FUNCTIONAL GENOMIC ANALYSIS OF DEVELOPMENTAL CONTROL GENE ACTION IN THE EMBRYONIC NERVOUS SYSTEM OF *DROSOPHILA*

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Haiqiong Montalta-He
aus Guangzhou (CHINA)

Basel 2003

Zoologisches Institut der Universität Basel
Pharmazentrum
Klingelbergstrasse 50
CH-4056 Basel

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof Dr. Heinrich Reichert (Fakultätsverantwortlicher)

Prof Dr. Esther Stoeckli (Korreferent)

Basel, den 21. Oktober 2003

Dekan Prof. Dr. Marcel Tanner

CONTENTS

1. SUMMARY

In the past two decades, developmental genetic analysis of the molecular control elements involved in early embryonic brain patterning has uncovered the existence of structurally and functionally homologous genes that have comparable, and indeed interchangeable functions in vertebrates and invertebrates. The cephalic gap gene family *orthodenticle(otd)/Otx* is expressed in the anterior brain of *Drosophila* and mouse. These genes play an important role during the formation of the anterior brain since mutation of *otd/Otx2* causes the loss of entire rostral brain in both phyla. Reciprocal gene replacement experiments have demonstrated the functional equivalence of *otd* and *Otx* genes. The homeotic genes are expressed in a virtually co-linear anteroposterior pattern in the developing posterior brain of *Drosophila* and mouse, where they are required for the patterning of the region and the specification of segmental neuronal identity. These findings indicate the evolutionary conservation of cephalic gap gene and homeotic gene action in embryonic brain development and propose the conserved genetic network composed of genes controlled by these genes.

In order to gain more information about the molecular basis of the genetic network underlying the observation of evolutionary conservation of key developmental control gene action, it is interesting and important to investigate the downstream targets of these control genes. To this end, this thesis takes advantage of the sequenced genome of *Drosophila* and the availability of high-density oligonucleotide array techniques to identify downstream genes at a genome wide level:

As an initial part of this thesis, microarray analysis of differential gene expression after heat shock revealed substantial changes in gene expression level for known heat-shock genes and identified numerous heat shock-inducible genes. These results demonstrated that high-density oligonucleotide arrays are sensitive, efficient, and quantitative instruments for the analysis of large-scale gene expression in *Drosophila* embryos. Based on this, in two subsequent parts of this thesis, this functional genomic approach was used to probe for candidate target genes of *otd* and *labial*(*lab*).

In a first part, microarray experiments focused on the *lab* gene. High-density oligonucleotide arrays with probe sets representing 1,513 identified and sequenced genes were used to analyze differential gene expression following *lab* overexpression in *Drosophila* embryos. A number of novel candidate downstream target genes for *lab* were identified, suggesting that LAB differentially regulates a limited and distinct set of

embryonically expressed *Drosophila* genes. This provides preliminary information for further mechanism-orientated experiments.

In a second part, microarray experiments focused on *otd/Otx* genes. In order to understand the functional equivalence of the *Drosophila otd* gene and the vertebrate *Otx* gene and gain insights into potential downstream genes of *otd* gene in the fly, a first genome wide quantitative transcript imaging experiment was carried out. This experiment was designed to study differential gene expression in flies in which either the *Drosophila otd* gene or the human *Otx2* gene was overexpressed under the control of heat shock. These experiments indicated that 93 genes, approximately one third of the *otd*-regulated transcripts, also respond to overexpression of the human *Otx2* gene in *Drosophila*. We postulate that these transcripts are common downstream targets of the fly *otd* gene and the human *Otx2* gene in *Drosophila* which are likely to represent the molecular basis of the functional equivalence of *otd* and *Otx2* gene action in *Drosophila*.

A final part of the thesis was aimed at reducing false positive results of microarray experiments. For this, methods were developed using the magnetic cell sorting technique to isolate specific cell population from *Drosophila* embryos for specific expression profiling. These methods were then applied to identify new candidate downstream genes of the gene *glial cells missing* (*gcm*) which is a key regulator during gliogenesis. The *GAL4-UAS* system was used to direct expression of a transmembrane protein, mCD8-GFP, exclusively to the neuroectoderm of stage 11 embryos, which permitted a high rate of purification of viable cells from the neuroectoderm as assayed by both cellular and molecular methods. Based on the sorted neuroectodermal cells, differential gene expression was analyzed in wildtype embryos versus embryos in which *gcm* was misexpressed throughout the neuroectoderm. Follow-up validation studies of genes identified as differentially expressed by in situ **hybridization revealed a rate of confirmation for the sorted cell-based microarray experiments of more than 80%. This strongly contrasts to the high false positive rate revealed by microarray experiments based on wholemount embryos. Our results strongly suggest that reduction of cell heterogeneity through cell sorting techniques leads to a marked increase in the ability of microarrays to reveal differential gene expression in the developing nervous system.**

2. Introduction

2.1 The embryonic brain of *Drosophila*: cephalic gap genes and homeotic genes

The anterior brain of *Drosophila* is subdivided into protocerebrum, deutocerebrum and tritocerebrum which develop from the procephalic neuroectoderm (Younossi-Hartenstein et al., 1996). The head gap genes such as *tailless* (*tll*), *orthodenticle*(*otd*), *empty spiracles*(*ems*), *sloppy paired*(*slp*) and *buttonhead*(*btd*), which are expressed in overlapping regions, are involved in the patterning of this region (Cohen and Jurgens, 1990; Finkelstein and Perrimon, 1990; Grossniklaus et al., 1994; Younossi-Hartenstein et al., 1997). The posterior brain of *Drosophila* is subdivided into mandibular, maxillar and labial neuromeres which derive from the ventral neuroectoderm of the three gnathal cephalic segments (Younossi-Hartenstein et al., 1996). The homeotic genes as well as the gap genes, which act differently compared to their roles in the anterior brain, mediate the metamerization of this part of the brain (McGinnis and Krumlauf, 1992; Pankratz, 1993; Hirth et al., 1998).

2.1.1 Cephalic gap gene action

Two criteria specify the term of cephalic gap gene. First, these genes transmit maternal positional information to the zygotic segmentation gene hierarchy and second, mutation of any of these genes results in a gap-like phenotype affecting several adjacent segments in the anterior head, including defects in cuticles, sensory organs and expression of segmentation markers like *engrailed* (*en*) and *wingless* (*wg*) in the respective domains (Grossniklaus et al., 1994; Schmidt-Ott et al., 1994; Cohen et al., 2003). Classically, gap genes expressed in the trunk region tightly regulate each others' expression domains and show only little overlap (Pankratz, 1993). Conversely, most of the cephalic gap genes are expressed in largely overlapping domains and seem not to interact with each other, except the regulation by terminal gap genes: *otd* is regulated by the gap gene *huckebein*; *ems* and *btd* are under the control of *tll* (Cohen and Jurgens, 1990; Wimmer et al., 1995; Gao et al., 1996; Hartmann et al., 2001). Due to differences between the cephalic and the trunk gap genes, the so-called combinatorial model was proposed (Cohen and Jurgens, 1990). According to this model, the first step in segmentation is accomplished by the fact that the overlapping domains of head gap gene expression are exactly one segment out of phase at their posterior ends. Specification of segmental identity is achieved by the individual combination of active gap

genes within a given segment. The validity of the combinatorial model was challenged by the fact that ectopic expression of *btd* anterior and posterior to its normal expression domain did not interfere with normal metamerization of the cephalic segments. It was, rather, speculated that the extant of overlap between the blastodermal expression domains of *wingless* (*wg*) and *hedgehog* (*hh*) ultimately determined the number of anterior cephalic segments (Cohen and Jurgens, 1990; Finkelstein and Perrimon, 1990; Mohler, 1995; Wimmer et al., 1997). Nevertheless, it is conceivable that segmentation in the anterior cephalic region including the brain is achieved by interactions of segment polarity genes which are directly turned on by specific head gap genes (Gallitano-Mendel and Finkelstein, 1997; Hartmann and Reichert, 1998; Sprecher, 2003).

