den RNA Marker enthielt, wurde vom restlichen Gel abgetrennt, mit EtBr (0.5  $\mu$ g/ml in 0.5 M Ammoniumazetat) für 40 min gefärbt und fotografiert. Das restliche Gel wurde für 30 min in 20 x SSC inkubiert.

### 5.1.17.2 Blotten

Eine Transferapparatur wurde nach den Standardprotokollen aufgebaut. Das Gel wurde umgekehrt auf Whatmanpapiere gelegt, deren Enden in 2 x SSC eintauchten. Auf die glatte Seite des Gels wurde die mit 2 x SSC angefeuchtete Membran (Hybond NX) gelegt; die Seiten der Apparatur wurden mit Parafilm abgedichtet. Auf die Membran wurden angefeuchtete Whatmanpapiere gelegt und darüber 5-10 cm hoch Apurablätter, deren Größe genau der Membran angepasst war. Zum Schluß wurden die Apurablätter mit einer Glasplatte fixiert und auf die Glasplatte ein Behälter zum Beschweren gelegt. Der Transfer erfolgte über Nacht. Am nächsten Morgen wurde der Blot in 20 x SSC gewaschen, um Reste der Agarose zu entfernen.

Danach wurde die Membran auf ein Whatmanpapier gelegt und im Stratalinker (Autocross) fixiert. Anschließend wurde die Membran zusätzlich für 30 min bei 80°C gebacken. Um das Glyoxal zu entfernen, wurde die Membran 10 min bei 65°C in 20 mM Tris pH 8.0 inkubiert.

### **5.1.17.3 Hybridisierung**

Die Prähybridisierung erfolgte bei 42°C für 2 Stunden. Die Proben wurden mit Hilfe des RediPrime (Amersham) nach Angaben des Herstellers radioaktiv mit <sup>32</sup>P markiert. Die Probe wurde für 5 min bei 95°C denaturiert, zu dem Hybridisierungspuffer dazugegeben und über Nacht bei 42°C inkubiert. Am nächsten Morgen wurde die Membran zweimal mit 2 x SSC, 0,1% SDS gewaschen, dann dreimal für je 30 min mit 0,1 x SSC, 0,1% SDS bei 65°C gewaschen. Die Membran wurde in Klarsichtfolie gewickelt und 5-6 h bei -80°C exponiert und anschließend dokumentiert.

## **5.2 Genetische Methoden**

### **5.2.1 Fliegenzucht**

Sofern keine anderen Angaben gemacht werden, wurden alle Fliegenstämme auf Futtermedium in Standard Plastiktöpfen bei 25°C gehalten. Für Kreuzungen wurden frisch geschlüpfte, jungfräuliche Weibchen (max. 10 h nach Verlassen des Pupariums bei 18°C) und Männchen verschiedener Altersstufen verwendet.

### **5.2.2 Futtermasse**

Die Fliegen wurden auf Futtermasse gehalten, die nach folgendem Rezept zubereitet wurde:

8 l Wasser

64 g Fadenagar

144 g Trockenhefe

80 g Sojamehl

176 g Rübensirup

640 g Maismehl

640 g Malzextrakt

19,2 g Nipagin (Methyl-4-hydroxybenzoat; Merck), in Ethanol (95%) gelöst

Propionsäure; Volumen abhängig vom pH-Wert: 16,8 ml bei pH 5, 61,2 ml bei pH 5,5, 105 ml bei pH 6.

Der Agar wurde in kochendem Wasser gelöst; alle Zutaten bis auf die Propionsäure und das Nipagin wurden dazugegeben, und 10 min gekocht. Nach Abkühlen des Breis auf ca. 60°C wurde die Propionsäure zusammen mit dem Nipagin hinzugefügt und der fertige Brei in Polystyrolröhrchen (Greiner) portioniert. Nach Trocknung über Nacht wurden die Futterröhrchen bei 4°C gelagert.

### **5.2.3 Essig-Agarplatten**

Fliegen in Ablagekäfigen wurden auf Essig-Agarplatten gehalten, die nach folgendem Rezept zubereitet wurden:

18 g Agar-Agar

in 500 ml Wasser aufkochen, bis der Agar gelöst ist.

10 g Saccharose

2,5 ml 100% Essigsäure

auf 1 l auffüllen und in Petrischalen gießen.

### **5.2.4 Sammeln von Embryonen**

Für Einzel-Eiablagen wurden die Fliegen in ein leeres, mit Luftlöchern versehenes Plastikgefäß überführt. Abgedeckt wurden diese Gefäße mit Deckeln, die zuvor mit Essig-Agar ausgegossen und mit etwas frischer Hefe bestrichen wurden. Nach der gewünschten Ablagezeit wurden die Essig-Agarplatten gewechselt. Die Embryonen wurden mit einem Pinsel unter fließendem Leitungswasser in Siebe überführt, gewaschen, und entsprechend der jeweiligen Anwendung weiter behandelt.

## 5.3 *Histologische Methoden*

### 5.3.1 *In-situ* Hybridisierungen von Embryonen

Die Proben für RNA basieren auf den Amplikons der Heidelberg Collection. Die notwendige T7 Promoterbindungsstelle wurde durch Reamplifikation des ersten PCR Produktes des Microarray Sets mit Hilfe von Tag-Primern generiert, welche die Erkennungsstelle des T7 Promoter besitzt.

#### 5.3.1.1 *In-vitro* Transkription

Die *in-vitro* Transkription wurde in 96-well Platten in folgendem Ansatz pro well durchgeführt:

1  $\mu$ g aufgereinigtes PCR Produkt wurde mit

10 mM rNTPs, (Roche)

3.5 mM Digoxigenin-11-UTP (Roche)

1 x Transkriptionspuffer (Roche)

20 U RNAsin (Promega)

40 U T7 Polymerase (Roche) gemischt und bei 37° C über Nacht inkubiert.

Nach Zugabe von 40 U DNase und 20 minütiger Inkubation bei 37° C wurde eine Lithiumfällung durchgeführt.

Die Platten wurden für 20 min bei 3500 Upm bei 4°C zentrifugiert, die Pellets mit 80% Ethanol gewaschen, nochmals für 5 min bei 3500 Upm zentrifugiert, die Flüssigkeit abgenommen. Danach wurden die Platten kurz auf dem Kopf bei 500 Upm zentrifugiert, um überschüssige Flüssigkeit zu entfernen. Nach kurzem Trocknen wurden die Pellets in 50  $\mu$ l H<sub>2</sub>O aufgenommen, ein 5  $\mu$ l Aliquot wurde auf einen Agarosegel getestet und die verbleibende Probe mit 50  $\mu$ l Formamid versetzt. Die fertige Probe kann bei –80° C für mehrere Monate aufbewahrt werden.

#### 5.3.1.2 Vorbereitung der Embryonen und Hybridisierung

Oregon R Embryos aus über Nacht Ablagen wurden mit 0-4 h Ablagen gemischt. Embryonen wurden chemisch mit 4% NaOCl dechorionisiert und mit 1 ml Fix-Lösung und 6 ml Heptan für 20 min fixiert. Durch Zugabe von 10 ml Methanol und kräftiges

Schütteln für 1 min wurde die Vitellinmembran entfernt. Die Embryonen wurden danach dreimal mit Methanol gewaschen, und bis zur Hybridisierung in Methanol bei  $-20^{\circ}\text{C}$  gelagert. Für die Hybridisierung wurden die Embryonen mit 3:1 Methanol/PBS und danach mit 1:3 Methanol/PBS rehydriert, bevor sie mit 4% Formaldehyd/PBS ein weiteres Mal für 20 min fixiert wurden. Die Embryonen wurden anschließend 3x für 20 min mit PBS gewaschen, anschließend in 3 ml Hybridisierungspuffer aufgenommen und für 1 h bei  $60^{\circ}\text{C}$  vorinkubiert. Je 50  $\mu\text{l}$  Embryonen wurden danach in Eppendorfgefäße bei  $60^{\circ}\text{C}$  überführt, 30  $\mu\text{l}$  Hybridisierungspuffer und 5  $\mu\text{l}$  der auf  $60^{\circ}\text{C}$  vorgewärmten Probe wurden hinzugefügt. Die Hybridisierung erfolgte über Nacht bei  $60^{\circ}\text{C}$ .

### 5.3.1.3 Nachweis der DIG-markierten RNA Sonde durch Immunfärbung

Die DIG-markierte RNA Sonde wurde über das Anti-Digoxygenin-AP Fab Fragment von Roche und einer anschließenden NBT/BCIP-Färbung nachgewiesen. Um den Hintergrund des Antikörpers zu minimieren, wurde dieser mit nicht hybridisierten Embryonen in einer 1:200 Verdünnung mit PBT für 1,5 h präabsorbiert. Für die eigentliche Immunfärbung wurde der präabsorbierte Antikörper nochmals 1:10 in PBT verdünnt und für 1,5 h auf dem Überkopftrotierer inkubiert. Nach zweimaligen Spülen mit PBT und 3 Waschschritten von je 15 min konnte mit der Färbereaktion begonnen werden.

Die Färbereaktion fand im AP (alkalischer Phosphatase)-Puffer statt. Nach dreimaligem Waschen der Embryonen in AP-Puffer wurde das Substrat für die alkalische Phosphatase NBT/BCIP hinzugegeben. Die NBT/BCIP-Stammlösung (Roche) wurde 1:50 in AP-Puffer verdünnt. Die Färbereaktion wurde in 1,9  $\text{cm}^2$  Gewebekulturschalen vorgenommen und gelegentlich unter der Stereolupe beobachtet (durchschnittliche Färbezeit ca. 30 min - 5 h, Färbereaktion ist lichtempfindlich). Die Färbung wurde durch dreimaliges Waschen in PBT abgebrochen. Unspezifischer Hintergrund wurde durch waschen in 100% Ethanol entfernt. Auf dem Objektträger wurden die Embryonen in 80% Glycerin eingebettet und im Durchlichtmikroskop analysiert.

### 5.3.2 Untersuchung des Alters von Embryonen

Um das genau Alter der Embryonen zu untersuchen, die für die Synthese von RNA für spezifische Entwicklungsstufen dienen, wurde ein Aliquot von Embryonen entnommen und mit Hoechst 33258 gefärbt. Dazu wurden die Embryonen von der Essigagarplatte in ein Waschkörbchen gewaschen, mit 3% Natriumhypochlorid chemisch dechorionisiert, gründlich mit Eggwash gewaschen und in ein Eppendorfgefäß überführt. Das Eggwash wurde abgenommen und die Embryonen mehrmals mit 1 ml PBS gewaschen. Danach wurden die Embryonen mit HeptanFix/Methanol PBS für 15 min fixiert. Das PBS wurde anschließend abgenommen und durch Methanol ersetzt. Zur Devitellinisierung der Embryonen wurden die Eppendorfgefäße für 15 Sekunden kräftig geschüttelt und dabei devitellinisiert. Danach wurden die Embryonen 2 x mit Methanol und anschließend 3 x mit PBS gewaschen. Nach dem letzten Waschschrift wurde das Volumen mit PBS auf 0,5 ml eingestellt, 250  $\mu$ l 1  $\mu$ g/ml Hoechst in PBS zugegeben und für 4 min im Dunkeln färben gelassen. Die Embryonen wurden danach 3 x für 5 min mit je 500  $\mu$ l PBS gewaschen. Anschließend wurden die Embryonen in 80% Glycerin eingebettet, auf einen Objektträger gegeben und lichtmikroskopisch untersucht. Je 100 Embryonen wurden auf ihr Alter überprüft. Die Embryonen wurde nur dann für die Synthese von RNA verwendet, wenn mehr als 95% im entsprechenden Entwicklungsstadium waren.

## 5.4 *Microarray-spezifische Methoden*

### 5.4.1 Fusion zweier unabhängiger Genvorhersagen

Für die Herstellung eines Transkriptom-Micorarrays wurden in einer Kollaboration mit dem Max-Planck Institut für molekulare Genetik in Berlin die Genvorhersagen der Version 2 des Berkeley *Drosophila* Genom Projektes mit der Vorhersage der fgenesh Software vereint. Da beide Programme unterschiedliche Algorithmen verwenden und ursprünglich auf verschiedene Organismen optimiert wurden, variieren sie in der Vorhersage der Anzahl der Gene und der Genmodelle. Für den Abgleich der beiden Datensätze wurden zunächst alle vorhergesagten Exons entsprechend ihrer Transkriptionsrichtung miteinander verglichen. Bei mehr als 30% Überschneidung der exonischen Bereiche wurden die Vorhersagen als ein Gen angesehen, und nur ein

Primerpaar für diese Region berechnet. Zusätzlich zu den Genen des Berkeley *Drosophila* Genom Projektes wurden auch die Gene der Berkeley *Drosophila* Gene Collection (Stapleton et al. 2002) in diesen Vergleich einbezogen. Noch während der Herstellung des *Drosophila* Microarrays wurde vom Berkeley *Drosophila* Genom Projekt eine neue Version ihrer Genvorhersage herausgegeben (Celniker et al. 2002). Diese neuen Daten wurden übernommen und unsere Berechnungen auf diese neue Anzahl und Vorhersage von Genen abgeglichen. Im Folgenden werden sich alle genannten Zahlen auf diese neue Version beziehen.

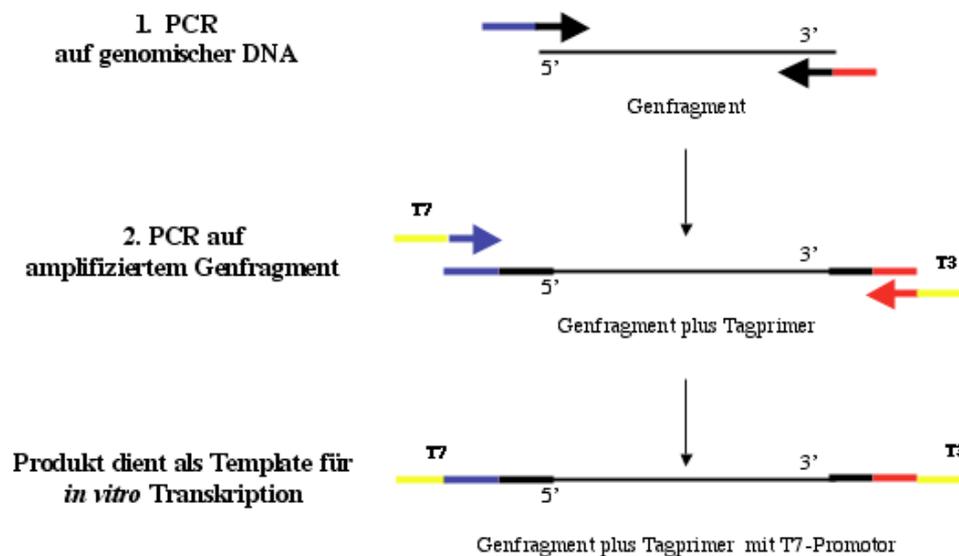
#### **5.4.2 Die Berechnung von Primern zur Synthese von PCR-Fragmenten für vorhergesagte Gene**

Das Primerdesign wurde ebenfalls am Max-Planck Institut mit der dort entwickelten GenomePride Software (Haas et al. 2003) durchgeführt. Um zu vermeiden, daß Primer mehr als eine Zielsequenz auf dem Genom haben, wurde ein „all-against-all“ Vergleich der exonischen Sequenz aller Gene durchgeführt. So konnten alle Regionen, die mehr als 70% Ähnlichkeit zueinander aufweisen, detektiert, und damit Domänen, konservierte Regionen oder repetitive Elemente von dem Primerdesign ausgeschlossen werden. Zusätzliche Ansprüche an das Programm waren – sofern möglich – der Ausschluss von Introns sowie eine optimale Gesamtproduktlänge von etwa 500 bp. Des Weiteren sollten die Primer möglichst auf einem mittleren Exon liegen. Für fehlende Primerpaare, die aufgrund repetitiver Sequenzen bzw. wegen nicht ausreichender Länge der einzelnen Exons nicht berechnet werden konnten, wurden, ebenfalls mit GenomePride, lange Oligomere von 65 bp Länge berechnet. Die Primer wurden von der Firma Eurogentec synthetisiert und in 96-well Platten mit einer Konzentration von 100  $\mu$ M ausgeliefert.

#### **5.4.3 Die PCR und das PrimerTag-System**

Die Amplifikation aller Genfragmente erfolgte über zwei Schritte (Abb. 27). Im ersten Schritt wurde eine PCR mit den spezifisch für jedes Gen berechneten Primern auf genomischer DNA durchgeführt. Jedem 5' Primer war zuvor bei der Synthese eine einheitliche Tagsequenz, sowie jedem 3' Primer eine von neun unterschiedlichen Tagsequenzen angehängt worden. In einer zweiten PCR konnten nun alle Fragmen-

te mit nur zehn Primern amplifiziert werden. Zusätzlich zu dem Vorteil, daß der gesamte Satz an Genen schnell und mit wenig Aufwand reamplifiziert werden kann und ein zweiter PCR Schritt auf vorselektiertem Template mehr Produkt liefert, vermeidet man Kontamination mit genomischer DNA in den Proben, die auf den Microarray gespottet werden sollen. Dabei wird die mögliche Kreuzkontamination zwischen benachbarten Proben durch den Gebrauch der neun verschiedenen Primer limitiert. Die Synthese des ersten PCR Produktes erfolgte in enger Kollaboration mit der Gruppe Hoheisel am DKFZ, die zweite PCR wurde sowohl am DKFZ als auch am ZMBH für unterschiedliche Applikationen mit leicht unterschiedlichen Protokollen durchgeführt. Für die Microarrays, die zur Validierung der Genvorhersagen dienten, wurden Produkte der am DKFZ durchgeführten zweiten PCR verwendet, wohingegen alle weiteren verwendeten Microarrays am ZMBH mit den eigenen Produkten gespottet wurden.



**Abb. 27 Das Tag-System.** Während einer ersten PCR werden über die speziellen Primer an jedes Produkt Tagsequenzen angehängt. Spezifische Primer sind als schwarze Pfeile dargestellt, die einheitliche 5'-Tagsequenz ist in blau, die neun unterschiedlichen 3'-Tagsequenzen sind in rot dargestellt (obere Zeile). Eine Reamplifikation des gesamten Gensatzes kann in einer zweiten PCR Runde mit Primern für diese Tagsequenzen erfolgen (blauer Pfeil, mittlere Zeile). Zusätzlich wurden für weitere Applikationen T3 und T7 Sequenzen an die Tagprimer angefügt (T3 und T7 Sequenzen, beide in gelb dargestellt, mittlere Zeile). Das entstehende Amplikon kann direkt auf den Microarray gespottet werden oder alternativ für *in-vitro* Transkription verwendet werden.

#### 5.4.4 Die erste PCR

Die PCR Amplifikation erfolgte in 96-well Platten in einem Gesamtvolumen von 50  $\mu$ l. Die Fragmente wurden auf genomischer DNA (Oregon R) mit je 20 pmol der spezifischen Primer für jedes Fragment durchgeführt. 100 ng genomische DNA, 1 x Qiagen PCR Puffer, je 40  $\mu$ M dNTPs, 1U Taq-Polymerase (Qiagen) wurden gemischt und in einem Thermocycler für 5 min bei 95°C inkubiert, gefolgt von 35 Zyklen a 95°C für 30 sek zur Denaturierung, 30 sek Annealing und 90 sek bei 72°C zur Elongation des Produktes. Die Annealingtemperatur wurde dabei in den ersten 10 Zyklen von 65°C auf 55°C reduziert, um die Spezifität der Amplifikation zu erhöhen. In den letzten 20 Zyklen wurde die Elongationszeit um 5 sek pro Zyklus erhöht, um die nachlassende Aktivität des Enzyms zu kompensieren.

#### 5.4.5 Die zweite PCR

1  $\mu$ l des Produktes aus der ersten PCR wurde in einer zweiten PCR mit einem Gesamtvolumen von 100  $\mu$ l eingesetzt. 1 x Qiagen PCR Puffer, je 0,1 mM dNTPs, 5 pmol pro Primer und 2 U Qiagen Taq-Polymerase wurden dazugegeben und der Ansatz in einem Thermocycler bei 95°C für 5 min inkubiert. Anschließend folgten 35 Zyklen a 30 sek bei 95°C zur Denaturierung, 30 sek bei 60°C zum Annealing der Primer an die DNA, und 90 s bei 72°C zur Elongation des Fragmentes. In den letzten 20 Zyklen wurde die Elongationszeit um 5 sek in jedem Zyklus verlängert, um die nachlassende Aktivität des Enzyms zu kompensieren. Die PCR Produkte für beide PCRs wurde auf 1% Agarosegelen überprüft und negative PCRs wurden einzeln wiederholt.

#### 5.4.6 Kontrollen

Als Kontrollen wurden Fragmente von *C. elegans* Genen sowie weiteren in *Drosophila* nicht vorhandenen Genen sowie genomische *Drosophila* DNA verwendet. Alle Kontrollen wurden mit spezifischen Primern in einer PCR amplifiziert. Anschließend wurden die *C. elegans* Kontrollen in den Psp64 Expressionsvektor (Promega), und alle weiteren Kontrollen in den PCR II - TOPO Vektor kloniert. Die *C.elegans* Gene können als sogenannte spike- Kontrollen zur quantitativen Analyse dienen, indem eine bekannte Menge *C.elegans* RNA zu der zu untersuchenden RNA

gegeben wird. Eine Liste der Kontrollen sowie deren Häufigkeit auf dem Microarray ist der Tabelle 8 zu entnehmen.

Name	Organismus/Verwendungszweck	Anzahl der Spots auf dem Microarray
F58g4	C. elegans (spike-Kontrolle)	24
Zk909	C. elegans (spike-Kontrolle)	24
T19d12	C. elegans (spike-Kontrolle)	24
C09g5	C. elegans (spike-Kontrolle)	24
Ah6	C. elegans (spike-Kontrolle)	24
T16h12	C. elegans (spike-Kontrolle)	24
F54g8	C. elegans (spike-Kontrolle)	24
C05b5	C. elegans (spike-Kontrolle)	24
T07c4	C. elegans (spike-Kontrolle)	24
Zk1010	C. elegans (spike-Kontrolle)	24
CAT/promega	Kontrolle	12
EGFP	Kontrolle	12
Luciferase firefly	Kontrolle	12
VP16-ActDM	Kontrolle	12
Amp r	Kontrolle	12
Beta-Gal	Kontrolle	12
Tet-r	Kontrolle	12
white	Kontrolle	12
Gal4 AD	Kontrolle	12
Dnmt3a	Kontrolle	24
hrGFP	Kontrolle	24

**Tabelle 8 Kontrollen für den Transkriptomarray.** Verschiedene Sequenzen von *C. elegans*, die keine Ähnlichkeit zu *Drosophila* Sequenzen haben, wurden als Spike Kontrollen verwendet. Weitere in *Drosophila* nicht vorkommenden Sequenzen dienten als Negativ-Kontrollen.

### 5.4.7 Microarrayherstellung

Je 50  $\mu$ l des Materials aus der zweiten PCR wurde in 384-well Platten (V-bottom, Genetix) überführt. Dazu wurden je 4 96-well Platten reihenweise versetzt in einer 384er Platten kombiniert. Die Platten wurden im Abzug für zwei Wochen gelagert und eingetrocknet, bevor sie in 10  $\mu$ l 1x Spottingsolution aufgenommen wurden. Das Spotten des Transkriptomarrays erfolgte in einem Biogrid II (Biorobotics).

Die DNA wurde mit 2 x 12 Nadeln (SMP3 pins, TeleChem) auf aminosilanisierte QMT Aminoslides (Quantifoil) aufgetragen. Der Spotdurchmesser betrug etwa 80  $\mu$ m, und der Abstand der Mittelpunkte der Spots betrug 140  $\mu$ m. Das Spotten erfolgte bei RT und 30-40% Luftfeuchtigkeit.

Die 21,306 Amplikons und die Kontroll-DNAs sind in Duplikaten auf dem Microarray vertreten; insgesamt enthält die erste Version des HeidelbergFlyArrays 47,616 Frag-

mente, die zweite Version aufgrund einer größeren Anzahl an Kontrollen 52,224 Fragmente.

Nach dem Spotten wurden die Microarrays über Nacht bei RT gelagert. Am nächsten Morgen wurden die Arrays in einem UV-Crosslinker zweimal mit einer Energie von  $250 \times 100 \text{ J/cm}^2$  behandelt und in einem Excikator mit Trockenmittel gelagert.

#### **5.4.8 Entwicklungsprofil von *Drosophila melanogaster***

Für das Entwicklungsprofil wurde der Lebenszyklus von *Drosophila* in neun Stadien aufgeteilt. Vier Embryonalstadien mit je einer Zeitspanne von vier Stunden wurden aus grossen Käfigen gesammelt. Von jedem dieser Stadien wurden Aliquots entnommen und morphologisch auf ihr Alter überprüft. Ablagen, in denen weniger als 95% der Embryonen dem erwarteten Alter entsprachen, wurden verworfen. Drei Larvalstadien wurden von Hand gesammelt, wobei diese Stadien zeitlich unterschieden wurden. Das erste Stadium umfasste die Zeiten von 0 bis 24 h nach Verlassen der Eihülle, das zweite Stadium von 24 bis 48 h und das dritte Stadium bestand aus Larven, die bereits an der Gefässwand hochwanderten. Auch drei Pupalstadien wurden von Hand abgesammelt. Puppen des ersten Stadiums wurden direkt nach der Verpuppung bis zu 24 h gesammelt, Puppen des zweiten Stadiums am zweiten Tag bis zu 48 h und Puppen des dritten Stadiums befanden sich kurz vor dem Schlüpfen, d.h. Augen und Flügel waren bereits gut sichtbar ausgebildet. Adultstadien wurden, nach Männchen und Weibchen sortiert, in virginelle Tiere, Fliegen im Alter bis zu 5 Tagen und ältere Tiere unterteilt. Von allen Stadien wurde RNA isoliert und mittels des Agilent Systems auf ihre Qualität überprüft. RNAs mit hoher Qualität wurden gesammelt und bei  $-80^\circ\text{C}$  gelagert.

Für das Entwicklungsprofil wurden jeweils gleiche Mengen der drei larvalen Stadien bzw. der verschiedenen Adultstadien zu einer Probe vereint, so daß insgesamt neun Stadien für die folgenden Versuche zur Verfügung standen.

#### **5.4.9 Untersuchung Pc-mutanter Fliegen auf Microarrays**

Zunächst wurde die Mutante mit einem Balancerchromosom gekreuzt, welches das grün leuchtende Protein GFP unter der Kontrolle des *Twist* Promoters exprimiert. Ab dem Stadium neun ist in allen Embryonen, die das Balancerchromosom tragen, GFP

sichtbar, und homozygot mutante Embryonen können so unter dem Fluoreszenzmikroskop aussortiert werden. Da auch alle unbefruchteten Embryonen und solche, die jünger als vier Stunden sind nicht leuchten, müssen die Embryonen beim sammeln auf ihr Alter überprüft werden. Aus 800 Embryonen des Stadiums 10, die per Hand gesammelt wurden, wurde total-RNA isoliert. Die gleiche Vorgehensweise wurde zur Kontrolle mit wildtypischen Embryonen wiederholt.

Je 1,5  $\mu$ g der RNA wurde über das Eberwine Protokoll linear amplifiziert und für Microarrayexperimente verwendet (Van Gelder et al. 1990; Eberwine et al. 1992).

#### 5.4.10 Modifiziertes Eberwine Protokoll

In einem PCR Reaktionsgefäß wurden 1-2  $\mu$ g RNA mit 9  $\mu$ l 1,7 M Trehalose und H<sub>2</sub>O auf 9  $\mu$ l gebracht, 1  $\mu$ l Eberwine oligodT/T7Primer (1 $\mu$ g/ $\mu$ l) wurde dazugegeben und das Gemisch für 10 min bei 65°C inkubiert. Anschließend wurden auf Eis 4  $\mu$ l Erststrangpuffer, 2  $\mu$ l DTT (0,1 M) 1  $\mu$ l RNAsin (20U/ $\mu$ l), 1  $\mu$ l dNTPs (je 10 mM), 1  $\mu$ l lineares Acrylamid (0,1  $\mu$ g/ $\mu$ l) und 1  $\mu$ l Reverse Transkriptase (200 Units/ $\mu$ l) dazugegeben. In einer PCR Maschine wurde die Reaktion zunächst für 10 min bei 37°C, dann für 10 min bei 65°C und dann für jeweils 10 Zyklen wechselnd bei 60°C für 2 min und 55°C für 2 min inkubiert.

Auf Eis wurden dann die Komponenten für die cDNA Zweitstrangsynthese hinzugefügt. Dazu wurden nacheinander 106  $\mu$ l H<sub>2</sub>O, 15  $\mu$ l Zweitstrangpuffer, 3  $\mu$ l 10 mM dNTPs, 1  $\mu$ l E. coli DNA Ligase (10 U/ $\mu$ l), 4  $\mu$ l E.coli DNA Polymerase I—Holoenzym (10 U/ $\mu$ l) und 1  $\mu$ l RNase H (2 U/ $\mu$ l) zu dem cDNA Erststrang gegeben und für 2 h bei 16°C inkubiert. Anschließend wurde die cDNA über Phenol-Chloroform extrahiert, und mit 4 M Ammoniumacetat präzipitiert. Die gefällte, mit 80% Ethanol gewaschene und in 30  $\mu$ l H<sub>2</sub>O aufgenommene cDNA wurde mit Biospinsäulen ein weiteres Mal gereinigt. Das Produkt wurde in der Speedvac auf 16  $\mu$ l eingengt und für eine *in-vitro* Transkription eingesetzt. Die *in-vitro* Transkription erfolgte in einem doppelten Ansatz entsprechend den Herstelleranweisungen (ambion).

Am nächsten Morgen wurde die RNA über Phenol - Chloroform Extraktion aufgereinigt und die wässrige Phase direkt auf Biospinsäulen gegeben. Die aufgereinigte RNA wurde auf einem Agarosegel überprüft (1  $\mu$ l RNA auf einem 1%

Agarosegel). mindestens 5 µg RNA wurden für ein Microarrayexperiment verwendet. (Van Gelder et al. 1990; Eberwine et al. 1992).

#### 5.4.11 Hybridisierung der Microarrays

In Kollaboration mit der Gruppe Hoheisel am DKFZ wurden die Hybridisierungen der Microarrays durchgeführt. Für alle Stadien wurden Zweifarbenexperimente durchgeführt, wobei das Embryonalstadium 0-4 h als Kontrollstadium für die acht weiteren Stadien diente. 20 µg Total RNA wurden – entsprechend den Anweisungen des Herstellers - indirekt über eine Aminoallylbindung mit Cy3 oder Cy5 markiert (Invitrogen). Die zusammenpipettierten cDNAs wurden mit 2,5 µl Heringssperma-DNA versetzt und in einer Speedvac eingetrocknet. Zur gleichen Zeit wurde der Microarray zunächst in 0,1 x SSC für 1min gewaschen, dann einmal in H<sub>2</sub>O getaucht und danach 3 min in kochendem H<sub>2</sub>O denaturiert. Der Microarray wurden dann direkt in 50 ml einer Prähybridisierungslösung bei 42°C für eine Stunde geblockt. Anschließend wurde der Array kurz in H<sub>2</sub>O geschwenkt und durch anschließende Zentrifugation getrocknet. Die eingetrockneten Proben wurden in Hybridisierungspuffer aufgenommen, bei 95°C für 5 min denaturiert und danach auf ein Deckglass gegeben, auf welches der prähybridisierte Microarray vorsichtig aufgelegt werden. Der Array wurde in Atlas Hybridisierungskammern (TeleChem) gelegt; in die seitlichen Wells wurden jeweils 18 µl 6 x SSC gegeben, und die Kammer gut verschlossen. Die Hybridisierung erfolgte in einem Wasserbad bei 55°C für 16 h. Nach 16 h wurden die Arrays kurz in Waschpuffer A geschwenkt, um das Deckglas zu entfernen. Anschließend erfolgten Waschschrte für jeweils 10 min in Waschpuffer A, Waschpuffer B und Waschpuffer C gewaschen und in einer Zentrifuge für 1 min bei 2,000 Upm getrocknet. Für jedes der acht Stadien wurden mindestens drei unabhängige Wiederholungen durchgeführt. Aufgrund der unterschiedlichen Einbau-rate der beiden Farbstoffe wurde mindesten in einer der Wiederholungen der Farbstoff der Probe mit dem der Kontrolle ausgetauscht.