As a cephalic gap gene, *otd* encodes a homeodomain transcription factor that is required for brain development and segmental patterning in *Drosophila*. At early blastoderm stages, *otd* transcripts first appear in a broad circumferential stripe in the anterior region of the embryo. This region gives rise to several cephalic segments as judged from blastoderm fate map (Cohen and Jurgens, 1990; Finkelstein and Perrimon, 1990; Finkelstein et al., 1990). Subsequently, expression of *otd* is restricted to the procephalic region and to a second expression domain, the ventral midline. During neuroectoderm formation, procephalic *otd* expression covers most of the protocerebral and an adjacent part of the deutocerebral brain anlagen, and during subsequent brain regionalization neuronal *otd* expression occurs throughout most of the protocerebrum and adjacent deutocerebrum (Hirth et al., 1995; Younossi-Hartenstein et al., 1997). Mutation of *otd* leads to pattern perturbations and deletions in the cuticular structures, and the peripheral nervous system of this cephalic region (Finkelstein and Perrimon, 1990; Finkelstein et al., 1990). Moreover, in the *otd* null mutants, protocerebral anlage is missing due to defective neuroblast formation in these regions, and in the developing ventral nerve cord, specific midline neurons and glia are defected which results in deranged or missing commissures (Finkelstein et al., 1990; Klämbt and Goodman, 1991; Younossi-Hartenstein et al., 1997).

ems encodes a homeobox gene transcription factor. During embryogenesis, expression of *ems* is first observed as a broad circumferential stripe at the cellular blastoderm stage (Walldorf and Gehring, 1992). Later, the *ems* gene is expressed in the developing cephalic region and in

a metameric expression pattern in ectodermal and neural cell patches in all trunk segments (Walldorf and Gehring, 1992). Results from *ems* mutant analysis indicate that *ems* is necessary for regionalized neurogenesis in the anterior brain; in the mutant, the deutocerebrum and tritocerebrum are deleted (Younossi-Hartenstein et al., 1997; Hartmann et al., 2000). Furthermore, *ems* is necessary for correct axonal path-finding of specific interneurons in the ventral nerve cord (Hartmann et al., 2000).

2.1.2 Function of the homeotic genes in the embryonic CNS

The homeotic genes encode homeodomain transcription factors and specify the identity of segments along the anteroposterior axis in insects (Lewis, 1978; McGinnis and Krumlauf, 1992). In *Drosophila*, they are arranged along the chromosome in two gene clusters: *Antennapedia*(*ATN-C*) complex consisting of *labial* (*lab*), *proboscipedia* (*pd*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), and *Bithorax* (*BX-C*) complex which includes *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*abd-B*)(Lewis, 1978). It has been shown that the expression pattern of the eight homeotic genes in the developing *Drosophila* embryo is related to the relative position of the genes within the cluster, such that genes located at the 3' in the cluster are expressed earlier and more anteriorly in the embryo than genes located more 5' in the cluster. This is the so-called spatial and temporal colinearity (Duboule and Morata, 1994). Furthermore, there seems to be a functional hierarchy among *Hox* gene products in that more posteriorly expressed *Hox* genes are functionally dominant over more anteriorly expressed *Hox* genes which is termed "posterior prevalence"(Duboule and Morata, 1994).

In the embryonic CNS of *Drosophila*, *Hox* gene expression is excluded from the anterior region of the brain where *otd* and *ems* are expressed and function to establish the corresponding neuromeres. With one exception, homeotic genes in the posterior regions of the developing brain and in the ventral nerve cord are expressed in an anteroposterior order that agrees with the spatial colinearity. The homeotic gene that is expressed in the most defined anterior region in the embryonic brain is *lab*. *lab* is expressed in the posterior tritocerebrum. Posterior to *lab* expression, there are non-overlapping domains of *Dfd*, *Scr* and *Antp* expression in the mandibular, maxillary and labial neuromeres, respectively. Genes of the *BX-C* are expressed in the more posterior thoracic and abdominal neuromeres(Hirth et al., 1998).

Detailed analysis on loss-of-function mutations for *lab* and *Dfd* reveal severe defects in the embryonic brain. In the *lab* null mutant, cells are generated in the posterior tritocerebral domain and correctly located. Nevertheless, these mutant cells do not appear to differentiate into neurons and do not extend axons, nor are they contacted by axons from other parts of the brain as they normally are in the wild type situation. In contrast to the absence of neuronal cell fate in the *lab* mutant domain, the glial patterning remains unaffected. In conclusion, the expression of the homeotic *lab* gene is necessary for neuron, but not glia, to adopt their proper differentiated cell fate in the developing tritocerebrum. Similar phenotypes were observed in *Dfd* mutants in the corresponding mandibular neuromere, but not in other homeotic gene mutants.

Interestingly, under the control of CNS-specific *labial* regulatory element, it has been demonstrated that all other *Drosophila Hox* gene products, except *Abd-B*, are able to efficiently rescue *lab* mutant phenotype. It is also noted that there is a correlation between the rescue efficiency of the *Hox* proteins and the chromosomal arrangement of their gene loci which is in agreement with the "posterior prevalence" (Duboule and Morata, 1994). Thus, genes located more 3' have a higher rescue efficiency than those located more 5'. Taken together, this indicates that most *Hox* proteins are functionally equivalent in their capability to replace Lab in the specification of neuronal identity although they have diverged gene sequences. Therefore, differences of *Hox* gene action most probably rely on *cis*-regulatory elements but not on protein specificity (Hirth et al., 2001).

2.2 Evolutionary conservation of the brain development in Drosophila and mouse

It has been debated since decades whether the brains of deuterostomes such as vertebrates and protostome invertebrates are phylogenetically unrelated or have the same origin. In the last decade, a large amount of comparative molecular, genetic and developmental evidence has accumulated that strongly support the monophyletic origin of the bilaterian brain. Among these findings, a developmental genetic analysis of the molecular control elements involved in early embryonic brain patterning is uncovering the existence of structurally and functionally homologous genes that have comparable, and indeed interchangeable, functions in key aspects of brain development in invertebrates and vertebrates (Reichert and Simeone, 2001; Reichert, 2002). Here, I focus on experimental data from studies carried out on two gene families, the

cephalic gap genes and the homeotic genes that might provide evidence for the evolutionary conservation of the CNS development.

2.2.1 Functional equivalence of cephalic gap genes

The homologs of the *Drosophila* cephalic gap genes have been shown to be involved in embryonic brain development in vertebrates. *Otx1* and *Otx2*, the two vertebrate homologs of the *otd* gene, are involved in fundamental processes of anterior neuroepithelium patterning (Simeone, 1998; Acampora et al., 2001a; Boyl et al., 2001b). Mutants of *Otx2* are early embryonic lethal and lack the rostral neuroectoderm which gives rise to the forebrain, midbrain, and rostral hindbrain. Similarly, the homologs of *ems* in vertebrates, *Emx*1 and *Emx*2 are expressed embryonically in the presumptive cerebral cortex and have been shown to play a role in the establishment of the cerebral cortex(Cecchi, 2002; Shinozaki et al., 2002).

In addition to the extensive similarities in expression patterns and mutant phenotypes of the *otd*/*Otx* gene family, *in vivo* gene replacement experiments provide remarkable evidence for the conservation of *otd*/*Otx* functional properties. Human *Otx1* and *Otx2* genes were overexpressed in *Drosophila otd* mutants under heat-shock control. This rescued the mutant brain phenotype and other defects observed in the *Drosophila otd* mutants (Leuzinger et al., 1998; Nagao et al., 1998). Similarly, the *Drosophila otd* gene was introduced into the *Otx1* and *Otx2* locus in the corresponding murine mutant background and was able to rescue mostly the defects of *Otx1* mutant as well as that of *Otx2* mutant if provided with the *Otx2* regulatory control elements (Acampora et al., 1998; Acampora et al., 2001b).

To determine whether the murine homologues of *ems* are capable of restoring the brain phenotype of *ems* mutant flies, genetic rescue experiments involving ubiquitous overexpression of the mouse *Emx*2 gene were carried out in *Drosophila* and substantial restoration of brain morphology was observed (Hartmann et al., 2000). It will be interesting to carry out the reciprocal genetic rescue experiments in the mouse to see if and to what extent the *Drosophila ems* gene can rescue those defects in the embryonic murine brain(Reichert and Simeone, 2001).

2.2.2 Evolutionary conservation of homeotic genes in brain patterning

Hox gene complexes that are structurally and functionally homologous to the *HOM-C* genes of *Drosophila* have been identified in many vertebrate species (Gellon and McGinnis, 1998). In many cases, spatial and temporal colinearity also applies, meaning that the order of the *Hox* genes in their chromosomal arrangements correlates with their expression pattern along the anteroposterior body axis, especially in the developing hindbrain where *Hox* gene expression patterns are anteroposteriorly ordered along the neuraxis (Duboule and Morata, 1994; Ruddle et al., 1994; Capecchi, 1997; Reichert and Simeone, 1999).