#### 5.4.12 Filterkriterien

Die Glassarrays wurden in der Gruppe Hoheisel mit dem ScanArray 5000 (Perkin Elmer, Wellesley, USA) und der entsprechenden Software (Version 3.1.) gescannt.

Externe, fluoreszent markierte Kontrollen dienten als Referenz für die Einstellungen des Lasers und des Photomultipliers. Die erhaltenen Abbildungen der einzelnen Kanäle von Cy3 (532 nm) und Cy5 (632 nm) wurden mit der Genepix Software (Version 4.0, Axon Instruments) weiter bearbeitet. Für jedes Gen wurde der Intensitätswert berechnet und im gpr Format gespeichert. So entstandene Ergebnisdokumente enthalten Informationen über Gen- bzw. Klonnamen, die Lokalisation des entsprechenden Spots auf dem Array sowie deren Signalintensitätswerte (Mittelwert und Median der Pixel), Hintergrundintensitäten, Pixelzahl der Spots sowie die Standardabweichung der beiden Farbstoffe. Als erstes Qualitätsmerkmal dienten die Duplikate auf dem Microarray. Die Standardabweichung der Verhältnisse der beiden Farbstoffe (Cy5/Cy3) diente dabei als Kriterium für reproduzierbare Hybridisierungskriterien innerhalb der einzelnen Microarrays. Arrays wurden nur dann für weitere Auswertungen verwendet, wenn mindestens 30% aller Gene eine Standardabweichung der  $\log(632/532)$  von weniger als 1/3 zwischen den Replikaten hatte.

### **5.4.13 Normalisierung und Datenanalyse**

#### **5.4.13.1 Normalisierung der Entwicklungsprofil-Experimente**

Die Normalisierung der Microarrays erfolgte am DKFZ in der Gruppe von Jörg Hoheisel, mit der dort entwickelten Software M-CHiPS (Multi-Conditional Hybridization Intensity Processing Software (Fellenberg et al. 2002). Dazu wurde eine lineare Regressionsanalyse verwendet, sowie unterschiedliche Normalisierungen. Die gefilterten Daten wurden in der Korrespondenzanalyse weiter analysiert (Fellenberg et al. 2001).

Bei der Korrespondenzanalyse handelt es sich um eine explorative Projektionsmethode, welche die Assoziationen zwischen Genen und Hybridisierungen in einem Plotdiagramm darstellt. Es ist gut für die Analyse großer Datensätze geeignet, da jede Expressionsintensität mit der entsprechenden Hybridisierung in eine Dimension gespeichert wird. Daher kann die Qualität der Hybridisierungen in multidimensionalen Experimenten validiert werden, wenn die wiederholten Experimente einer Kondition (in unserem Fall Entwicklungsstadien) Cluster bilden. Gene mit ähnlichem Expressionsprofil haben geringe Distanzen in dem Plot und

werden mit diesen Clustern assoziiert. 24 Hybridisierungen, die den angesprochenen Konditionen genügen, und in M-CHiPS normalisiert worden waren, wurden für diese Analyse ausgewählt.

#### **5.4.13.2 Normalisierung der Pc-Zielgen Analysen**

Für die Normalisierung und Auswertung dieser Experimente wurde die Software „express yourself“ verwendet. Die Genliste mit den Spotintensitäten wurde im gpr-Format geladen. Normalisierung des Hintergrunds, Spot-filtering und Spot-normalisierung wurden nach den Standardbedingungen des Programms durchgeführt. Gene, die mindestens 2-fach in fünf aus sechs Datenpunkten hoch- bzw. herunterreguliert waren, wurden für die weitere Analyse ausgewählt. Die Gesamt-pixelintensität der Spots sollte in beiden Datenpunkten mindestens 500 betragen. Die Gen-IDs der regulierte Gene wurde unter Verwendung der Webseite EnMart auf ihre Go-Annotation und ihre Interpro-Einträge untersucht.

## 6 Referenzen

Aasland, R. und A. F. Stewart (1995). "The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1." Nucleic Acids Res **23**(16): 3168-73.

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, *et al.* (2000). "The genome sequence of *Drosophila melanogaster*." Science **287**(5461): 2185-95.

Akam, M. (1987). "The molecular basis for metamerism in the *Drosophila* embryo." Development **101**(1): 1-22.

Andrews, J., G. G. Bouffard, C. Cheadle, J. Lu, K. G. Becker und B. Oliver (2000). "Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis." Genome Res **10**(12): 2030-43.

Arbeitman, M. N., E. E. Furlong, F. Imam, E. Johnson, B. H. Null, B. S. Baker, M. A. Krasnow, M. P. Scott, R. W. Davis und K. P. White (2002). "Gene expression during the life cycle of *Drosophila melanogaster*." Science **297**(5590): 2270-5.

Baxevanis, A. D. (2001). "Predictive methods using DNA sequences." Methods Biochem Anal **43**: 233-52.

Beisel, C., A. Imhof, J. Greene, E. Kremmer und F. Sauer (2002). "Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1." Nature **419**(6909): 857-62.

Beuchle, D., G. Struhl und J. Müller (2001). "Polycomb group proteins and heritable silencing of *Drosophila* Hox genes". Development **128**: 993-1004

Bloyer, S., G. Cavalli, H. W. Brock und J. M. Dura (2003). "Identification and characterization of polyhomeotic PREs and TREs." Dev Biol **261**(2): 426-42.

Boutros, M., A. A. Kiger, S. Armknecht, K. Kerr, M. Hild, B. Koch, S. A. Haas, H. F. Consortium, R. Paro und N. Perrimon (2004). "Genome-wide RNAi analysis of growth and viability in *Drosophila* cells." Science **303**(5659): 832-5.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, *et al.* (2001). "Minimum information about a microarray experiment (MIAME)-toward standards for microarray data." Nat Genet **29**(4): 365-71.

- Breen, T. R. und P. J. Harte (1991). "Molecular characterization of the trithorax gene, a positive regulator of homeotic gene expression in *Drosophila*." Mech Dev **35**(2): 113-27.
- Breiling, A., B. M. Turner, M. E. Bianchi und V. Orlando (2001). "General transcription factors bind promoters repressed by Polycomb group proteins." Nature **412**(6847): 651-5.
- Brown, J. L., C. Fritsch, J. Mueller und J. A. Kassis (2003). "The *Drosophila* pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing." Development **130**(2): 285-94.
- Buchenau, P., J. Hodgson, H. Strutt und D. J. Arndt-Jovin (1998). "The distribution of polycomb-group proteins during cell division and development in *Drosophila* embryos: impact on models for silencing." J Cell Biol **141**(2): 469-81.
- Bullock, W. E. und S. D. Wright (1987). "Role of the adherence-promoting receptors, CR3, LFA-1, and p150,95, in binding of *Histoplasma capsulatum* by human macrophages." J Exp Med **165**(1): 195-210.
- Burge, C. und S. Karlin (1997). "Prediction of complete gene structures in human genomic DNA." J Mol Biol **268**(1): 78-94.
- Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones und Y. Zhang (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science **298**(5595): 1039-43.
- Casso, D., F. Ramirez-Weber und T. B. Kornberg (2000). "GFP-tagged balancer chromosomes for *Drosophila melanogaster*." Mech Dev **91**(1-2): 451-4.
- Celniker, S. E., D. A. Wheeler, B. Kronmiller, J. W. Carlson, A. Halpern, S. Patel, M. Adams, M. Champe, S. P. Dugan, E. Frise, *et al.* (2002). "Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence." Genome Biol **3**(12): RESEARCH0079.
- Chan, C. S., L. Rastelli und V. Pirrotta (1994). "A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression." Embo J **13**(11): 2553-64.
- Chang, Y. L., B. O. King, M. O'Connor, A. Mazo und D. H. Huang (1995). "Functional reconstruction of trans regulation of the Ultrabithorax promoter by the products of two antagonistic genes, trithorax and Polycomb." Mol Cell Biol **15**(12): 6601-12.
- Chinwalla, V., E. P. Jane und P. J. Harte (1995). "The *Drosophila* trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites." Embo J **14**(9): 2056-65.

- Claverie, J. M. (1997). "Computational methods for the identification of genes in vertebrate genomic sequences." Hum Mol Genet **6**(10): 1735-44.
- Claverie, J. M., O. Poirot und F. Lopez (1997). "The difficulty of identifying genes in anonymous vertebrate sequences." Comput Chem **21**(4): 203-14.
- Consortium, T. C. e. S. (1998). "Genome sequence of the nematode *C. elegans*: a platform for investigating biology." Science **282**(5396): 2012-8.
- Cornell, R. A. und T. V. Ohlen (2000). "Vnd/nkx, ind/gsh, and msh/msx: conserved regulators of dorsoventral neural patterning?" Curr Opin Neurobiol **10**(1): 63-71.
- Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof und V. Pirrotta (2002). "Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites." Cell **111**(2): 185-96.
- D'Alessio, M. und M. Frasch (1996). "msh may play a conserved role in dorsoventral patterning of the neuroectoderm and mesoderm." Mech Dev **58**(1-2): 217-31.
- Daly, M. J. (2002). "Estimating the human gene count." Cell **109**(3): 283-4.
- Davuluri, R. V., Y. Suzuki, S. Sugano und M. Q. Zhang (2000). "CART classification of human 5' UTR sequences." Genome Res **10**(11): 1807-16.
- Deak, P., M. M. Omar, R. D. Saunders, M. Pal, O. Komonyi, J. Szidonya, P. Maroy, Y. Zhang, M. Ashburner, P. Benos, *et al.* (1997). "P-element insertion alleles of essential genes on the third chromosome of *Drosophila melanogaster*: correlation of physical and cytogenetic maps in chromosomal region 86E-87F." Genetics **147**(4): 1697-722.
- DeCamillis, M., N. S. Cheng, D. Pierre und H. W. Brock (1992). "The polyhomeotic gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb." Genes Dev **6**(2): 223-32.
- Diehl, F., B. Beckmann, N. Kellner, N. C. Hauser, S. Diehl und J. D. Hoheisel (2002). "Manufacturing DNA microarrays from unpurified PCR products." Nucleic Acids Res **30**(16): e79.
- Diehl, F., S. Grahlmann, M. Beier und J. D. Hoheisel (2001). "Manufacturing DNA microarrays of high spot homogeneity and reduced background signal." Nucleic Acids Res **29**(7): E38.
- Dietzel, S., H. Niemann, B. Brückner, C. Maurange und R. Paro (1999). "The nuclear distribution of Polycomb during *Drosophila melanogaster* development shown with a GFP fusion protein." Chromosoma **108**(2): 83-94.

- Driever, W. und C. Nüsslein-Volhard (1988). "A gradient of bicoid protein in *Drosophila* embryos." Cell **54**(1): 83-93.
- Driever, W. und C. Nüsslein-Volhard (1989). "The bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo." Nature **337**(6203): 138-43.
- Eberwine, J., H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel und P. Coleman (1992). "Analysis of gene expression in single live neurons." Proc Natl Acad Sci U S A **89**(7): 3010-4.
- Eisen, M. B. und P. O. Brown (1999). "DNA arrays for analysis of gene expression." Methods Enzymol **303**: 179-205.
- Elfring, L. K., C. Daniel, O. Papoulas, R. Deuring, M. Sarte, S. Moseley, S. J. Beek, W. R. Waldrip, G. Daubresse, A. DePace, *et al.* (1998). "Genetic analysis of brahma: the *Drosophila* homolog of the yeast chromatin remodeling factor SWI2/SNF2." Genetics **148**(1): 251-65.
- Faucheux, M., S. Netter, S. Bloyer, M. Moussa, E. Boissonneau, F. Lemeunier, M. Wegnez und L. Theodore (2001). "Advantages of a P-element construct containing MtnA sequences for the identification of patterning and cell determination genes in *Drosophila melanogaster*." Mol Genet Genomics **265**(1): 14-22.
- Fellenberg, K., N. C. Hauser, B. Brors, J. D. Hoheisel und M. Vingron (2002). "Microarray data warehouse allowing for inclusion of experiment annotations in statistical analysis." Bioinformatics **18**(3): 423-33.
- Fellenberg, K., N. C. Hauser, B. Brors, A. Neutzner, J. D. Hoheisel und M. Vingron (2001). "Correspondence analysis applied to microarray data." Proc Natl Acad Sci U S A **98**(19): 10781-6.
- Fodor, S. P., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu und D. Solas (1991). "Light-directed, spatially addressable parallel chemical synthesis." Science **251**(4995): 767-73.
- Francis, N. J., A. J. Saurin, Z. Shao und R. E. Kingston (2001). "Reconstitution of a functional core polycomb repressive complex." Mol Cell **8**(3): 545-56.
- Franke, A., M. DeCamillis, D. Zink, N. Cheng, H. W. Brock und R. Paro (1992). "Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*." Embo J **11**(8): 2941-50.
- Franke, A., S. Messmer und R. Paro (1995). "Mapping functional domains of the polycomb protein of *Drosophila melanogaster*." Chromosome Res **3**(6): 351-60.

- Fritsch, C., J. L. Brown, J. A. Kassis und J. Müller (1999). "The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a *Drosophila* homeotic gene." Development **126**(17): 3905-13.
- Fukumoto, H., B. S. Cheung, B. T. Hyman und M. C. Irizarry (2002). "Beta-secretase protein and activity are increased in the neocortex in Alzheimer disease." Arch Neurol **59**(9): 1381-9.
- Gatta, L. B., A. Albertini, R. Ravid und D. Finazzi (2002). "Levels of beta-secretase BACE and alpha-secretase ADAM10 mRNAs in Alzheimer hippocampus." Neuroreport **13**(16): 2031-3.
- Gaul, U., E. Seifert, R. Schuh und H. Jäckle (1987). "Analysis of Kruppel protein distribution during early *Drosophila* development reveals posttranscriptional regulation." Cell **50**(4): 639-47.
- Gindhart, J. G., Jr. und T. C. Kaufman (1995). "Identification of Polycomb and trithorax group responsive elements in the regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*." Genetics **139**(2): 797-814.
- Gopal, S., M. Schroeder, U. Pieper, A. Sczyrba, G. Aytekin-Kurban, S. Bekiranov, J. E. Fajardo, N. Eswar, R. Sanchez, A. Sali, *et al.* (2001). "Homology-based annotation yields 1,042 new candidate genes in the *Drosophila melanogaster* genome." Nat Genet **27**(3): 337-40.
- Goto, T., P. Macdonald und T. Maniatis (1989). "Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues." Cell **57**(3): 413-22.
- Gray, N. K. (1998). "Translational control by repressor proteins binding to the 5'UTR of mRNAs." Methods Mol Biol **77**: 379-97.
- Haas, S. A., M. Hild, A. P. Wright, T. Hain, D. Talibi und M. Vingron (2003). "Genome-scale design of PCR primers and long oligomers for DNA microarrays." Nucleic Acids Res **31**(19): 5576-81.
- Halfon, M. S., S. Gisselbrecht, J. Lu, B. Estrada, H. Keshishian und A. M. Michelson (2002). "New fluorescent protein reporters for use with the *Drosophila* Gal4 expression system and for vital detection of balancer chromosomes." Genesis **34**(1-2): 135-8.
- Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." J Mol Biol **166**(4): 557-80.

- Harding, K., T. Hoey, R. Warrior und M. Levine (1989). "Autoregulatory and gap gene response elements of the even-skipped promoter of *Drosophila*." Embo J **8**(4): 1205-12.
- Hogenesch, J. B., K. A. Ching, S. Batalov, A. I. Su, J. R. Walker, Y. Zhou, S. A. Kay, P. G. Schultz und M. P. Cooke (2001). "A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes." Cell **106**(4): 413-5.
- Hohn, T., H. S. Park, O. Guerra-Peraza, L. Stavalone, M. M. Pooggin, K. Kobayashi und L. A. Ryabova (2001). "Shunting and controlled reinitiation: the encounter of cauliflower mosaic virus with the translational machinery." Cold Spring Harb Symp Quant Biol **66**: 269-76.
- Holsinger, R. M., C. A. McLean, K. Beyreuther, C. L. Masters und G. Evin (2002). "Increased expression of the amyloid precursor beta-secretase in Alzheimer's disease." Ann Neurol **51**(6): 783-6.
- Howard, K., P. Ingham und C. Rushlow (1988). "Region-specific alleles of the *Drosophila* segmentation gene hairy." Genes Dev **2**(8): 1037-46.
- Ingham, P. W. (1988). "The molecular genetics of embryonic pattern formation in *Drosophila*." Nature **335**(6185): 25-34.
- Ingham, P. W. (1998). "trithorax and the regulation of homeotic gene expression in *Drosophila*: a historical perspective." Int J Dev Biol **42**(3): 423-9.
- Jäckle, H., E. Seifert, A. Preiss und U. B. Rosenberg (1986). "Probing gene activity in *Drosophila* embryos." J Embryol Exp Morphol **97 Suppl**: 157-68.
- Jacobs, J. J. und M. van Lohuizen (1999). "Cellular memory of transcriptional states by Polycomb-group proteins." Semin Cell Dev Biol **10**(2): 227-35.
- Jimenez, F., L. E. Martin-Morris, L. Velasco, H. Chu, J. Sierra, D. R. Rosen und K. White (1995). "vnd, a gene required for early neurogenesis of *Drosophila*, encodes a homeodomain protein." Embo J **14**(14): 3487-95.
- Jones, C. A., J. Ng, A. J. Peterson, K. Morgan, J. Simon und R. S. Jones (1998). "The *Drosophila* esc and E(z) proteins are direct partners in polycomb group-mediated repression." Mol Cell Biol **18**(5): 2825-34.
- Jordan, B. R., R. Jourdan und B. Jacq (1976). "Late steps in the maturation of *Drosophila* 26 S ribosomal RNA: generation of 5-8 S and 2 S RNAs by cleavages occurring in the cytoplasm." J Mol Biol **101**(1): 85-105.
- Jürgens, G. (1985). "A group of genes controlling the spatial expression of the Bithorax complex in *Drosophila*." Nature **316**: 153--155.

- Karlin, S., A. Bergman und A. J. Gentles (2001). "Genomics. Annotation of the *Drosophila* genome." Nature **411**(6835): 259-60.
- Karlstrom, R. O., L. P. Wilder und M. J. Bastiani (1993). "Lachesin: an immunoglobulin superfamily protein whose expression correlates with neurogenesis in grasshopper embryos." Development **118**(2): 509-22.
- Kassis, J. A. (1994). "Unusual properties of regulatory DNA from the *Drosophila* engrailed gene: three "pairing-sensitive" sites within a 1.6-kb region." Genetics **136**(3): 1025-38.
- Kaufman, T. C., M. A. Seeger und G. Olsen (1990). "Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*." Adv Genet **27**: 309-62.
- Kennison, J. A. (1995). "The Polycomb and trithorax group proteins of *Drosophila*: trans-regulators of homeotic gene function." Annu Rev Genet **29**: 289-303.
- Kennison, J. A. (2004). "Introduction to Trx-G and Pc-G genes." Methods Enzymol **377**: 61-70.
- Kim, C. A., M. Gingery, R. M. Pilpa und J. U. Bowie (2002). "The SAM domain of polyhomeotic forms a helical polymer." Nat Struct Biol **9**(6): 453-7.
- Kochetov, A. V., I. V. Ischenko, D. G. Vorobiev, A. E. Kel, V. N. Babenko, L. L. Kisselev und N. A. Kolchanov (1998). "Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features." FEBS Lett **440**(3): 351-5.
- Koizumi, K., C. Lintas, M. Nirenberg, J. S. Maeng, J. H. Ju, J. W. Mack, J. M. Gruschus, W. F. Odenwald und J. A. Ferretti (2003). "Mutations that affect the ability of the vnd/NK-2 homeoprotein to regulate gene expression: transgenic alterations and tertiary structure." Proc Natl Acad Sci U S A **100**(6): 3119-24.
- Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst und D. Reinberg (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein." Genes Dev **16**(22): 2893-905.
- Lavigne, M., N. J. Francis, I. F. King und R. E. Kingston (2004). "Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin." Mol Cell **13**(3): 415-25.
- Lewis, P.H. (1947). New mutants report. D. I. S.

- Lin, X., S. Kaul, S. Rounsley, T. P. Shea, M. I. Benito, C. D. Town, C. Y. Fujii, T. Mason, C. L. Bowman, M. Barnstead, *et al.* (1999). "Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*." Nature **402**(6763): 761-8.
- Llimargas, M., M. Strigini, M. Katidou, D. Karagogeos und J. Casanova (2004). "Lachesin is a component of a septate junction-based mechanism that controls tube size and epithelial integrity in the *Drosophila* tracheal system." Development **131**(1): 181-90.
- Macias, A., J. Casanova und G. Morata (1990). "Expression and regulation of the *abd-A* gene of *Drosophila*." Development **110**(4): 1197-207.
- Maurange, C. und R. Paro (2002). "A cellular memory module conveys epigenetic inheritance of hedgehog expression during *Drosophila* wing imaginal disc development." Genes Dev **16**(20): 2672-83.
- Mazo, A. M., D. H. Huang, B. A. Mozer und I. B. Dawid (1990). "The *trithorax* gene, a trans-acting regulator of the *bithorax* complex in *Drosophila*, encodes a protein with zinc-binding domains." Proc Natl Acad Sci U S A **87**(6): 2112-6.
- Meijer, H. A. und A. A. Thomas (2003). "Ribosomes stalling on uORF1 in the *Xenopus* Cx41 5' UTR inhibit downstream translation initiation." Nucleic Acids Res **31**(12): 3174-84.
- Misra, S., M. A. Crosby, C. J. Mungall, B. B. Matthews, K. S. Campbell, P. Hradecky, Y. Huang, J. S. Kaminker, G. H. Millburn, S. E. Prochnik, *et al.* (2002). "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review." Genome Biol **3**(12): RESEARCH0083.
- Moazed, D. und P. H. O'Farrell (1992). "Maintenance of the engrailed expression pattern by Polycomb group genes in *Drosophila*." Development **116**(3): 805-10.
- Morris, D. R. und A. P. Geballe (2000). "Upstream open reading frames as regulators of mRNA translation." Mol Cell Biol **20**(23): 8635-42.
- Müller, J., C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, M. B. O'Connor, R. E. Kingston und J. A. Simon (2002). "Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex." Cell **111**(2): 197-208.
- Nüsslein-Volhard, C. und E. Wieschaus (1980). "Mutations affecting segment number and polarity in *Drosophila*." Nature **287**(5785): 795-801.
- Orlando, V., E. P. Jane, V. Chinwalla, P. J. Harte und R. Paro (1998). "Binding of *trithorax* and Polycomb proteins to the *bithorax* complex: dynamic changes during early *Drosophila* embryogenesis." Embo J **17**(17): 5141-50.

- Orlando, V. und R. Paro (1995). "Chromatin multiprotein complexes involved in the maintenance of transcription patterns." Curr Opin Genet Dev **5**(2): 174-9.
- Pankratz, M. J., M. Hoch, E. Seifert und H. Jäckle (1989). "Kruppel requirement for knirps enhancement reflects overlapping gap gene activities in the Drosophila embryo." Nature **341**(6240): 337-40.
- Pankratz, M. J., E. Seifert, N. Gerwin, B. Billi, U. Nauber und H. Jäckle (1990). "Gradients of Kruppel and knirps gene products direct pair-rule gene stripe patterning in the posterior region of the Drosophila embryo." Cell **61**(2): 309-17.
- Papoulas, O., S. J. Beek, S. L. Moseley, C. M. McCallum, M. Sarte, A. Shearn und J. W. Tamkun (1998). "The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes." Development **125**(20): 3955-66.
- Paro, R. (1995). "Propagating memory of transcriptional states." Trends Genet **11**(8): 295-7.
- Pesole, G., C. Gissi, G. Grillo, F. Licciulli, S. Liuni und C. Saccone (2000). "Analysis of oligonucleotide AUG start codon context in eukariotic mRNAs." Gene **261**(1): 85-91.
- Peterson, C. L. und I. Herskowitz (1992). "Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription." Cell **68**(3): 573-83.
- Petruk, S., Y. Sedkov, S. Smith, S. Tillib, V. Kraevski, T. Nakamura, E. Canaani, C. M. Croce und A. Mazo (2001). "Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene." Science **294**(5545): 1331-4.
- Poux, S., R. Melfi und V. Pirrotta (2001). "Establishment of Polycomb silencing requires a transient interaction between PC and ESC." Genes Dev **15**(19): 2509-14.
- Preece, P. und N. J. Cairns (2003). "Quantifying mRNA in postmortem human brain: influence of gender, age at death, postmortem interval, brain pH, agonal state and inter-lobe mRNA variance." Brain Res Mol Brain Res **118**(1-2): 60-71.
- Rastelli, L., C. S. Chan und V. Pirrotta (1993). "Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function." Embo J **12**(4): 1513-22.
- Rastelli, L., Chan, CS., Pirrotta V. (1993). "Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function." Embo J **12**: 1513-22.

- Ray, R. P. und T. Schüpbach (1996). "Intercellular signaling and the polarization of body axes during *Drosophila* oogenesis." Genes Dev **10**(14): 1711-23.
- Reboul, J., P. Vaglio, J. F. Rual, P. Lamesch, M. Martinez, C. M. Armstrong, S. Li, L. Jacotot, N. Bertin, R. Janky, *et al.* (2003). "C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression." Nat Genet **34**(1): 35-41.
- Reboul, J., P. Vaglio, N. Tzellas, N. Thierry-Mieg, T. Moore, C. Jackson, T. Shin-i, Y. Kohara, D. Thierry-Mieg, J. Thierry-Mieg, *et al.* (2001). "Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*." Nat Genet **27**(3): 332-6.
- Reese, M. G., D. Kulp, H. Tammanna und D. Haussler (2000). "Genie--gene finding in *Drosophila melanogaster*." Genome Res **10**(4): 529-38.
- Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S., Meisel, R.P., Schaeffer, S.W., Gelbart, W.M., Gibbs, R.A., (2004). Genome sequence and analysis of *Drosophila pseudoobscura*. 45th *Drosophila* Research Conference, Washington.
- Ringrose, L. und R. Paro (2004). "Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins." Annu Rev Genet.
- Ringrose, L., M. Rehmsmeier, J. M. Dura und R. Paro (2003). "Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*." Dev Cell **5**(5): 759-71.
- Rivera-Pomar, R., X. Lu, N. Perrimon, H. Taubert und H. Jackle (1995). "Activation of posterior gap gene expression in the *Drosophila* blastoderm." Nature **376**(6537): 253-6.
- Rogers, G. W., Jr., G. M. Edelman und V. P. Mauro (2004). "Differential utilization of upstream AUGs in the beta-secretase mRNA suggests that a shunting mechanism regulates translation." Proc Natl Acad Sci U S A **101**(9): 2794-9.
- Rogozin, I. B., A. V. Kochetov, F. A. Kondrashov, E. V. Koonin und L. Milanesi (2001). "Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon." Bioinformatics **17**(10): 890-900.
- Rozovskaia, T., S. Tillib, S. Smith, Y. Sedkov, O. Rozenblatt-Rosen, S. Petruk, T. Yano, T. Nakamura, L. Ben-Simchon, J. Gildea, *et al.* (1999). "Trithorax and ASH1 interact directly and associate with the trithorax group-responsive bxd region of the Ultrabithorax promoter." Mol Cell Biol **19**(9): 6441-7.
- Salamov, A. A. und V. V. Solovyev (2000). "Ab initio gene finding in *Drosophila* genomic DNA." Genome Res **10**(4): 516-22.

- Saurin, A. J., Z. Shao, H. Erdjument-Bromage, P. Tempst und R. E. Kingston (2001). "A Drosophila Polycomb group complex includes Zeste and dTAFII proteins." Nature **412**(6847): 655-60.
- Schena, M., D. Shalon, R. Heller, A. Chai, P. O. Brown und R. W. Davis (1996). "Parallel human genome analysis: microarray-based expression monitoring of 1000 genes." Proc Natl Acad Sci U S A **93**(20): 10614-9.
- Shalon, D., S. J. Smith und P. O. Brown (1996). "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization." Genome Res **6**(7): 639-45.
- Shao, Z., F. Raible, R. Mollaaghababa, J. R. Guyon, C. T. Wu, W. Bender und R. E. Kingston (1999). "Stabilization of chromatin structure by PRC1, a Polycomb complex." Cell **98**(1): 37-46.
- Shearn, A. (1989). "The ash-1, ash-2 and trithorax genes of Drosophila melanogaster are functionally related." Genetics **121**(3): 517-25.
- Shoemaker, D. D., E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, P. M. Loerch, A. Leonardson, P. Y. Lum, G. Cavet, *et al.* (2001). "Experimental annotation of the human genome using microarray technology." Nature **409**(6822): 922-7.
- Slifer, E.H. (1942). "A mutant stock of Drosophila with extra sex-combs." J. exp. Zool. **90**:31-40
- Simon, J., A. Chiang, W. Bender, M. J. Shimell und M. O'Connor (1993). "Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products." Dev Biol **158**(1): 131-44.
- Snow, P. M., A. J. Bieber und C. S. Goodman (1989). "Fasciclin III: a novel homophilic adhesion molecule in Drosophila." Cell **59**(2): 313-23.
- Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." J Mol Biol **98**(3): 503-17.
- St Johnston, D. und C. Nüsslein-Volhard (1992). "The origin of pattern and polarity in the Drosophila embryo." Cell **68**(2): 201-19.
- Stapleton, M., J. Carlson, P. Brokstein, C. Yu, M. Champe, R. George, H. Guarin, B. Kronmiller, J. Pacleb, S. Park, *et al.* (2002). "A Drosophila full-length cDNA resource." Genome Biol **3**(12): RESEARCH0080.

- Stapleton, M., G. Liao, P. Brokstein, L. Hong, P. Carninci, T. Shiraki, Y. Hayashizaki, M. Champe, J. Pacleb, K. Wan, *et al.* (2002). "The Drosophila gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes." Genome Res **12**(8): 1294-300.
- Stein, L. D., C. Mungall, S. Shu, M. Caudy, M. Mangone, A. Day, E. Nickerson, J. E. Stajich, T. W. Harris, A. Arva, *et al.* (2002). "The generic genome browser: a building block for a model organism system database." Genome Res **12**(10): 1599-610.
- Struhl, G. (1989). "Morphogen gradients and the control of body pattern in insect embryos." Ciba Found Symp **144**: 65-86; discussion 86-91, 92-8.
- Strutt, H., G. Cavalli und R. Paro (1997). "Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression." Embo J **16**(12): 3621-32.
- Suzuki, Y., D. Ishihara, M. Sasaki, H. Nakagawa, H. Hata, T. Tsunoda, M. Watanabe, T. Komatsu, T. Ota, T. Isogai, *et al.* (2000). "Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries." Genomics **64**(3): 286-97.
- Tautz, D. (1988). "Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres." Nature **332**(6161): 281-4.
- Tie, F., T. Furuyama, J. Prasad-Sinha, E. Jane und P. J. Harte (2001). "The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3." Development **128**(2): 275-86.
- Tie, F., J. Prasad-Sinha, A. Birve, A. Rasmuson-Lestander und P. J. Harte (2003). "A 1-megadalton ESC/E(Z) complex from Drosophila that contains polycomblike and RPD3." Mol Cell Biol **23**(9): 3352-62.
- Tijet, N., C. Helvig und R. Feyereisen (2001). "The cytochrome P450 gene superfamily in Drosophila melanogaster: annotation, intron-exon organization and phylogeny." Gene **262**(1-2): 189-98.
- Tomancak, P., A. Beaton, R. Weiszmam, E. Kwan, S. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, *et al.* (2002). "Systematic determination of patterns of gene expression during Drosophila embryogenesis." Genome Biol **3**(12): RESEARCH0088.
- Van Gelder, R. N., M. E. von Zastrow, A. Yool, W. C. Dement, J. D. Barchas und J. H. Eberwine (1990). "Amplified RNA synthesized from limited quantities of heterogeneous cDNA." Proc Natl Acad Sci U S A **87**(5): 1663-7.