Hoxa1 and *Hoxb1* are homologs of the *Drosophila labial* gene. They are activated in the early neural ectoderm, and by headfold stage their expression patterns have reached a sharp anterior boundary coinciding with the anterior rhombomere 4 (r4) border. In *Hoxa1*$^{-/-}$;*Hoxb1*$^{-/-}$ double loss-of-function mutants, r4 is formed but reduced in size and r4-specific markers fail to be activated indicating the presence of a territory between r3 and r5 with an unknown identity. *Hoxa1*$^{-/-}$; *Hoxb1*$^{-/-}$ double mutants also have a reduced number of facial motor neurons which appear to exit randomly from the neural tube without fasciculation. These results suggest that *Hoxa1* and *Hoxb1* act together in the specification of r4 neuronal identity and in the patterning of nerves during vertebrate hindbrain development. This is remarkably similar to the function of *lab* during embryonic brain development in *Drosophila* (Gavalas et al., 1998; Studer et al., 1998; Reichert, 2002).

2.2.3 Conserved genetic network and target genes

The identification and investigation of specific families of developmental control genes like cephalic gap genes and homeotic genes, which play central and evolutionarily conserved roles in patterning the embryoinc brain, have provided important information towards a comprehensive understanding of the molecular genetic networks involved in brain morphogenesis in animals as diverse as *Drosophila* and the mouse (Reichert, 2002).

The molecular nature of this extended conservation is still unclear. Nevertheless, based on current results, one can propose that key elements of the ancestral molecular genetic program that controlled the development of an urbilaterian brain are likely to be conserved. Thus, either these developmental control genes regulate a conserved set of downstream targets or

morphogenetic pathways, which will lead to the notion that brains of vertebrates and invertebrates share the same origin, or these targets are different but functionally equivalent (Simeone et al., 2002). One way to answer this question is to investigate the downstream targets of these developmental control genes which are all transcription factors. The rapidly progressing genome project and advanced microarray techniques facilitate these studies in a spectacular manner. Especially in *Drosophila*, where the full genome sequence is now available, it is already possible to combine extensive manipulative molecular genetic technology and large scale functional genomics with the long-term goal of deciphering the genetic network involved in brain development (Reichert and Simeone, 2001).

2.3 Gliogenesis in *Drosophila* embryonic CNS: *glial cells missing* (*gcm*) gene action

The formation of a functional nervous system requires the correct specification of a large number of different cell types. These cell types fall into two major categories, neurons and glial cells (Jones, 2001). Accordingly, an important issue in developmental neurobiology is to understand how this diversity is generated in the nervous system. *Drosophila* has proved to be an excellent genetic model to study the mechanisms involved in neurogliogenesis, and recently significant progress has been made in understanding the mechanisms underlying neuron-glia fate switch, symmetric-asymmetric division of multipotent precursors, and sublineage specification (Bossing et al., 1996; Schmidt et al., 1997; Schmid et al., 1999; Anderson, 2001).

In *Drosophila*, the gene *glial cell deficient/glial cells missing* (*gcm*) is the master regulator of glial cell fate determination. It encodes a transcription factor that is transiently expressed in glial precursors in the neuroectoderm (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996). In the *gcm* mutant, cells that normally develop into glia enter a neuronal differentiation pathway leading to a loss of glia and a gain of neurons. In contrast, misexpression of *gcm* in neural progenitors results in an increase of glial cells at the expense of neurons, (Akiyama-Oda et al., 1998; Bernardoni et al., 1998; Jones, 2001). The molecular mechanisms of *gcm* action in *Drosophila* are thought to be mediated through the regulation of *gcm* downstream target genes. However, until recently, molecular genetic analyses have identified only few genes as *gcm* targets that are involved in gliogenesis (Klämbt, 1993; Campbell et al., 1994; Klaes et al., 1994; Xiong and Montell, 1995; Granderath et al., 1999).

In order to identify additional *gcm* target genes, two genome-wide microarray experiments based on Affymetrix genechips or spotted cDNA arrays have been carried out recently using embryos in which *gcm* was misexpressed genetically throughout the entire neuroectoderm (Egger et al., 2002; Freeman et al., 2003). Both studies reported the identification a large number of differentially expressed candidate genes following *gcm* misexpression in embryos as compared to wild type-like control embryos. However, changes of gene expression could be validated by in situ hybridization or immunostaining for only a very limited number of these candidate genes (Egger et al., 2002; Freeman et al., 2003). Indeed, based on the low level of validation, an estimate of the number of false positive results in whole embryo microarray studies of this type has been given at 88% (Freeman et al., 2003). Clearly, such a high level of false positives results would hinder the further application of microarray technology to studies of neurogliogenesis in *Drosophila*. False positives may be due to upregulation of target gene expression outside the CNS as an indirect effect of ectopic Gcm within the CNS. Meanwhile, specific subtle changes of gene expression in the CNS in response to GCM misexpression could be 'diluted' by expression outside the CNS. Thus the use of homogeneous target tissue should substantially increase true positives and reduce false positives in future microarray expeiments.

2.4 High-density oligonucleotide arrays and their application

Since the molecular concept of the gene was established in the late 1960s, the single-gene approach has dominated molecular research and has been very successful. Changes in gene expression are associated with many important biological phenomena, including developmental processes, disease states, and adaptive responses to the environment. Over years, a variety of techniques have been developed to detect these changes. For example: differential display, subtractive hybridization and real time reverse transcription polymerase chain reaction (RT-PCR). Although these methods are effective and sensitive, there is a limit on the numbers of genes that can be analyzed simultaneously. With the rapid progress in the genome sequencing projects, it has become possible to take advantage of the sequenced genome to decipher biological questions from a global perspective that is essential for obtaining comprehensive pictures of cell function. Driven by all these needs, high-density microarrays were developed and have become more and more popular (Brown and Botstein,

1999). High density microarrays, for the first time, provide biologists with the tool to investigate simultaneously an almost unlimited number of genes from a given genome and allow parallel quantification of their expression levels. In a high-throughput manner, expression profiling using microarrays appears to be a powerful tool for correlating gene functions with DNA sequences as well (Schena et al., 1995).

The application of microarrays in expression profiling is based on two fundamental principles. First, for most genes, changes in expression are the direct result of alteration in the abundance of the cognate mRNA. Those biological questions involving posttranscriptional regulation are not amenable to microarray analysis. Second, only DNA strands possessing complementary sequences can hybridize to each other to form a stable, double-stranded molecule. Microarrays exploit this property through the immobilization of millions of single-strand copies of a gene as individual array elements on a solid support surface. This is then incubated with a mixture of labeled DNA molecules. Only the labeled molecules that represent the same gene as the immobilized DNA elements can form heteroduplexes. By measuring the amount of labels at the end of the hybridization, relative abundance levels for each gene can be determined (Deyholos and Galbraith, 2001).

Depending on the nature of the probes and the how probes are immobilized, there are two different types of microarrays available for expression profiling. These are spotted DNA microarrays and oligonucleotide arrays, also known as Affymetrix GeneChips. Normally, probes on the spotted arrays are DNA fragments of ~ 400-2000bp generated by PCR whereas those on the oligonucleotide array are oligonucleotide sets with the length of ~25bp representing the genes (Schena et al., 1995). While probes on the spotted arrays are immobilized on the solid surface by printing, oligonucleotide probes are produced *in situ* on the wafer using a combination of photolithography and combinatorial chemistry (Lipshutz et al., 1999). Since only oligonucleotide array were used in this thesis, details about spotted array will not be discussed here.

2.4.1 High-density oligonucleotide arrays

Oligonucleotide arrays for expression profiling are designed and synthesized based on sequence information alone, without the need for physical intermediates such as clones, PCR

products, cDNAs, etc. Using as little as 200 to 300 bases from the coding region or 3' untranslated region, independent 25-mer oligonucleotides are selected (non-overlapping if possible, or minimally overlapping) to serve as sensitive, unique, sequence-specific probes. According to a set of empirically derived rules, probe design is based on complementarities to the selected genes, uniqueness relative to family members and other genes, and an absence of high homology to other RNAs that may be highly abundant in the sample (for example, rRNAs, tRNAs, Alu-like sequences, housekeeping genes, repetitive sequences) (Lockhart et al., 1996). Each gene sequence is represented on the array by a set of 14-20 oligonucleotide (probes) perfectly matching reference sequences. The same set of probes, containing a single nucleotide mismatch in a central position, is also represented on the array. The mismatch probes act as specificity controls that allow the direct subtraction of both background and cross-hybridization signals, and allow discrimination between 'real' signals and those due to non-specific or semi-specific hybridization (Lipshutz et al., 1999). In the first generation of oligonucleotide arrays, all the probes for one specific gene are aligned next to each other whereas, in new oligonucleotide arrays, probes for the same genes are distributed randomly on the array. This is specially designed to control the position effect during hybridization.