- Weigmann, K., R. Klapper, T. Strasser, C. Rickert, G. Technau, H. Jäckle, W. Janning und C. Klämbt (2003). "FlyMove-a new way to look at development of *Drosophila*." Trends Genet **19**(6): 310-1.
- Weiss, J. B., T. Von Ohlen, D. M. Mellerick, G. Dressler, C. Q. Doe und M. P. Scott (1998). "Dorsoventral patterning in the *Drosophila* central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity." Genes Dev **12**(22): 3591-602.
- White, K. (2004). Systematic mapping of developmental genomic regulatory networks. 45th Annual *Drosophila* Research Conference, Washington.
- White, K., N. L. DeCelles und T. C. Enlow (1983). "Genetic and developmental analysis of the locus *vnd* in *Drosophila melanogaster*." Genetics **104**(3): 433-48.
- Yamamoto, Y., F. Girard, B. Bello, M. Affolter und W. J. Gehring (1997). "The cramped gene of *Drosophila* is a member of the Polycomb-group, and interacts with *mus209*, the gene encoding Proliferating Cell Nuclear Antigen." Development **124**(17): 3385-94.
- Yang, L. B., K. Lindholm, R. Yan, M. Citron, W. Xia, X. L. Yang, T. Beach, L. Sue, P. Wong, D. Price, *et al.* (2003). "Elevated beta-secretase expression and enzymatic activity detected in sporadic Alzheimer disease." Nat Med **9**(1): 3-4.
- Yasojima, K., E. G. McGeer und P. L. McGeer (2001). "High stability of mRNAs postmortem and protocols for their assessment by RT-PCR." Brain Res Brain Res Protoc **8**(3): 212-8.
- Zalokar, M. (1976). "Autoradiographic study of protein and RNA formation during early development of *Drosophila* eggs." Dev Biol **49**(2): 425-37.
- Zink, B., Y. Engstrom, W. J. Gehring und R. Paro (1991). "Direct interaction of the Polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of *Drosophila melanogaster*." Embo J **10**(1): 153-62.

## 7 Veröffentlichungen

Teile dieser Arbeit wurden veröffentlicht in:

Hild, M., Beckmann, B., Haas, S.A., Koch, B., Solovyev, V., Busold, C., Fellenberg, K., Boutros, M., Vingron, M., Sauer, F., Hoheisel, J.D. und Paro, R. (2003). An integrated gene annotation and transcriptional profiling towards the full gene content of the *Drosophila* genome. Genome Biology 5:R3.

Boutros, M., Kiger, A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Heidelberg Fly Array Consortium, Paro, R., Perrimon, N. (2004). Genome-Wide RNAi Analysis of Growth and Viability in *Drosophila* Cells. Science **303**(5659): 832-5

## 8 Abkürzungen

### 8.1 Genetik

<i>abd-A</i>	<i>abdominal-A</i>
<i>Abd-B</i>	<i>Abdominal-B</i>
<i>Antp</i>	<i>Antennapedia</i>
CMM	Cellular Memory Module
<i>Dfd</i>	<i>Deformed</i>
fas3	fasciclin-3
Gal4	Hefegen zum Wachstum auf Galaktose
GFP	Green Fluorescent Protein
Hox	Homeo Box (Homöobox)
<i>hth</i>	<i>homothorax</i>
<i>Lac</i>	<i>Lachesin</i>
<i>pb</i>	<i>proboscipedia</i>
PcG	Polycomb-Gruppe
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PRE	Polycomb Response Element
trxG	trithorax-Gruppe
<i>trx</i>	<i>thrithorax</i>
<i>Ubx</i>	<i>Ultrabithorax</i>
<i>w</i>	<i>white</i>

**8.2 Sonstige**

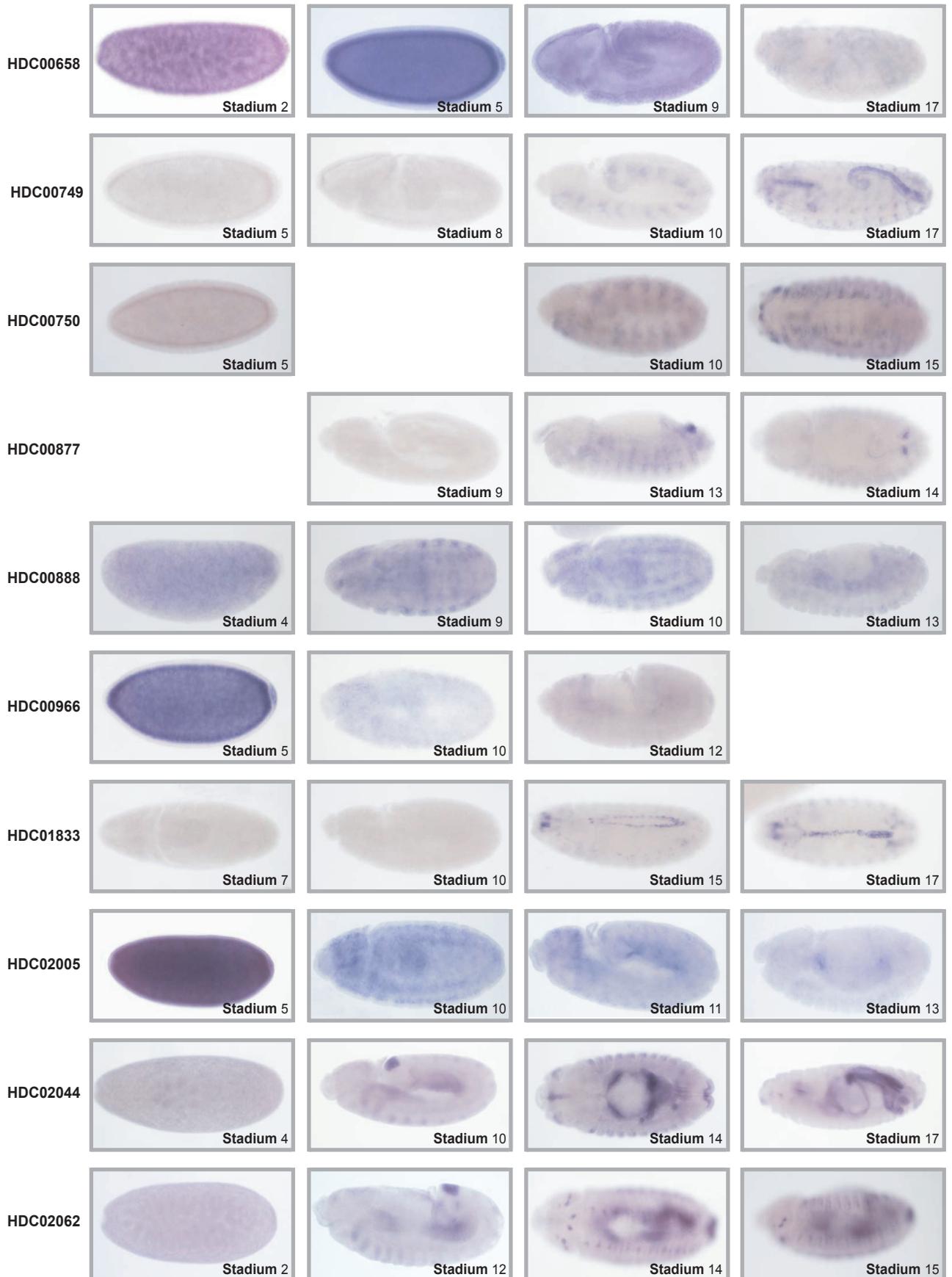
Amp	Ampicillin
AP	Alkalische Phosphatase
APS	Ammoniumpersulfat
ATP	Adenosin-5'-triphosphat
BCIP	5-Bromo-4-Chloro-3-Indolyl-phosphat
BSA	Albumin aus Rinderseren
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DTT	Dithiothreitol
EDTA	Ethylendiaminotetraessigsäure
EtBr	Ethidiumbromid
GS	Goat Serum
GST	Glutathion-S-Transferase
Hepes	4-(2-hydroxyethyl)-piperazin-1-ethansulfonsäure
IPTG	Isopropyl-thio- $\beta$ -galaktopyranosid
NBT	Nitro-Blue-Tetrazolium-Chlorid
OD	optische Dichte
PAA	Polyacrylamid
PBS	Phosphat gepufferte Salzlösung
PCR	Polymerase Kettenreaktion (Polymerase chain reaction)
RNA	Ribonukleinsäure
RT	Raumtemperatur
RT-PCR	Reverse Transkription-Polymerase Kettenreaktion
TBS	Tris gepufferte Salzlösung
TBST	Tris gepufferte Salzlösung mit Tweenzusatz
Tris	Tris-(hydroxymethyl)-aminomethan
ün	über Nacht
Upm	Umdrehungen pro Minute

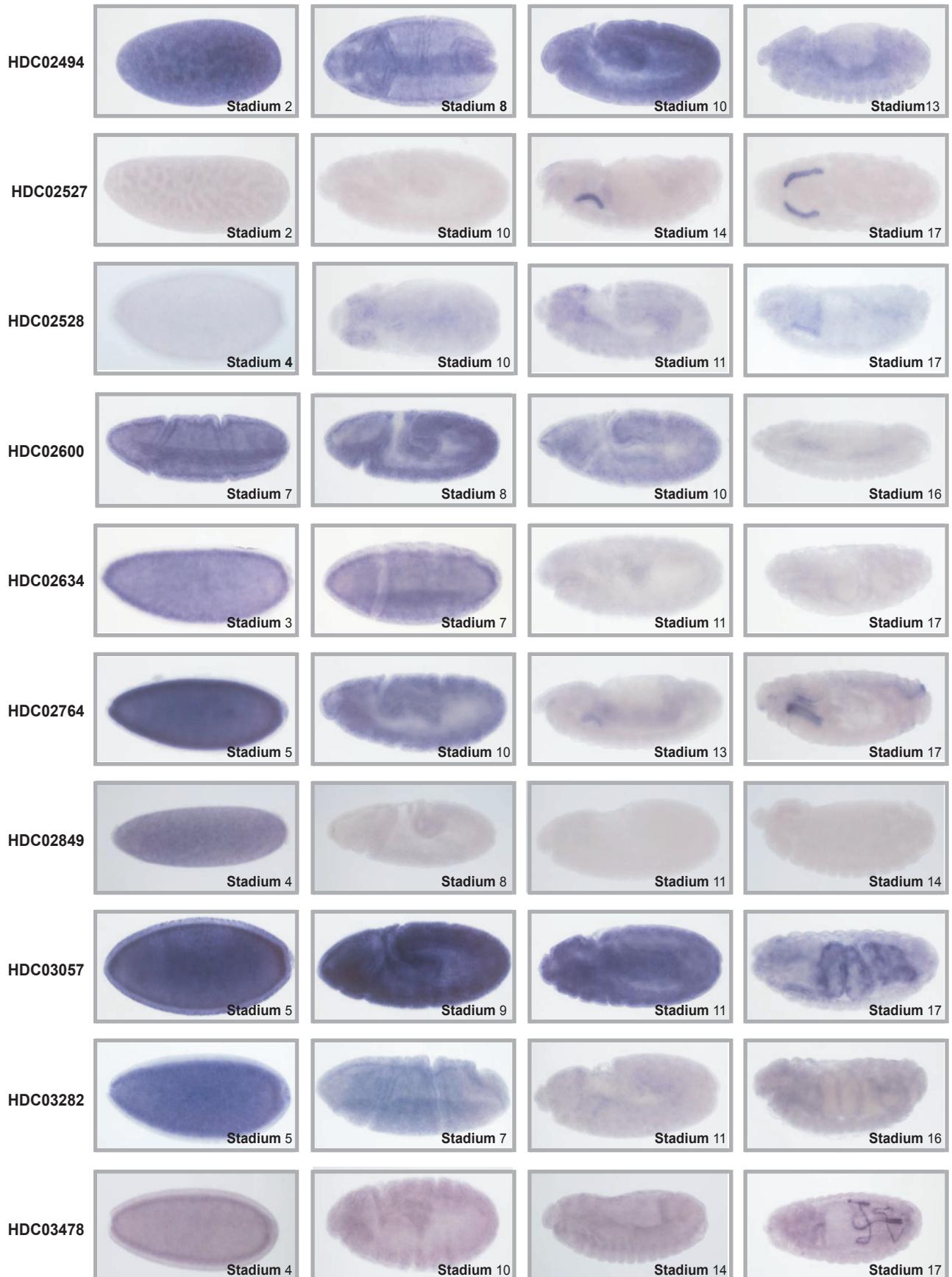
## 9 Anhang

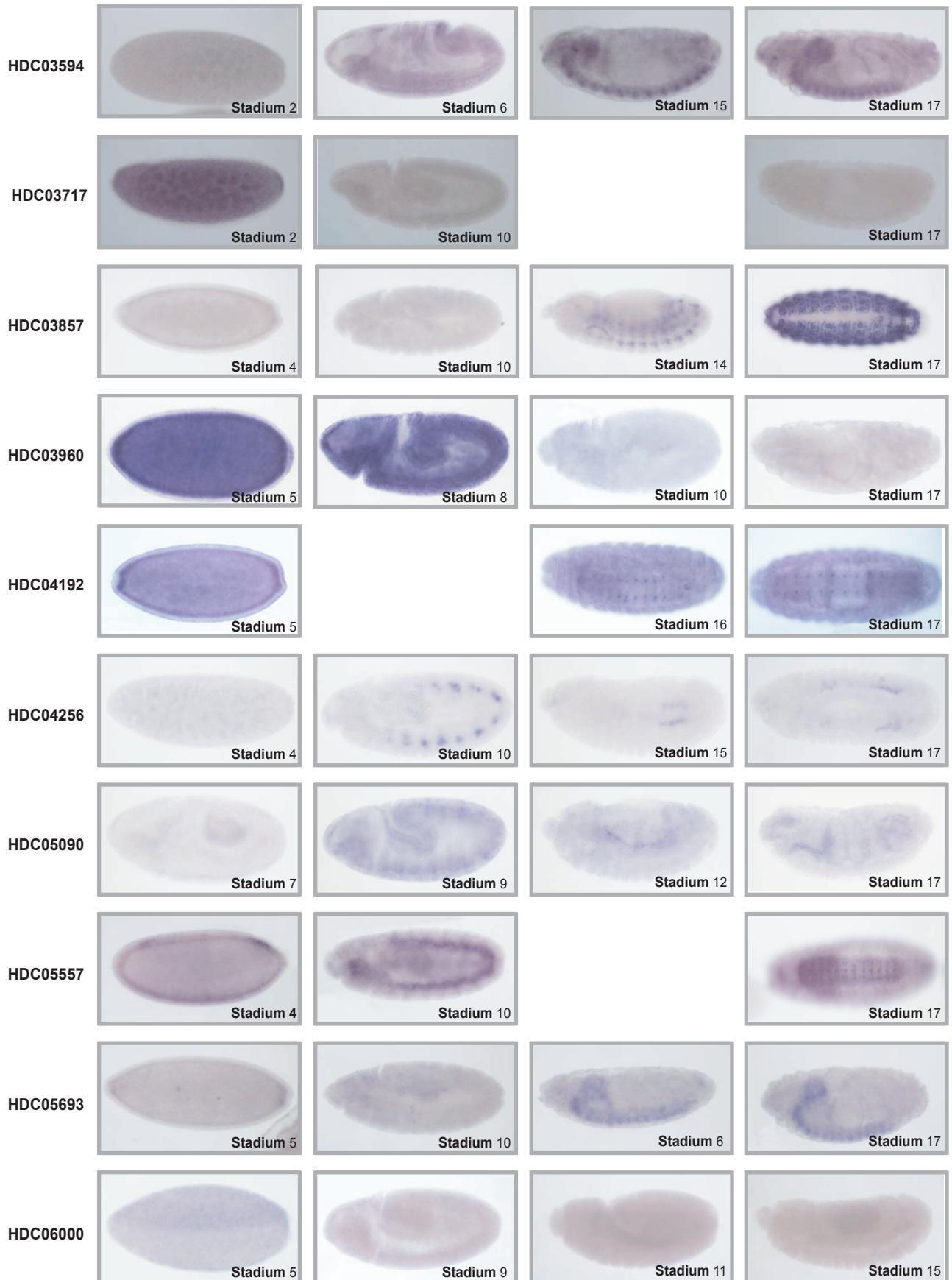
**Abbildungen der *in-situ* Hybridisierungen.** Gene, die über *in-situ* Hybridisierungen nachgewiesen werden konnten, wurden zu verschiedenen Zeitpunkten der Embryogenese fotografiert. Für jedes Gen ist die Expression zu verschiedenen, angegebenen Zeitpunkten während der Embryogenese gezeigt.

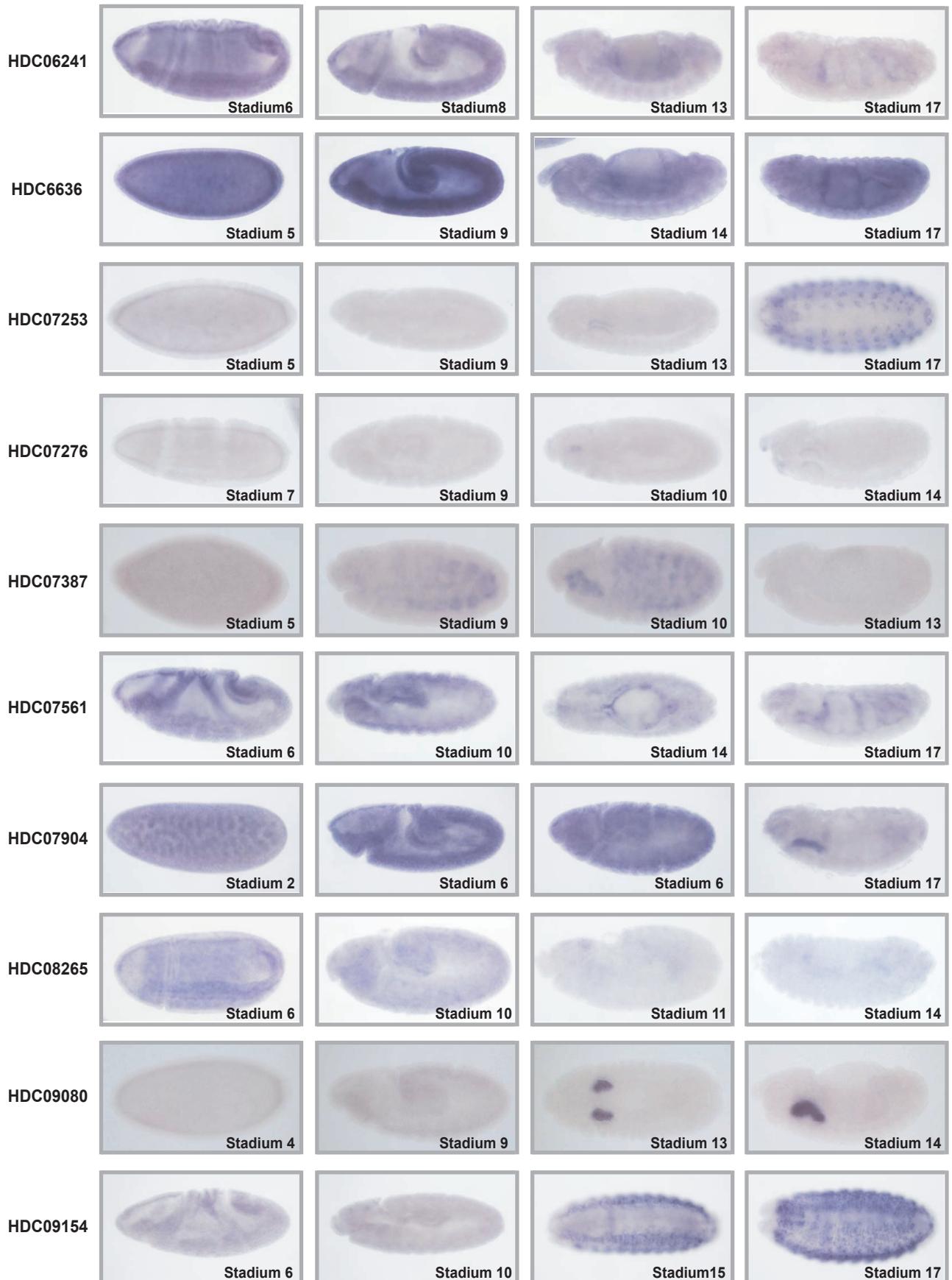
**Potentielle Polycomb Zielgene.** Eine Liste der Gene, die in der Polycomb Mutante mindestens 2fach hochreguliert (Tabelle 1) bzw. herunterreguliert (Tabelle 2) sind, die eindeutige Gen-ID, der Gennamen, die GO-Beschreibung, vorhandene Domänen, und die Regulation im Vergleich zur Kontrolle sind dargestellt.

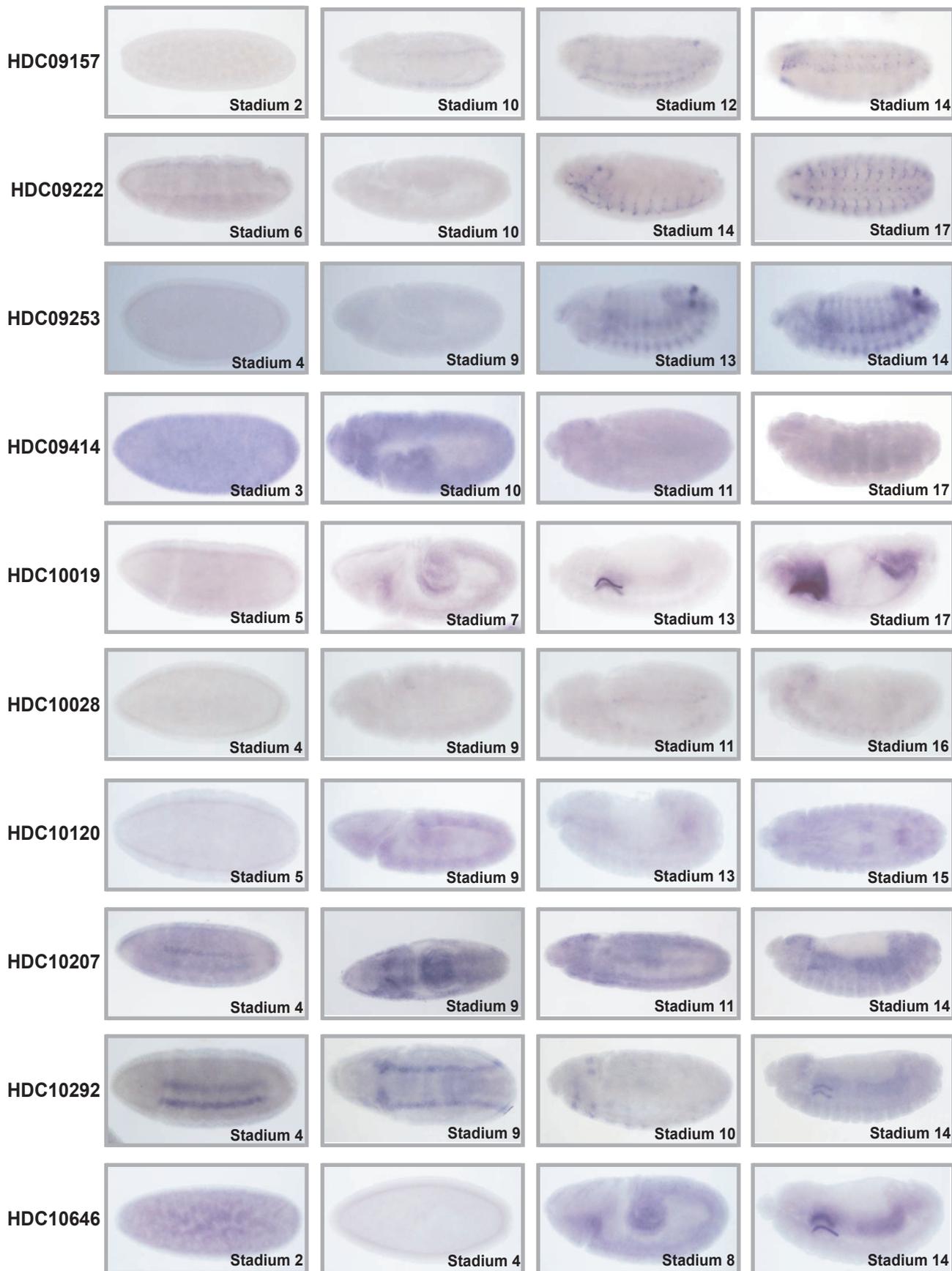
**Die HeidelbergFlyAnnotation.** Alle vorhergesagten Gene, die Primersequenzen für alle im Microarrayprojekt verwendeten Primer, die eindeutigen Amplikon-Nummern (HFA) für jedes Fragment, vorhandene CG-Nummern der Version 2 und 3.1 sowie die Bestellnummern der Primer und der Name der HeidelbergCollection (HDC), die Nukleotid- und Peptidlängen der vorhergesagten Gene, Informationen über den Erfolg in den PCRs, auf dem Microarray (M-CHiPS), den *in-situs* und den RT-PCRs, die Platteninformation der einzelnen Fragmente in den 96-well Platten und vorhandene Domänen sind der beigelegten CD zu entnehmen.