Taken together, oligonucleotide arrays have several specific advantages compared to spotted cDNA arrays:

- They can be designed and made directly from sequence information without physical intermediates.
- Large numbers of probes are used to increase detection redundancy, meaning there are many 'detectors' per gene so that saturation of hybridization can be avoided.
- Shorter probes can be targeted to the most unique regions of genes, therefore reducing cross-hybridization so that closely related members of gene families can be discriminated.
- Involvement of semiconductor techniques and light directed oligonucleotide synthesis allows the construction of arrays with extremely high information content (Lipshutz et al., 1999). With the development of these techniques, it is perceived that one hundred million non-overlapping 30-mer probes presenting the whole human genome would fit on a 2x2 cm array.
- Because of how the arrays are manufactured, it is very easy to handle them and the reproducibility of hybridization using the same batch of array is high.

The disadvantages of oligonucleotide arrays of the type Affymetrix GeneChip are the high cost and the lack of flexibility inherent in the synthesis process. However, the overall cost for experiments using oligonucleotide arrays has been decreasing in the past years. Moreover, there may be a misperception regarding inflexibility since many custom arrays have been made and more will be designed on a regular base by Affymetrix. Nevertheless, an alternative means of oligonucleotide array synthesis using ink-jet deposition has been developed and appears to reduce costs and increase the flexibility of oligonucleotide arrays (Agilent Technologies, Palo Alto, CA). With this technology, the time from completion of the design to fabrication of the array can take less than 1 day. It permits the creation of arrays with more than 25,000 features (individual oligos up to 60-mer in length) on a 1x3 inch microscope slide (Deyholos and Galbraith, 2001).

In this thesis, three different Affymetrix Genechips were used for global gene expression profiling in *Drosophila*. (1) The genes represented on a first array (ROEZ003A; Affymetrix Inc.) corresponded to 1,513 sequenced and annotated *Drosophila* genes deposited in SWISS-PROT/TrEMBL database as of spring 1998 (Leemans et al., 2000). (2) The first full genome-Genechip available was custom-designed *Drosophila* Genechip (roDROMEGAa; Affymetrix Inc.). It contained 14,090 sequences representing 13,369 genes from the Release1 of the annotated *Drosophila* genome (Montalta-He et al., 2002). (3) Subsequently, a second full genome array, commercial DrosGenome1 (Affymetrix, cat# 900 335) was used. This array was also based on the Release 1.0 of the *Drosophila* genome. (Sequences were downloaded from the Flybase database on August 25th, 2000.) Sequences on the array represented more than 13,500 predicted transcripts as well as different control genes. The difference between the two full genome GeneChips is: Probes on roDROMEGAa were selected from the coding region of the genes. In contrast, for DrosoGenome1, probes are specially chosen from the 3' untranslated region for two reasons: firstly, sequences in the 3' UTR has been proved to be more gene-specific which will theoretically low the chance of unspecific and cross hybridization. Secondly, this complements very well to the 3' bias of target preparation that contributes to the accuracy of the microarray experiments.

2.4.2 Concerns about microarray experiments

On the one hand, as a very powerful and efficient tool to view changes in expression level of thousands of genes at a time, microarrays allow biologists to address old questions in new ways and to generate new hypotheses. On the other hand, not only are they very costly in terms of equipment, consumables and time as are other large-scale experiments, but there are also inherent biological factors that influence whether the resulting experiment is maximally informative. In most cases, careful attention to experimental design will ensure that good use is made of the available resources, obvious biases will be avoided and it will be possible to answer the primary questions of interest (Yang and Speed, 2002). Many aspects should be considered during the design for microarray experiment, for example, consideration about the biological question, choice of arrays, replicates used, ways of sampling and data analysis and interpretation. Here we mainly focus on two aspects: sample heterogeneity and independent confirmation.

2.4.2.1 Sample heterogeneity

The major problem that hinders the further application of microarrays is the low level of validation attained. One of the main reasons accounting for this drawback appears to be the complexity of the tissue when multicellular organisms are used for the microarray experiments. Consequently, biologically relevant changes in gene expression level may be very subtle so that small differences may be averaged out in the overall signal and missed. This is especially prominent when studying neural tissue because the intrinsic heterogeneity of the tissue samples used causes a signal-to-noise problem for the specific detection of gene expression in a given microarray experiment (Barlow and Lockhart, 2002; Griffin et al., 2003; Henry et al., 2003). One way to solve the roblem of tissue heterogeneity is to purify specific cell types from complex tissue such as a developing nervous system.

There have been several successful examples of microarray experiments based on purification of specific cell types. These include the application of Laser Captured Microdissection (LCM), Fluorescent Associated Cell Separation (FACS), single cell transcript profiling or mRNA-tagging (Bryant et al., 1999b; Mills et al., 2001; Roy et al., 2002; Luzzi et al., 2003; Tietjen et al., 2003). These all demonstrate that access to a homogeneous population of specific cell types facilitates the application of microarray analysis in developmental biology.

In this thesis, we for the first time applied the technique of magnetic cell separation (MACS) to isolate neuroectoderm cells from *Drosophila* embryos for microarray analysis of *gcm* action in neurogliogenesis. To this end, neuroectodermal cells were genetically labeled with a transmembrane fusion protein consisting of murine CD8 fused with GFP (mCD8-GFP); following cell dissociation, mCD8-positive neuroectodermal cells were isolated using anti-mCD8 microbeads. Validation studies of genes identified as differentially expressed by in situ hybridization revealed high rates of confirmation for the sorted cell-based microarray experiments (Montalta-He et al., 2003).

2.4.2.2 Independent confirmation of microarray data

Typically, microarray studies tend to provide the biologist with a list of genes with quantitative changes in expression of genes that span a large number of functional categories. As a biologist, these data are only of value when validated with independent *in vivo* follow-up experiments. Currently, several methods have been popular for the validation of microarray data. They are Northern blot, Western blot, real time RT-PCR, in situ RNA hybridization and antibody immunostaining. Among these, Northern blot, Western blot and RT-PCR are more quantitative and high-throughput than in situ hybridization and antibody immunostaining. But in situ hybridization and antibody immunostaining not only can confirm the quantitative changes qualitatively but also give biological information concerning the spatial and temporal expression pattern of the genes, which might lead directly to the function of the gene and help to exclude experimental artifacts (Barlow and Lockhart, 2002). Given that changes in gene expression measured by microarrays can be spatially ubiquitous or ectopic, it seems that the combination of quantitative methods combined with in situ hybridization or antibody immunostaing leads to much better validation results. However, the follow-up experiments are currently comparably low throughput hence the efficiency of microarray studies is reduced.

In the fly community, this drawback has been recognized. Efforts towards a global solution have been directed to use high-throughput RNA in situ hybridization to assemble a database of gene-expression patterns for embryonic development of *Drosophila* (Tomancak et al., 2002). Until the beginning of 2003, this database comprised embryonic expression patterns

for about one-sixth of all *Drosophila* genes. Considering the production rate, it should cover 70% of *Drosophila* genome in the end of 2003. All the data are freely available to the scientific community through interactive web pages: www.fruitfly.org/cgi-bin/ex/insitu.pl. This database will definitely facilitate the validation of microarray data and consequently the application of microarray in research (Montalta-He and Reichert, 2003).

2.4.3 Organization and public sharing of microarray data

Microarray expression studies generated and are still producing unprecedented amount of functional genomics data, which promise to provide key insights into gene function. However, how to organize and maximize the value of the data to the entire scientific community becomes an important but difficult issue (Brazma et al., 2001; Brazma et al., 2003).

Several factors account for the difficulty in sharing microarray data. Gene expression data are very complex in that they are meaningful only in the context of particular conditions under which they were generated. In addition, there are as many transcriptomes as there are cell types multiplied by conditions although there is only one genome. Because microarrays only measure the relative changes in gene expression depending on conditions that are rarely standardized, it is often hard to compare gene expression data. Finally, involvement of different microarray platforms, experimental designs and normalization makes comparison and integration of these data a problematic exercise. Thus, a standard for microarray data - Minimum Information About a Microarray Experiment (MIAME) - was proposed (Brazma et al., 2001). According to MIAME, the minimum information about a published microarray study should include a description of the following six sections: 1) Experimental design; 2) Array design: each array used and each element (spot, feature) on the array; 3) Samples: nature of the sample, extract preparation and target labeling; 4) Hybridizations: procedures and parameters; 5) Measurements: images and quantification; 6) Normalization controls: logics and values. Based on MIAME, which was proposed to serve both research biologists and software developers, and a data exchange format, Microarray Gene Expression Markup Language, two public databases have been developed aiming at microarray data organization and data sharing.

Gene Expression Omnibus (GEO) provides a flexible and open design that facilitates submission, storage and retrieval of data sets. These data include single and dual channel microarray-based experiments measuring mRNA, genomic DNA and protein abundance, as well as non-array techniques such as serial analysis of gene expression (SAGE), and mass spectrometry proteomic data. Platforms, Samples and Serials are the three central data entities of GEO. The GEO repository has been publicly accessible through the web at http://www.ncbi.nlm.nih.gov/geo since 2002. Data in this thesis from 2002 on are all submitted to GEO (Montalta-He et al., 2002; Montalta-He et al., 2003).

ArrayExpression, a new public database of microarray gene expression data at the EBI, is available since 2003(http://www.ebi.ac.uk/arrayexpress). It is designed to hold well-annotated data from all microarray platforms in a structured way. The infrastructure of ArrayExpression consists of the database itself, data submissions as well as the Expression Profiler online analysis tool. Three types of submission are accepted by ArrayExpression: arrays, experiments, and protocols, each of which is assigned an accession number. A curation team provides help on data submission and annotation. The database can be queried on parameters such as author, laboratory, organism, experiment or array types.

2.5 This Thesis

The identification and investigation of specific families of developmental control genes like cephalic gap genes and homeotic genes, which play central and evolutionarily conserved roles in patterning the embryoinc brain, have suggested that the nervous systems of vertebrates and invertebrates are evolutionarily related. In order to gain more information about the molecular basis of the genetic network underlying these observations of evolutionary conservation, it is very interesting and very important to investigate the downstream targets of these developmental control genes. To this end, in this thesis, we take advantage of the sequenced genome of *Drosophila* and the high-density oligonucleotide array techniques to address four issues at a genome while level:

Firstly, oligonucleotide arrays were used for quantitative transcript imaging of embryonically expressed genes under standard conditions and in response to heat shock. Analysis of differential gene expression after heat shock revealed substantial gene expression level

changes for known heat-shock genes and identified numerous heat shock-inducible genes. These results demonstrate that high-density oligonucleotide arrays are sensitive, efficient, and quantitative instruments for the analysis of large scale gene expression in *Drosophila* embryos (Leemans et al., 2000).

Second, to probe for downstream genes of the homeotic gene labial, high-density oligonucleotide arrays with probe sets representing 1,513 identified and sequenced genes were used to analyze differential gene expression following labial overexpression in *Drosophila* embryos. A number of novel candidate downstream target genes for labial were identified, suggesting that LAB differentially regulates a limited and distinct set of embryonically expressed *Drosophila* genes. This provides preliminary information for further mechanism -orientated experiments (Leemans et al., 2001)

Third, in order to understand the functional equivalence of the *Drosophila otd* gene and the vertebrate *Otx2* gene demonstrated in reciprocal rescue experiments(Acampora et al., 1998; Leuzinger et al., 1998)and to gain insight into potential downstream genes of *otd* gene in the fly, a genome wide quantitative transcript imaging experiment of otd/Otx2 target genes was carried out. Oligonucleotide arrays representing 13,400 annotated *Drosophila* genes were used to study differential gene expression in flies in which either the *Drosophila otd* gene or the human *Otx2* gene was over expressed under the control of heat shock. Our experiments indicate that 93 genes, approximately one third of the *otd*-regulated transcripts, also respond to overexpression of the human *Otx2* gene in *Drosophila*. We postulate that these transcripts are common downstream targets of the fly *otd* gene and the human *Otx2* gene in *Drosophila*. These common *otd/Otx2* downstream genes are likely to represent the molecular basis of the functional equivalence of *otd* and *Otx2* gene action in *Drosophila*(Montalta-He et al., 2002).

Finally, to develop methods for reduction of false positive results of microarray experiment as well as to identify new *gcm* downstream genes involved in gliogenesis during *Drosophila* embryogenesis, a combination of genetic labeling and magnetic cell sorting was used for isolating neuroectodermal cells from *Drosophila* embryos for microarray analysis of gcm action in neurogliogenesis. The *GAL4-UAS* system was used to direct expression of mCD8-GFP, a molecular label suitable for magnetic cell isolation, exclusively to the neuroectoderm

of stage 11 embryos, which permitted a high rate of purification of viable cells from the neuroectoderm as assayed by both cellular and molecular methods. Based on the sorted neuroectodermal cells, differential gene expression was analyzed in wild type embryos versus embryos in which *gcm* were misexpressed throughout the neuroectoderm. Follow-up validation studies of genes identified as differentially expressed by in situ hybridization revealed a rate of confirmation for the sorted cell based microarray experiments of more than 80%. This strongly contrasts to the predicted false positive rate of 88% based on wholemount embryos. Our results strongly suggest that reduction of cell heterogeneity through cell sorting techniques leads to a marked increase in the ability of microarrays to reveal differential gene expression in the developing nervous system(Montalta-He et al., 2003).

3. Quantitative transcript imaging in normal and heat shocked *Drosophila* embryos using high-density oligonucleotide arrays

Ronny Leemans, Boris Egger, Thomas Loop, Lars Kammermeier, Haiqiong He, Beate Hartmann, Ulrich Certa*, Frank Hirth and Heinrich Reichert

Institute of Zoology, University of Basel, Basel, Switzerland
*PRGT, Hoffmann-La Roche Ltd., Basel, Switzerland

Proc. Natl. Acad. Sci. U S A. (2000) **97(22)**: 12138-12143.

SUMMARY

Embryonic development in *Drosophila* is characterized by an early phase during which a cellular blastoderm is formed and gastrulation takes place, and by a later postgastrulation phase in which key morphogenetic processes such as segmentation and organogenesis occur. We have focused on this later phase in embryogenesis with the goal of obtaining a comprehensive analysis of the zygotic gene expression that occurs during development under normal and altered environmental conditions. For this, a novel functional genomic approach to embryogenesis has been developed which uses high-density oligonucleotide arrays (GeneChips®) for large scale detection and quantification of gene expression. These oligonucleotide arrays were used for quantitative transcript imaging of embryonically expressed genes under standard conditions and in response to heat shock. In embryos raised under standard conditions, transcripts were detected for 37% of the 1519 identified genes represented on the arrays, and highly reproducible quantification of gene expression was achieved in all cases. Analysis of differential gene expression following heat shock revealed substantial expression level changes for known heat shock genes and identified novel heat shock-inducible genes. These results demonstrate that high-density oligonucleotide arrays are sensitive, efficient and quantitative instruments for the analysis of large scale gene expression in *Drosophila* embryos.

INTRODUCTION

Recently the genome of the first multicellular eukaryote *C. elegans* has been completely elucidated (*Caenorhabditis elegans* Sequencing Consortium, 1998). Sequencing of the *Drosophila melanogaster* genome has also been carried out and currently the corresponding putative open reading frames are being defined (Adams et al., 2000). On the basis of this complete genomic information, it will now be important to determine the complex expression of all encoded genes and analyse physiological as well as pathological phenomena from a global genetic perspective. Large scale transcript analysis is made possible by DNA micro- or oligonucleotide arrays (Granjeaud et al., 1999; Lipshutz et al., 1999), both of which allow the

simultaneous monitoring of hundreds of mRNA expression profiles (Lockhart et al., 1996; Lashkari et al., 1997). In this study, we used *Drosophila* high-density oligonucleotide arrays to monitor the simultaneous expression of zygotically active genes during the later postgastrulation stages of embryonic development (Akam, 1987; Pankratz, 1993; Campos-Ortega and Hartenstein, 1997). We analysed the relative abundance levels of hundreds of embryonically expressed genes under normal physiological conditions and in response to heat shock (Nover and Scharf, 1997). In embryos raised under normal conditions, we obtained highly reproducible quantification for 563 expressed genes corresponding to different functional classes. Following a 36°C heat shock, we detected increases in expression levels for known heat shock genes and identified novel heat shock-inducible genes.