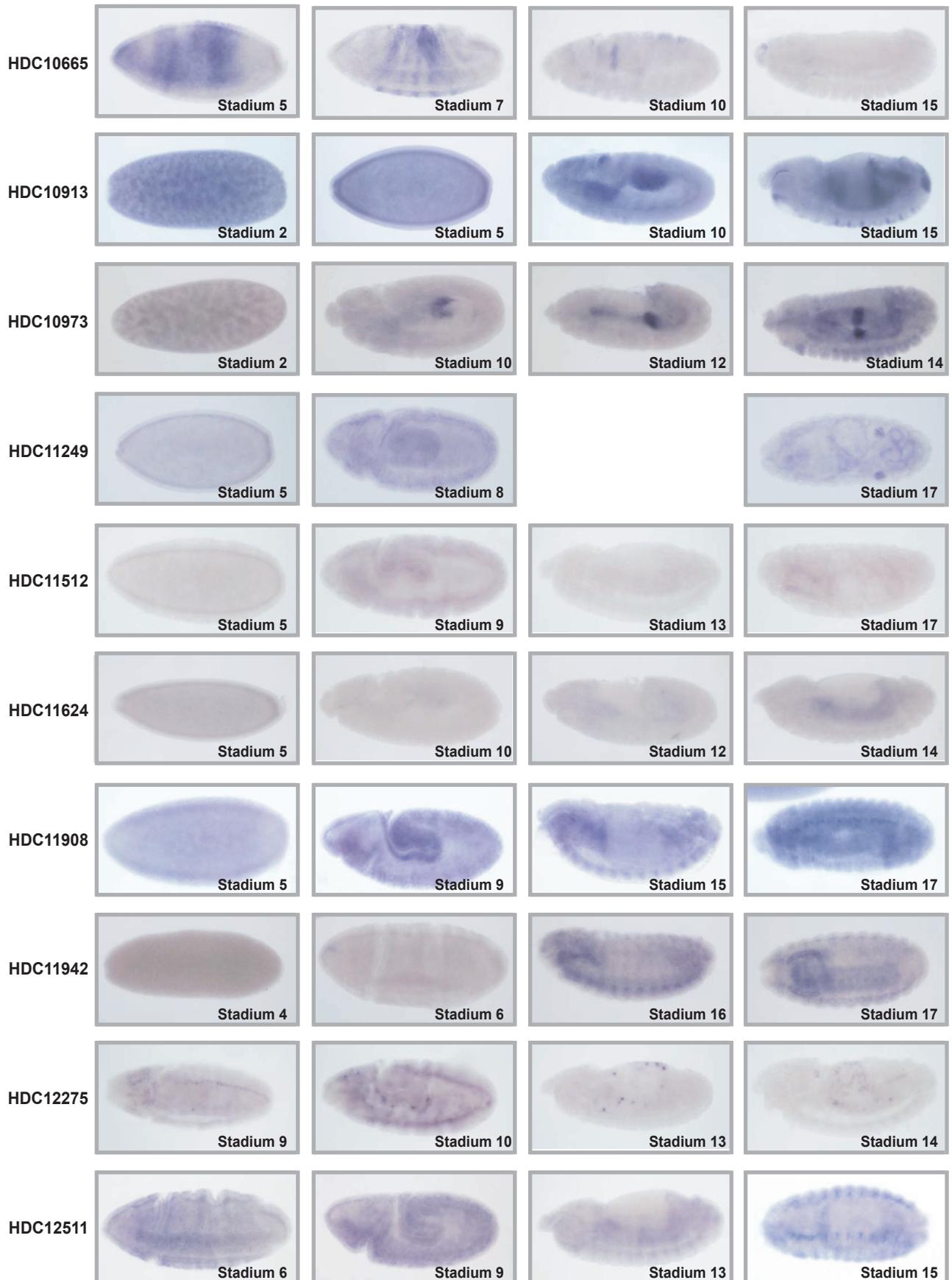


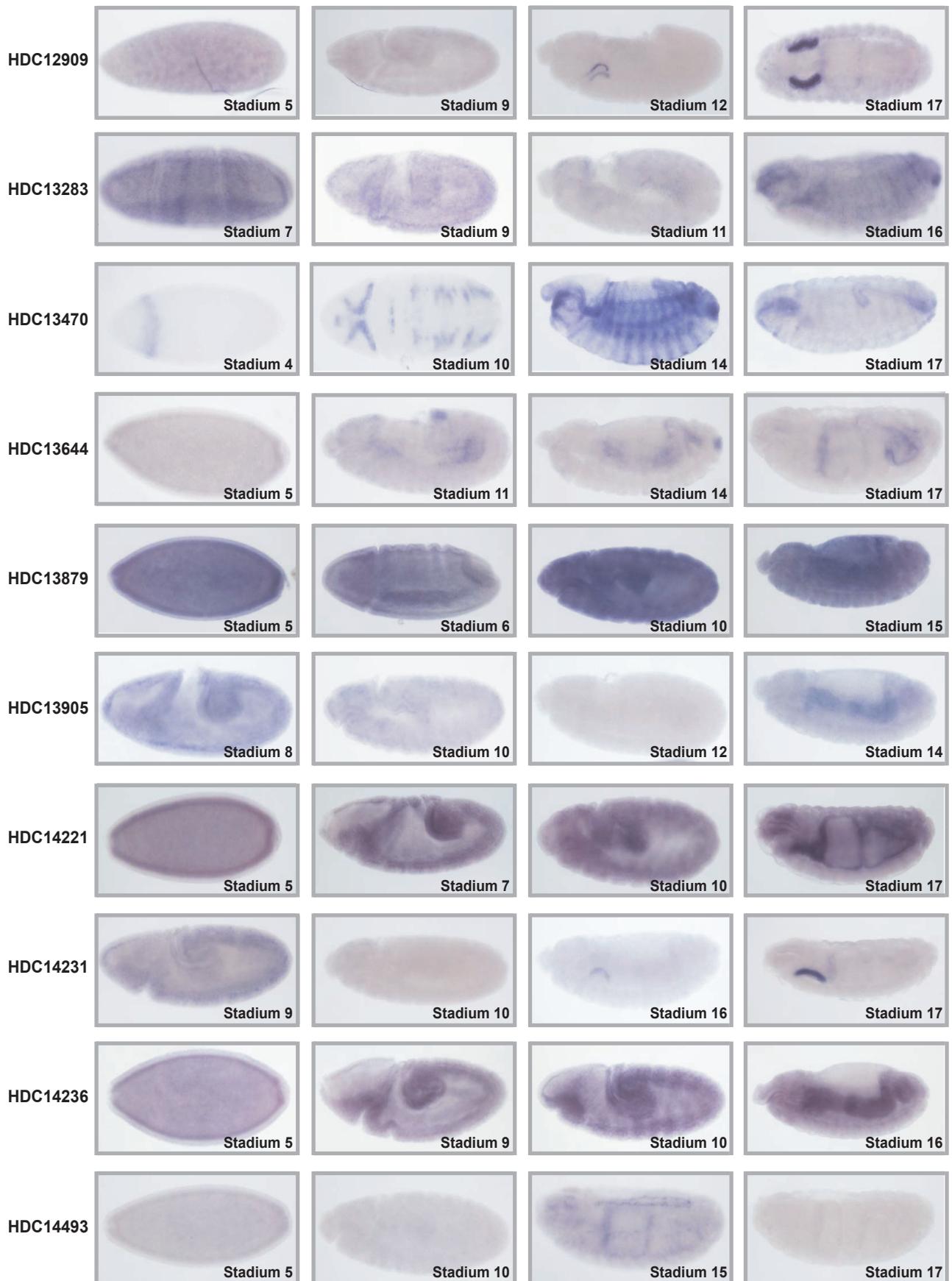


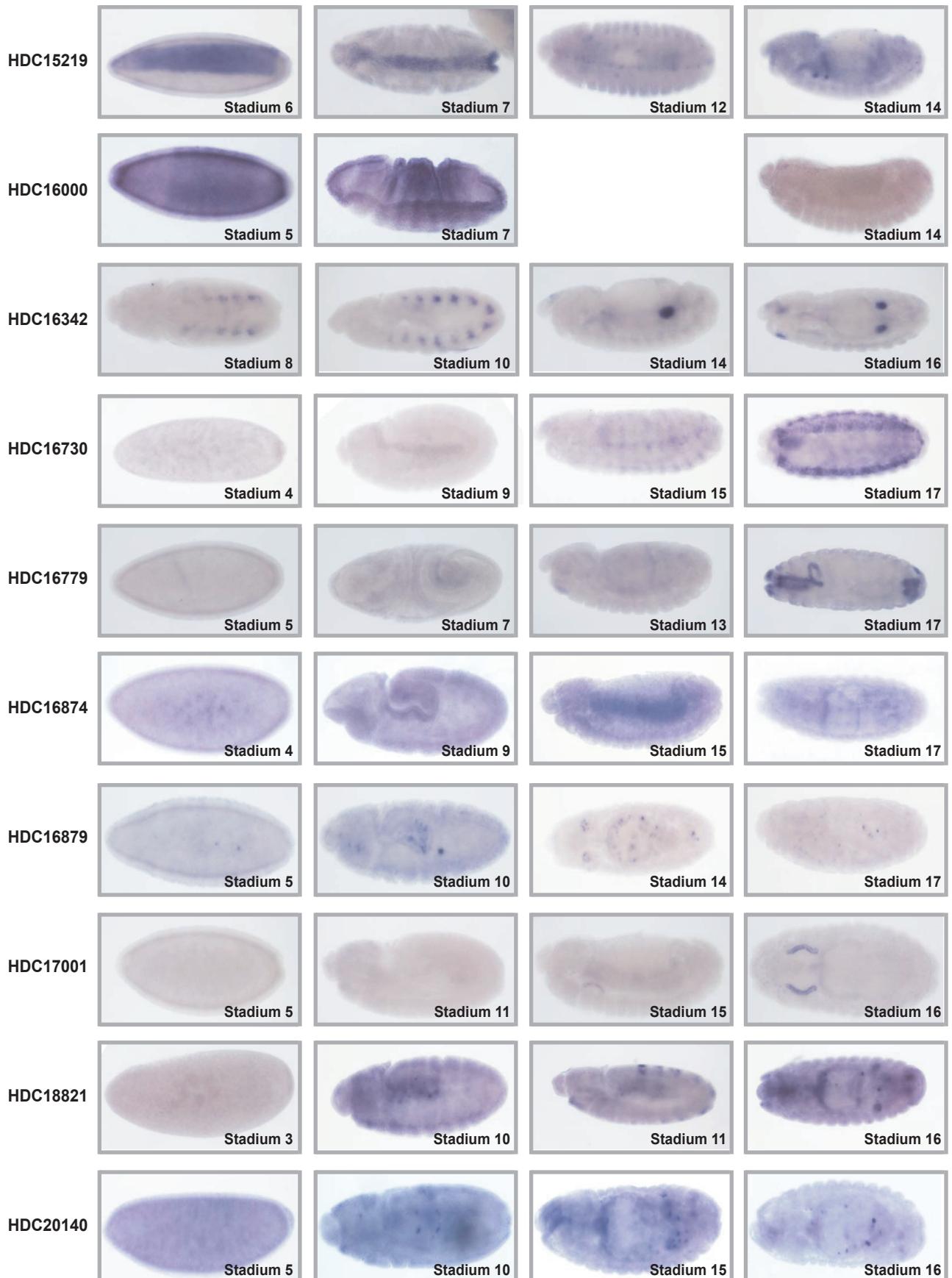












**Anhang Tabelle 1. In der Pc-Mutante hochregulierte Gene**

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA00011	21C	CG33127		Ser_protease_Try	2.85
HFA00068	-	unique			6.58
HFA00093	-	HDC00316			14.71
HFA00237	2L	HDC00756			2.97
HFA00391	22B	CG15361			3.11
HFA00464	25B	CG15629	oxidoreductase activity	Adh_short_C2	2.67
HFA00628	24D	CG3702			2.75
HFA00770	23F	CG8846	eukaryotic initiation factor 4E binding		3.65
HFA00794	24F	CG3048	Toll signaling pathway	Znf_TRAF	2.45
HFA00825	21A	CG2671	transforming growth factor beta receptor signaling pathway	Lethal2_giant	3.33
HFA01269	-	unique			19.23
HFA01274	-	unique			22.73
HFA01566	-	unique			4.59
HFA01585	-	HDC20376			12.66
HFA01838	39C	CG8677 CG31626		Chitin_bind_PerA	12.82
HFA01850	36C	CG6582	anti-apoptosis		2.82
HFA01862	35B	CG32954		Insect_adh_fam	3.85
HFA01878	34C	CG9031	small GTPase regulatory/interacting protein activity	LRR	3.70
HFA01914	34C	CG18507			2.46
HFA01983	35F	CG4455	protein disulfide isomerase activity		6.85
HFA02008	29A	CG8049	protein-tyrosine kinase activity	Tyr_pkinase	3.04
HFA02078	37B	CG10621	selenocysteine methyltransferase activity	S_methyl_trans	3.05
HFA02103	38B	CG10746			4.88
HFA02174	28D	CG12560		GCN5acetyltransf	4.05
HFA02202	29F	CG13102			2.48
HFA02318	26B	CG31640		FA58_C	2.68
HFA02379	29A	CG14275		NLS_BP	3.58
HFA02581	32A	CG17127			2.40
HFA02620	25D	CG11020	ion channel activity	ANK	3.22
HFA02775	30F	CG4747	3-hydroxyisobutyrate dehydrogenase activity	PWWP_domain	2.51
HFA02798	31D	CG4972			3.15
HFA02824	27F	CG5177	trehalose-phosphatase activity	Trehalose_PPase	16.39
HFA02971	33C	CG6579			2.83
HFA03099	39D	CG8666	integral to membrane	Transmem_4	7.25
HFA03107	39C	CG8677		Znf_PHD	2.84
HFA03171	39B	CG9253	ATP-dependent RNA helicase activity	DEAD	3.11
HFA03206	38F	CG9331	oxidoreductase activity, acting on CH-OH group of donors	2-Hacid_DH	2.65
HFA03211	38E	CG9336			5.52
HFA03213	38F	CG9338			4.90

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA03233	29F	CG9496	integral to membrane	Transmem_4	6.94
HFA03265	29F	CG9568			4.39
HFA03292	30B	CG3738	protein serine/threonine kinase activity	Cyc_dep_kin_rsub	5.43
HFA03337	31E	CG6176	structural constituent of cytoskeleton	Spc97_Spc98	2.47
HFA03350	36A	CG4475	imaginal disc growth factor activity	2S_Globulin	4.65
HFA03382	92C/27B	CG9258	sodium/potassium-exchanging ATPase activity	Na/K_ATPase_beta	3.07
HFA03413	31D	CG5371	DNA replication	Ribonucleo_red	2.86
HFA03466	29F	CG9564	trypsin activity	Chymotrypsin	5.38
HFA03473	26A	CG9046	eggshell formation (sensu Insecta)		2.75
HFA03702	40D	CR12628			2.87
HFA04001	59D	CG30413			2.83
HFA04116	59B	CG3800	nucleic acid binding	Znf_CCHC	3.94
HFA04187	59D	CG13545			6.37
HFA04317	60E	CG2803		UPF0131	3.58
HFA04369	60C	CG3419		PMP	2.91
HFA04435	59E	CG3941		Znf_C2H2	3.14
HFA04533	60D	CG9196			3.68
HFA04547	60E	CG9358		A10_OS-D	7.14
HFA04601	58F	CG5179	transcription elongation factor complex b	Tyr_pkinase	2.79
HFA04638	60A	CG11390		A10_OS-D	3.75
HFA04661	60A	CG11303	plasma membrane	Transmem_4	5.13
HFA04675	59A	CG3612	proton-transporting two-sector ATPase complex	ATPase_a/bN	5.10
HFA04698	59F	CG4533	heat shock protein activity	Crystallin_alpha	3.13
HFA04821	-	HDC20376			16.13
HFA04835	42A	CG12051	structural constituent of cytoskeleton	Actin_like	3.19
HFA04982	42D	CG18817			3.37
HFA05045	-	unique			15.38
HFA05221	47A	CG12052	RNA polymerase II transcription factor activity, axon guidance	Znf_C2H2	4.07
HFA05458	-	HDC09590			2.42
HFA05944	68D	CG8536	beta-N-acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase activity	Galactosyl_T_2	3.32
HFA05947	42E	CG10106	integral to membrane	Transmem_4	4.12
HFA06085	56E	CG11797		PBP_GOBP	2.58
HFA06105	46D	CG12129		KH_dom	2.75
HFA06117	42E	CG12172	serpin	Serpin	3.53
HFA06488	44D	CG14752			4.10
HFA06589	54B	CG30101			4.93
HFA06600	43D	CG1600		Adh_zn_family	2.61
HFA06609	43C	CG30502		Cyt_B5	4.03
HFA06661	54A2	CG30457		Gyr	16.67
HFA06673	55C	CG17524	glutathione transferase activity	GST_Cterm	3.51
HFA06756	54C	CG33130			3.29

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA06792	42D/E	CG18817	integral to membrane	Transmem_4	2.64
HFA06854	49D	CG3845	mRNA cap complex	eIF5C	3.22
HFA06866	53A	CG4311	hydroxymethylglutaryl-CoA synthase activity	HMG_CoA_synt	3.09
HFA06880	49F	CG4679			2.65
HFA06900	54C	CG4878	RNA binding,translation initiation factor activity	RNA_rec_mot	4.29
HFA06942	55C	CG30118			2.51
HFA06980	54E	CG6370	dolichyl-diphosphooligosaccharide-protein glycotransferase activity		2.78
HFA06985	53D	CG6426			2.86
HFA07018	47B	CG7220	ubiquitin conjugating enzyme activity	UBQ_conjugat	2.94
HFA07052	53B	CG7997	alpha-galactosidase activity	Glyco_hydro_27	4.03
HFA07058	45B	CG8029	hydrogen/potassium-exchanging ATPase complex		4.33
HFA07076	51E	CG8092		Znf_C2H2	5.41
HFA07085	51F	CG8157		Eggshell	5.10
HFA07310	48C	CG8986		PRO_rich	5.75
HFA07354	48F	CG8472	calcium ion binding	Recoverin	4.12
HFA07360	53D	CG5210	chitinase activity	Glyco_hydro_18	3.41
HFA07365	49B	CG8587	mitochondrion	EP450I	5.62
HFA07452	49A	CG12369	axonogenesis	Ig-like	2.87
HFA07474	49F	CG4696	actin binding	SM22_calponin	2.40
HFA07564	47E	CG9073	calcium ion binding	EF-hand	4.17
HFA07837	-	HDC07680			4.24
HFA07874	61C	HDC07759			4.69
HFA07881	61C	HDC07758			2.94
HFA08024	63B	CG32280			4.63
HFA08102	64B	CG32248		Gyr	10.99
HFA08126	-	HDC09590			2.83
HFA08159	63C	CG10812	defense/immunity protein activity		6.76
HFA08180	64B	CG1135	nucleolus	FHA	2.39
HFA08202	64A	CG32253		Brix	2.45
HFA08284	64B	CG1259	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	10.75
HFA08287	64B	CG12607		NLS_BP	17.86
HFA08313	64C	CG13704			6.94
HFA08314	64C	CG13705		YLP_motif	5.21
HFA08340	62D	CG32302		Chitin_bind_PerA	2.49
HFA08341	62D	CG13806		Chitin_bind_PerA	2.62
HFA08476	64B	CG15022		PRO_rich	7.19
HFA08488	63B	CG32486		Chitin_bind_PerA	2.56
HFA08586	61A	CG7047		Arrestin	2.82
HFA08649	64B	CG14996	actin binding	SM22_calponin	5.15
HFA08657	63C	CG10810	antifungal peptide activity	Gamma-thionin	4.81

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA08715	61D	CG1007	negative regulation of transcription from Pol II promoter, protein binding	HLH_basic	2.68
HFA08723	61A/B	CG18321		PTN_MK	4.93
HFA08902	66A	CG7892	protein serine/threonine kinase activity	Prot_kinase	3.42
HFA08928	-	HDC02062			2.40
HFA09096	-	HDC09590			2.77
HFA09115	68C	HDC09623			2.79
HFA09116	68C	HDC09626			3.12
HFA09253	70A	HDC09973			2.68
HFA09328	-	unique			6.62
HFA09484	74B	CG13731		PRO_rich	10.87
HFA09608	-	unique			2.77
HFA09613	76B	CG32214			2.88
HFA09615	76B	CG32213			3.50
HFA09640	76D	HDC11058			2.99
HFA09700	70A	CG10116			4.05
HFA09723	69A	CG10426	dephosphorylation	Exo_endo_phos	2.56
HFA09727	65A	CG32405		Znf_ring	5.35
HFA09891	76B	CG12519		NOR1_receptor	6.58
HFA09913	73A	CG13033		YLP_motif	4.46
HFA09928	72E	CG13047			4.98
HFA09949	72E	CG13066			5.35
HFA09950	72D	CG13067			7.63
HFA09951	72E	CG13068			7.87
HFA09952	72E	CG13069			9.43
HFA10010	71B	CG13463			16.67
HFA10033	66C	CG13674			3.53
HFA10036	21C	CG13678			5.92
HFA10098	76B	CG14095			14.71
HFA10099	76B	CG14096			3.44
HFA10212	64C	CG15212			8.47
HFA10213	64C	CG15213			8.40
HFA10302	67B	CG18178			5.21
HFA10311	76B	CG18294			3.51
HFA10315	67F	CG18349		Insect_cuticle	3.48
HFA10334	71B	CG18649		Granin	2.79
HFA10340	64A	CG32400	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	4.02
HFA10341	65A	CG18777			4.31
HFA10343	65A	CG32400		Insect_cuticle	3.02
HFA10344	65A	CG18777		Insect_cuticle	3.04
HFA10345	65A	CG18779		Insect_cuticle	10.53