MATERIALS AND METHODS

Embryos.

Drosophila melanogaster Oregon R stocks were kept on standard cornmeal/yeast/agar medium at 25°C. Embryos were collected overnight on grapejuice plates for 12 hours and were kept for further 5 hours at 25°C before RNA isolation. Therefore, at the time of RNA isolation these embryos were at embryonic stages 10-17 (Campos-Ortega and Hartenstein, 1997). In heat shock experiments, embryos were collected overnight in the same way, kept for further 4 hours at 25°C and then subjected to a 36°C heat shock for 25 minutes followed by a recovery period of 25 min at 25°C before RNA isolation. Embryos younger than embryonic stage 10 were not used, since heat shock in these earlier stages results in lethality (Walter et al., 1990). Embryos used for *in situ* hybridization studies were collected and heat shock treated in the same way.

Preparation of biotinylated cRNA.

Initial experiments designed to determine the sensitivity and reproducibility of hybridization showed that the use of total RNA versus poly(A)$^+$ RNA as template for cDNA synthesis and subsequent amplification (synthesis of cRNA) gave comparable results, despite the fact that we consistently detected 5S RNA and histone genes present on the array with cRNA derived

from total RNA. Based on these findings, all experiments were carried out using a total RNA protocol (Mahadevappa and Warrington, 1999; Certa et al., 2001).

Total RNA was isolated from 200 mg of embryonic tissue, using guanidinium isothiocyanate in combination with acidic phenol (pH 4.0) (fast RNA tube green kit from BIO101) in a fast prep homogenizer FP120 (Bio 101). After precipitation the RNA was dissolved in DEPC-treated water (Ambion) and spectrophotometrically quantified using a GeneQuant RNA/DNA calculator (Pharmacia Biotech). The quality of the RNA was checked on a 0.5x TBE agarose gel and the samples were stored at -80°C. cDNA was synthesized upon total RNA as a template, using the SuperScript Choice System for cDNA synthesis (Gibco/BRL) with a T7-(T)24 DNA primer. This primer (5'-GGCCAGTGAATTGTAATACGACTCACTATAG GGAGGCGG-(T)24VN-3') was PAGE-gel purified. For first strand cDNA synthesis, a typical 40 µl reaction contained 25 µg RNA, 200 pmoles T7-(T)24 primer, 500 µM of each dNTPs and 800 units reverse transcriptase (AMV Superscript II). The reaction was incubated for one hour at 42°C. Second strand cDNA synthesis was carried out at 18°C for two hours in a total volume of 340 µl, using 20 units *E. coli* DNA ligase, 80 units *E. coli* DNA polymerase I and 4 units RNase H in the presence of 250 µM of each dNTP. After 2nd strand cDNA synthesis, 0.5 µl RNase A (100mg/ml) (Qiagen) was added and the samples were incubated at 37°C for half an hour. Thereafter 7.5 µl proteinase K (10mg/ml) (Sigma) was added and the samples were further incubated at 37°C for another half hour. After cDNA synthesis was completed, samples were phenol-chloroform extracted (3 times) using Phase Lock Gel (5 Prime-3 Prime, Inc.) and precipitated overnight at -20°C with 2.5 volumes 100 % ethanol. After precipitation, the samples were stored at -20°C. Biotinylated antisense cRNA was synthesized from the dsDNA template, using T7 RNA polymerase (MEGAscript T7 Kit, Ambion, Inc.). A 20 µl reaction volume contained between 0.3-1.5 µg cDNA, 7.5 mM of both ATP and GTP, 5.6 mM of both UTP and CTP and 1.8 mM of both biotinylated Bio-16-UTP and Bio-11-CTP (ENZO diagnostics) and 2 µl 10x T7 enzyme mix. The reaction was incubated at 37°C for 8 hours. Thereafter the unincorporated NTPs were removed by putting the sample over an RNeasy spin column (Qiagen). Aliquots of the reaction before and after cRNA synthesis were analysed by agarose gel electrophoresis. Samples were precipitated overnight at -20°C, taken up in 20 µl DEPC treated water and spectrophotometrically quantified. Thereafter, 40 µg of the biotinylated antisense cRNA was fragmented by heating

the sample to 95°C for 35 min in a volume of 25 µl, containing 40 mM tris-acetate (pH 8.1), 100 mM KOAc, 30 mM MgOAc. After the fragmentation, the samples were placed on ice.

High-density oligonucleotide arrays.

In this study, a custom designed *Drosophila* oligonucleotide array (Affymetrix Inc., ROEZ003A) was used. The genes represented on the array correspond to 1519 sequenced *Drosophila* genes encoding open reading frames deposited in SWISS-PROT/TrEMBL databases as of spring 1998 (a complete list of these genes will be given on our web-site). Each gene is represented on the array by a set of 20 oligonucleotide probes (25-mers) matching the gene sequence. To control the specificity of hybridization the same probes are synthesized with a single nucleotide mismatch in a central position. As such, each gene is represented by 20 probe pairs comprised of a perfect match and a mismatch oligo. The difference between the perfect match hybridization signal and the mismatch signal is proportional to the abundance of a given transcript (Lipshutz et al., 1999). *Drosophila* genes, which were not unambiguously represented by a probe set of 20 probe pairs on the array, were excluded from further analysis (23 probe sets were not used). The oligonucleotide probe selection corresponding to each *Drosophila* gene and the array fabrication was performed by Affymetrix Inc.

Hybridization and scanning.

Gene Chips (stored at 4°C) were allowed to warm up to room temperature and were pre-hybridized with 220 µl hybridization buffer (1x MES (pH 6.7), 1 M NaCl, 0.01 % triton, 0.5 µg/µl acetylated BSA, 0.5 µg/µl sonicated herring sperm DNA) for 15 min at 45°C on a rotisserie at 60 rpm. Hybridization was done in a final volume of 220 µl hybridization buffer, containing 40 µg fragmented biotinylated cRNA. The samples were heated to 95°C for 5 min and briefly spun down. Hybridizations were carried out for 16 hours at 45°C with mixing on a rotisserie at 60 rpm. After hybridization, the solutions were removed, arrays were briefly rinsed with 6x SSPE-T (0.9 M NaCl, 0.06 M NaH_2PO_4, 6 mM EDTA, 0.01 % triton) and washed on a Fluidics station (Affymetrix Inc.). Hybridized arrays were stained with 220 µl detection solution (1x MES buffer, containing 2.5 µl streptavidin-R phycoerythrin conjugate (1mg/ml) (Molecular Probes) and 2.0 mg/ml acetylated BSA (Sigma) at 40°C for 15 min and washed again (Certa et al., 2001).

Data analysis.

Probe arrays were scanned with a commercial confocal laser scanner (Hewlett-Packard). Pixel intensities were measured and expression signals were analysed with commercial software (Genechip 3.1, Affymetrix Inc.). Detailed data analysis was carried out using RACE-A (Roche), Access 97 and Excel 97 (Microsoft) software. For quantification of relative transcript abundance the average difference value (Avg Diff) was used. Four replicates for wild type (condition 1) as well as heat shock treated wild type (condition 2) embryos were carried out. All chips were normalized against the mean of the total sums of Avg Diff values across all 8 chips. For the analysis of expression profiling of condition 1 embryos, two filter operations were combined. First, all genes with a mean Avg Diff over the 4 replicate chips that was below 50 were excluded from further analysis. Second, a transcript was judged as present only if the standard deviation of its mean Avg Diff value over the 4 replicate chips was below 25% of its mean Avg Diff. For differential transcript imaging, only genes with a change factor quality above 1 were considered in this analysis, meaning that the difference of the means of the Avg Diff values over the 4 replicates between condition 1 and condition 2 was larger than the sum of the standard deviations of the mean Avg Diff values of condition 1 and condition 2 (RACE-A software, Neeb and Broger, unpublished results). In addition, for downregulation, the mean Avg Diff value of a gene had to be above or equal to 50 in condition 1; for upregulation, the mean Avg Diff value of a gene had to be above or equal to 50 in condition 2.