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA10387	75E	CG3902	short-branched-chain-acyl-CoA dehydrogenase activity	Acyl-CoA_dh_C	3.62
HFA10407	73A	CG4229		YLP_motif	3.65
HFA10517	75A	CG5567	4-nitrophenylphosphatase activity	Hydrolase	2.39
HFA10549	72C	CG5830		NIF	5.78
HFA10597	72A	CG6169		NLS_BP	6.58
HFA10632	66D	CG6416		PDZ	8.00
HFA10647	74B	CG6479			2.48
HFA10651	66D	CG6486	peroxisome targeting signal-2 receptor activity	WD40	4.29
HFA10710	66D	CG6776	glutathione transferase activity	GST_Cterm	2.44
HFA10905	73E/D	CG7724	oxidoreductase activity, acting on CH-OH group of donors	3Beta_HSD	4.05
HFA10932	67F	CG7941	structural constituent of larval cuticle	Insect_cuticle	3.13
HFA10961	70C	CG8100	extracellular	hemocyanin_N	3.86
HFA11007	65F	CG8600			5.95
HFA11025	65E	CG8634		Insect_cuticle	5.85
HFA11027	65E	CG8640	structural constituent of cuticle (sensu Insecta)	Insect_cuticle	2.84
HFA11036	76C	CG32209		LDL_receptor_A	8.26
HFA11038	76C	CG8756		LDL_receptor_A	15.63
HFA11055	70E	CG9238	protein phosphatase 1 binding	CBM_21	2.67
HFA11073	76B	CG9449	acid phosphatase activity	HisAc_phsphtse	2.92
HFA11186	88E	CG6603	co-chaperone activity	Hsp70	3.83
HFA11210	65A	CG6956	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	3.73
HFA11211	65A	CG6955	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	2.80
HFA11215	65A	CG10530	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	4.35
HFA11226	70E/F	CG4994	carrier activity	Mitoch_carrier	2.51
HFA11289	64C	CG4659	7S RNA binding	SRP54	3.12
HFA11335	68C	CG7533			2.70
HFA11390	75C	CG4319	apoptosis activator activity		2.77
HFA11392	68B	CG7590			7.52
HFA11416	77C	HDC11156			4.55
HFA11417	77c	HDC11158			2.43
HFA11567	-	unique			5.10
HFA11593	80A	CG11100			7.75
HFA11694	79F	CG13239			5.46
HFA11729	84F	CG14568			6.94
HFA11730	78F/79A	CG14569			2.56
HFA11734	78F	CG14573			5.43
HFA11846	78C	CG7658	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	7.81
HFA11869	78D	CG7540	plasma membrane	Myelin_PLP	2.40
HFA11917	80A-B	CG17947	actin binding	Alpha_catenin	2.45

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA12128	2B	CG1151			4.52
HFA12172	83D	CG31559		YLP_motif	3.14
HFA12174	79D	CG1148			4.57
HFA12183	82B	CG11739	tricarboxylate carrier activity	Mtc	3.66
HFA12229	82A	CG14643		DUF243	8.93
HFA12242	82D	CG14656			2.52
HFA12252	82F	CG14666	protein translocase activity	Tim17_Tim22	2.67
HFA12292	82E	CG2016		Odorant_binding	3.08
HFA12314	83A	CG2663	tocopherol binding	RetBind/tocTrans	3.69
HFA12371	83B	CG1088	hydrogen-transporting ATPase V1 domain	ATPsynt_Esub	2.95
HFA12516	7F	HDC12113			2.60
HFA12541	84D	CG14598			2.48
HFA12568	83E	CG15188			4.83
HFA12587	69E	CG1252	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	24.39
HFA12588	84A	CG1327	structural constituent of larval cuticle (sensu Insecta)	Insect_cuticle	2.98
HFA12596	70A	CG1019	nucleus	LIM	5.03
HFA12749	85E	HDC12717			5.13
HFA12992	87F	HDC13470			6.76
HFA13078	-	HDC20376			14.93
HFA13273	90D	CG31243	RNA binding	RNA_rec_mot	3.82
HFA13276	-	HDC04773			13.33
HFA13305	-	unique			3.01
HFA13334	91E	HDC14405			3.76
HFA13400	-	unique			14.29
HFA13447	-	HDC14713			6.25
HFA13562	94B	CG13852		Mob1_phocein	2.71
HFA13687	96A	HDC15447			2.79
HFA14123	97D	CG6378	extracellular matrix		2.39
HFA14324	100D	CG11525	DNA repair, cyclin-dependent protein kinase regulator activity	Cyclin_N	5.59
HFA14350	1B	HDC13367			2.49
HFA14387	85B	CG31349			4.12
HFA14389	96B	CG11786		P_rich_extensn	6.13
HFA14441	96B	CG33095			4.17
HFA14503	90D	CG31243	guanylate cyclase activity	PDZ	2.72
HFA14512	85E	CG12418			2.75
HFA14637	96B	CG31357	serine-type endopeptidase activity	Chymotrypsin	2.96
HFA14738	97C	CG14243		DUF243	6.29
HFA14748	97D	CG14253			2.65
HFA14753	97E	CG14258		Odorant_binding	4.33
HFA14770	91D	CG14290		UPF0041	3.14
HFA14806	90D	CG14327		Eggshell	3.65

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA15124	87E	CG15887			3.10
HFA15147	85D	CG33188		Znf_A20	2.67
HFA15257	87B	CG17738			4.03
HFA15430	99F	CG2229	serine-type endopeptidase activity	Chymotrypsin	4.29
HFA15545	87B	CG4115		Lectin_C	23.26
HFA15589	86C	CG4511	ATP binding	Thioredox_dom2	3.13
HFA15810	97C	CG5468		DUF243	5.43
HFA15815	97C	CG5480		DUF243	8.00
HFA15868	95F	CG5746			2.41
HFA16045	97C	CG6460		DUF243	2.43
HFA16071	86C	CG6584	methionine-S-oxide reductase activity	DUF25	2.68
HFA16120	88E	CG6803			3.09
HFA16330	84E	CG7878	ATP-dependent RNA helicase activity	KH_dom	2.45
HFA16364	87D	CG8031		DUF52	2.79
HFA16418	85D	CG8420			2.67
HFA16554	99E	CG9747	acyl-CoA delta11-desaturase activity	Desaturase	4.15
HFA16570	85B	CG9821			4.69
HFA16657	90B	CG7539	structural constituent of pupal cuticle (sensu Insecta)	Eggshell	8.70
HFA16672	86C	CG1469	iron ion homeostasis	Ferritin	3.51
HFA16722	86E	CG17216	protein kinase activity	Tyr_pkinase	5.29
HFA16781	96B	CG7748	oligosaccharyl transferase activity	Oligo_trans_STT3	2.42
HFA16831	88A	CG3180	transcription from Pol II promoter	RNA_pol_Rpb2_6	2.63
HFA16859	99C	CG31362		Glutaredoxin	3.42
HFA16891	89E	CG10388	specific RNA polymerase II transcription factor activity regulation of transcription, DNA-dependent	Homeobox	3.51
HFA16896	87C	CG17369	hydrogen-transporting ATPase V1 domain	ATPase_a/bcentre	6.54
HFA16897	89E	CG10325	specific RNA polymerase II transcription factor activity, specification of segmental identity, abdomen,	Homeobox	3.28
HFA16944	86C	CG6544			3.57
HFA16959	89A	CG9734	binding	Globin	4.48
HFA16967	86B	CG17117	specific RNA polymerase II transcription factor activity	Homeobox	2.43
HFA16999	93D	CG2050	DNA binding, protein binding, RNA binding	RNA_rec_mot	4.52
HFA17061	61F	CG18525	serpin	Serpin	3.30
HFA17113	102B	HDC16691			2.39
HFA17114	102B	HDC16692			2.82
HFA17127	102A	CG1651	structural constituent of cytoskeleton	ANK	2.89
HFA17167	56E	CG1836	damaged DNA binding	UBA_domain	3.92
HFA17214	102F	CG18069	learning and/or memory	Prot_kinase	3.85
HFA17308	4B	CG3665	fasciculation of neuron	FnIII_subd	2.70
HFA17403	-	unique			7.87
HFA17481	6E	CG32733			3.73

ID	Zytologie	Gen-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA17705	3A	CG32797			3.86
HFA17897	1B	CG13362		PRO_rich	6.41
HFA18242	4A	CG2934	hydrogen-transporting ATPase V0 domain	ATPSynt_AC39sub	2.46
HFA18246	9B	CG2961		PRO_rich	2.45
HFA18423	8C	CG7267			3.80
HFA18476	1E	CG14629			3.11
HFA18565	-	unique			2.66
HFA18584	1F	CG11415	integral to membrane	Transmem_4	4.59
HFA18660	4A	CG4944	actin binding	Thymosin_b4	3.05
HFA19096	X	HDC19048			2.72
HFA19246	17A	HDC19530			2.39
HFA19546	18A	CG14191		Prion	3.61
HFA19671	13F	CG15602			3.02
HFA19686	11B	CG1572		Marvel	6.76
HFA19698	11B	CG15731		YLP_motif	3.44
HFA19742	11B	CG15927			2.99
HFA19932	15E	CG5070			3.34
HFA20136	13E	CG32581		Znf_ring	3.28
HFA20202	15A	CG9673	serine-type endopeptidase activity	Chymotrypsin	4.24
HFA20216	14B	CG9914		3HCDH_N	3.04
HFA20246	13F	CG8893	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity	GAP_dhydrogenase	2.65
HFA20253	10E	CG4147	heat shock protein activity	Hsp70	2.40
HFA20274	15A	CG9108	regulation of G-protein coupled receptor protein signaling pathway	Regl_Gprotein	2.99
HFA20298	13F	CG9038			6.33
HFA20332	11D	CG4396	RNA binding	Hud_Sxl_RNA	6.17
HFA20568	19E	CG1742 CG32510	glutathione transferase activity	MAPEG	3.05
HFA20623	20A	CG32499			7.14
HFA20923	42C	CG3265	microtubule associated complex	Calponin-like	7.35

**Anhang Tabelle 1 Microarrayanalyse der in der Pc-Mutante hochregulierten Gene.** 308 Gene zeigten nach stringenter Auswertung mit der Software „express yourself“ eine mindestens 2-fache Hochregulation in der Polycomb-Mutante im Vergleich zum Wildtyp. Die in der Heidelberg Annotation verwendete HFA-Nummer, die zytologische Kartierung, der CG-Name, vorhandene Gene Ontology-(GO) und Interpro- Beschreibungen sowie die auf dem Microarray beobachtete Regulation sind angegeben.

**Anhang Tabelle 2. In der Pc-Mutante herunterregulierte Gene**

ID	Gene-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA00160	CG31950		snRNP_Sm	-2.57
HFA00299	CG10880		NLS_BP	-2.50
HFA00358	CG13689			-3.20
HFA02003	CG9543	retrograde transport, Golgi to ER	Coatomer_E	-3.76
HFA02421	CG14940	calmodulin-dependent cyclic-nucleotide phosphodiesterase	PDEase	-2.50
HFA02463	CG15162			-2.69
HFA02808	CG5056			-3.03
HFA03073	CG7818	transcription factor activity	MT-A70	-2.41
HFA03190	CG9298		Sybindin	-2.50
HFA03421	CG5271	protein degradation tagging activity	Ubiquitin	-2.99
HFA03512	CG5855	oocyte axis determination	Cornichon	-2.93
HFA03738	CG2207	chromatin		-2.43
HFA04105	CG11301	transcription factor activity	CBFA_NFYB_domain	-2.41
HFA04458	CG4279	spliceosome complex	snRNP_Sm	-2.60
HFA04979	CG17266	mRNA splicing	CSA_PPase	-3.41
HFA06016	CG10933		SH3	-2.43
HFA06199	CG12859			-2.23
HFA06274	CG13018			-2.39
HFA06599	CG1599	SNAP receptor activity	Synaptobrevin	-2.41
HFA06939	CG5497	protein biosynthesis		-3.56
HFA07295	CG8888	oxidoreductase activity, acting on CH-OH group of donors	ADH_short	-2.91
HFA07554	CG8905	superoxide dismutase activity	SODismutase	-5.54
HFA07569	CG1818	uroporphyrinogen decarboxylase activity	Uro_decarbxylys	-3.16
HFA07662	CG11205	DNA repair protein	FAD_binding_7	-2.50
HFA08371	CG13901			-3.94
HFA08497	CG16971			-2.48
HFA09987	CG13379			-2.03
HFA09989	CG32202		NLS_BP	-2.06
HFA10474	CG5057	transcription coactivator activity		-2.35
HFA10634	CG32139	transcription factor activity DNA bending activity	HMG_12_box	-2.35
HFA10760	CG7011	COPII-coated vesicle		-2.19
HFA10835	CG7394		DnaJ_N	-3.25
HFA10948	CG8005	deoxyhypusine synthase activity	DS	-2.57
HFA11097	CG9710		Nuclear_move	-2.01
HFA11789	CG5969			-2.22
HFA12257	CG14671			-2.16
HFA14185	CG10091	glutathione transferase activity	GST_Cterm	-3.68
HFA14412	CG11858	peptidyl-prolyl cis-trans isomerase activity	Rotamase	-3.58
HFA14649	CG31111			-2.14
HFA14856	CG14463			-3.34

ID	Gene-ID	GO Beschreibung	InterPro Beschreibung	Regulation
HFA14988	CG14865			-2.66
HFA15128	CG15922			-2.50
HFA15442	CG31472		Pyridox_oxidase	-2.04
HFA15552	CG4217	transcription factor activity	HMG_12_box	-2.10
HFA15602	CG4570			-2.14
HFA16124	CG6830		DUF227	-2.14
HFA16279	CG7671		WD40	-2.10
HFA16499	CG9471	biliverdin reductase activity		-2.07
HFA16538	CG9688	structural constituent of ribosome	Ribosomal_S18	-2.71
HFA16572	CG9836		NifU_N	-2.10
HFA16802	CG6011	spliceosome complex	Prp18	-2.39
HFA16873	CG6476	chromatin binding	EF_GTPbind	-5.90
HFA16940	CG7425	chromosome organization and biogenesis	UBQ_conjugat	-4.17
HFA17020	CG7929		Ocnus	-4.44
HFA17282	CG3176			-2.55
HFA17749	CG10648		Mak16	-2.19
HFA17922	CG32736			-5.54
HFA18121	CG16892		WD40	-2.13
HFA18136	CG17776			-2.20
HFA18618	CG3038	UDP-galactose	Glyco_trans_31	-2.41
HFA18667	CG2227			-5.28
HFA18734	CG3796	transcription factor activity	HLH_basic	-2.97
HFA19360	CG10990		IF_eIF4G_MA3	-3.03
HFA19392	CG11164			-2.03
HFA19425	CG11757		Acyltransferase	-2.28
HFA19437	CG12177		I/U_NHdlase	-3.76
HFA19681	CG15645			-3.48
HFA19840	CG2200	cytoplasm	Peptidase_S51	-2.03
HFA20220	CG9921		HIG_1_N	-2.93
HFA20487	CG11748	odorant binding	PBP_GOBP	-2.19

**Anhang Tabelle 2 Microarrayanalyse der in der Pc-Mutante herunterregulierten Gene.** 74 Gene zeigten nach stringenter Auswertung mit der Software „express yourself“ eine mindestens 2-fache gulation in der Polycomb-Mutante im Vergleich zum Wildtyp. Die in der Heidelberg Annotation verwendete HFA-Nummer, die zytologische Kartierung, der CG-Name, vorhandene Gene Ontology- (GO) und Interpro- Beschreibungen sowie die auf dem Microarray beobachtete Regulation sind angegeben.