Whole mount in situ **hybridization.**

Digoxigenin-labeled sense and antisense RNA probes were generated *in vitro*, with a DIG labeling kit (Roche diagnostics), using commercially available templates (Research Genetics, Inc) and hybridized to *Drosophila* whole mount embryos following standard procedures (Tautz and Pfeifle, 1989). Hybridized transcripts were detected with an alkaline phosphatase conjugated anti-digoxigenin Fab fragment (Roche diagnostics) using Nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (Sigma) as chromogenic substrates. Embryos were mounted in Canada balsam (Serva) and photographed with a Prog/Res/3008 digital camera (Kontron Electronic) on a Zeiss Axioskop microscope with differential interference contrast optics.

Functional classification.

The *Drosophila* genes represented on the high-density oligonucleotide array were classified into 14 functional classes according to the function of the gene product and currently available genetic data. For this, notations in Flybase, Interactive Fly, and SWISS-PROT/TrEMBL databases were used. Representative genes for each of the functional classes are listed as follows. *Signal transduction*: cytoplasmic proteins involved in intracellular signalling/ MAP-kinases/ cAMP, cGMP dependent kinases/ small GTP-ases/ ras oncogene-like proteins/ SH3-SH2-SH3 domain proteins; *Transcriptional regulation*: Transcription factors such as homeodomain proteins, zinc finger proteins, ETS proteins, Pou domain proteins/ nuclear hormone and steroid receptors/ Polycomb- and Trithorax group proteins; *Cell cycle*: cyclins/ cyclin dependent kinases; *Cytoskeleton/ structural proteins*: proteins involved in cytoskeletal organization such as actin, actin filament-associated proteins, microtubule-associated proteins, dynein, kinesin/ proteins involved in muscle contraction such as myosin, tropomyosin/ yolk proteins/ chorion proteins/ nuclear envelope proteins; *Metabolism*: general "house-keeping" proteins/ enzymes/ soluble calcium binding proteins/ pheromone binding and odorant binding proteins/ ABC transporters/ pigment proteins/ antibacterial peptides/ proteins involved in nucleotide synthesis/ cytochromes; *Translation*: ribosomal proteins/ proteins involved in translational regulation/ tRNA synthetases; *Heat shock proteins*: Heat shock proteins and Heat shock cognate proteins; *Transcription/ replication/ repair*: RNA polymerases/ TATA binding factors/ DNA polymerases/ DNA helicases/ proteins involved in DNA damage and repair; *Proteolytic systems/ apoptosis*: ubiquitinases/ ubiquitin-activated enzymes/ proteasome subunits/ trypsin/ serine proteases/ proteins involved in apoptotic pathways; *Cell surface receptors/ CAMs/ ion channels*: transmembrane signalling receptors/ glutamate receptors/ GABA receptors/ acetylcholine receptors/ membrane associated antigens/ transmembrane phosphatases and kinases/ ion channel subunits/ cell adhesion molecules/ rhodopsins; *Transposable elements*: F-, copia-, HET-A-, gypsy-, P-elements, transposable element-encoded ORFs; *Chromatin structure*: DNA binding proteins not involved in transcription/ histones/ nucleosome associated proteins/ centrosome associated proteins/ proteins involved in chromosomal segregation; *RNA binding*: RNA helicases/ proteins involved in RNA localization/ RNA binding proteins; *Secreted proteins*: secreted signalling proteins/ ligands.

RESULTS AND DISCUSSION

Quantitative transcript imaging of genes expressed in postgastrulation embryogenesis under standard conditions.

The oligonucleotide array used contains probe sets that are complementary to 1,519 identified sequenced *Drosophila* genes. Most of these genes (96%) can be grouped into 14 functional categories according to the nature of the encoded protein (Table 1). In a first set of experiments, we used this oligonucleotide array to identify transcripts expressed in wild type embryos raised under standard conditions (25°C). Transcript imaging revealed a total of 563 (37%) of the 1,519 *Drosophila* genes as expressed in embryonic stages 10-17. To document the quantitative reproducibility of the relative expression levels, average difference intensity values, Avg Diff (see *Materials and Methods*) and corresponding standard deviations for the detected transcripts were determined over four experimental replicates (Fig. 1). Over two thirds of the detected transcript types encode proteins involved in metabolism (19.8%), transcriptional regulation (13.1%), cell surface receptors/CAMs/ion channels (11.1%), translation (9.2%) cytoskeleton/cell structure (8.5%) or signal transduction (7.2%).

Marked differences were observed in the range of relative expression levels for the different functional categories (Fig. 2). Highest expression levels were seen for specific genes encoding proteins involved in translation. Thus, of the 21 transcripts with Avg Diff >5,000, 18 encode ribosomal proteins. High expression levels with Avg Diff >4000 are also seen for specific individual transcripts encoding proteins involved in chromatin structure and protein degradation. For example the highest Avg Diff in the functional class protein degradation/ apoptosis is the transcript encoding the Cystatin-like protein (Avg Diff=4792). Some transcripts for proteins involved in signal transduction, DNA transcription/replication/repair, metabolism, as well as the transcript encoding the Heat shock cognate protein 70-4 have maximal Avg Diff in the 3000-4000 range. Surprisingly, elevated expression levels are observed for transcripts encoded by specific transposable elements; in three cases Avg Diff were above 2000, namely for two open reading frames, encoded by the transposon I element and a putative reverse transcriptase, encoded by an F element. Remarkably elevated expression levels are also seen for the transcription factor Box B-binding factor 1 (1,315); for other genes encoding transcription factors such as *snail* (Avg Diff = 394), *glial cells missing*

(237), *islet* (136), and *paired* (64) transcript levels were in the intermediate to low range (Avg Diff <550).

Quantitative transcript imaging of heat-shocked compared to non heat-shocked embryos.

Oligonucleotide arrays were next used to determine transcript profile changes following heat shock exposure. For this, transcript imaging was carried out on stage 10-17 embryos subjected to a 36°C heat shock for 25 min (see *Materials and Methods*). The expression profile from embryonically expressed genes after heat shock was quantitatively compared to the expression profile from embryos raised under standard conditions. Comparative transcript imaging identified 74 genes, distributed among 12 functional classes, whose relative expression level changed in response to heat shock; 36 genes had increased and 38 genes had decreased expression levels (Fig. 3).

Heat shock is known to induce the expression of an evolutionary conserved family of genes, encoding the heat shock proteins (Hsps) (Lindquist and Craig, 1988; Schlesinger, 1990; Nover and Scharf, 1997). Accordingly, in our comparative screen we observed a prominent increase in relative transcript abundances for all genes encoding Hsps represented on the chip and which have been reported to be highly upregulated by heat shock. Transcript imaging detected increases above 3-fold in relative expression levels for 9 genes encoding *Drosophila* heat shock proteins: *Hsp22, 26, 27, 23, DnaJ-1, Hsp67Bc, 83, 70Ab, 70Bb* (Pauli, 1990; Michaud et al., 1997). The largest changes (>10-fold) were observed for *Hsp22*, *Hsp26*, *Hsp27*, and *Hsp23*. This is in accordance with several studies that report that these four small Hsps are expressed during normal fly development and are upregulated under heat shock (Haass et al., 1990; Vazquez et al., 1993). For five other genes known to encode heat shock proteins, *DnaJ-1, Hsp67Bc, 83, 70Ab, 70Bb*, we detect an increase in expression in the 3-6 fold range. All of these genes are known to be responsive to heat shock (Vazquez et al., 1993). The heat shock cognate genes (Hsc) have been reported to be expressed at normal temperatures but are not further induced by heat shock (Craig et al., 1983; Rubin et al., 1993). In accordance with this, we observed no marked change in expression level for *Hsc70-1, Hsc70-4 and Hsc70-5*. We did, however, detect a small increase in expression level for *Hsc70-3*.

Two other genes with increases in relative expression levels above 3-fold are *Shark*, involved in a signalling pathway for epithelial cell polarity (Ferrante et al., 1995) and *anon-23Da*, encoding a protein with currently unknown function. 25 other genes show increased expression levels in the 1.5 to 3-fold range. Heat shock induced expression of these genes in *Drosophila* has not been reported before. However, *Cdc37* is known to interact genetically with *Hsp83* in a common signalling pathway in *Drosophila* (Cutforth and Rubin, 1994), and in several other cases, homologous genes in other eukaryotes are known to be stress-inducible. The gene *kayak* (*kay*) for example is the *Drosophila* homologue of the mammalian *c-fos*. *c-fos* mRNA is induced following exposure to noxious stimuli such as heat, arsenite and heavy metals and recently it has been reported that the human and rodent *c-fos* promoters contain heat shock element consensus sequences, which enhance transcription in response to heat (Ishikawa et al., 1999). A second example is *Tenascin major* (*Ten-m*), encoding a protein implicated in patterning the early fly embryo. The mammalian homologue of *Tenascin major* is the gene *DOC4*, which is known to act downstream of CHOP, a small nuclear protein that mediates changes in cell phenotype in response to stress (Wang et al., 1998).

Heat shock induced decreases in relative expression levels greater than 3-fold are seen for *mus210*, the *Drosophila* homologue of the *xeroderma pigmentosum complementation group C* gene, which is involved in DNA repair, and for *anon-X*, which encodes a novel WD repeat protein of unknown function (Henning et al., 1994; Kraemer et al., 1998). The remaining 36 genes with decreased relative expression levels are in the 1.5-fold to 3-fold range. A decrease in relative expression in response to heat shock has not been reported previously for any of these genes in *Drosophila*.

For most of the 74 identified genes, which show differential expression levels in response to heat shock, changes are in the 1.5- to 3-fold range. It was not possible to unambiguously reveal these small quantitative changes using qualitative detection techniques such as in situ hybridization. Changes in gene expression that are in higher ranges can, however, be detected with in situ hybridization. To document this, whole mount in situ hybridization was carried out for transcripts of *Hsp22* (19-fold increase), *Hsp26* (14-fold increase) and *DnaJ-1* (6-fold increase) (Fig. 4). In all three cases, in situ hybridization revealed clear increases in hybridization signal following heat shock.

Taken together, these results demonstrate that oligonucleotide arrays have the potential to analyse the relative expression levels of hundreds of known genes in a complex RNA sample

of the multicellular *Drosophila* embryo. In addition, they allow a quantitative assessment of differential gene expression under normal versus heat shock conditions. Thus, the oligonucleotide probe arrays used in our study establish highly reproducible transcript images of *Drosophila* embryos and allow accurate comparisons of changes in gene expression under different environmental conditions. In this respect, they complement the DNA microarray technique that has recently been used to study gene expression during metamorphosis in *Drosophila* (White et al., 1999). With the imminence of whole genome sequence data for *Drosophila* (Adams et al., 2000) it will now be possible to expand quantitative transcript imaging to include all functional genes and set the stage for a complete genomic analysis of expression profiles in normal and environmentally or genetically manipulated *Drosophila* embryos.

Acknowledgments

We thank Jan Mous, Adrian Roth, Michel Tessier, Monika Seiler and Reto Brem for essential contributions and helpful advice. We are particularly grateful to Clemens Broger and Martin Neeb for allowing us to use their RACE-A Chip analysis software, prior to publication. This research was funded by grants from the SNSF and EU Biotech (to H.R.) and by Hoffmann-La Roche.

FIGURES

Figure 1. Gene expression monitoring of stage 10-17 wild type embryos raised under standard conditions (25°C). Compilation of the 100 genes expressed with the highest Average Difference values (Avg Diff; for details see *Materials and Methods*) and the corresponding standard deviations (SD; indicated by bars) over 4 experimental replicates.

Fig. 1

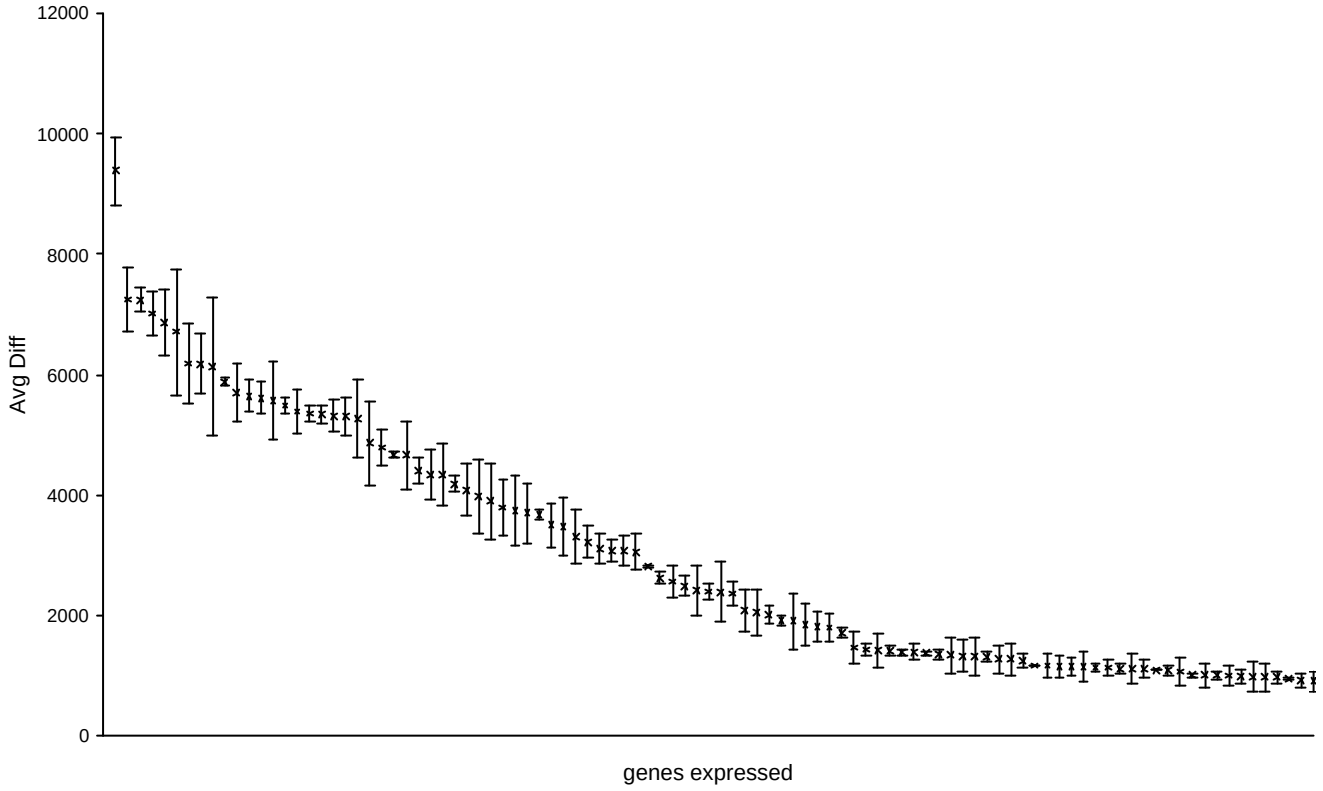


Table 1. Genes expressed in stage 10-17 wild type embryos raised under standard conditions (25°C), grouped according to functional classes. Number of genes within a functional group present on the chip (N); total number of genes represented on the chip ΣN=1519. Number of genes expressed within a functional group (n); total number of transcripts detected Σn=563. (n/N x 100 in %) Distribution of genes expressed within a functional group in relation to the total number of identified genes in this group present on the chip. Distribution of genes expressed within a functional group, given as percentage of the total number of genes expressed.

Drosophila oligonucleotide array: expression data for wildtype embryos

Functional class	Number of genes on the chip (N)	Number of transcripts detected (n)	n/N x 100 (%)	Transcripts detected (%)
Metabolism	315	112	35.5	19.8
Transcriptional regulation	268	74	27.6	13.1
Cell surface receptors/CAMs/ion channels	181	63	34.8	11.1
Translation	60	52	86.6	9.2
Cytoskeleton/structural proteins	149	48	32.2	8.5
Signal transduction	107	41	38.3	7.2
RNA binding	59	29	49.1	5.1
Transcription/replication/repair	73	28	38.3	4.9
Unknown function	85	23	27.0	4.0
Proteolytic systems/apoptosis	62	22	35.4	3.9
Cell cycle	37	18	48.6	3.1
Transposable elements	35	18	51.4	3.1
Chromatin structure	36	18	50.0	3.1
Heat shock proteins	18	10	55.5	1.7
Secreted proteins	34	7	20.5	1.2
	ΣN = 1519	Σn = 563		

Figure 2. Range of Avg Diff values for expressed genes, grouped according to their functional classes. Translation (min Avg Diff 56 - max Avg Diff 9394), Chromatin structure (78-5873), Proteolytic systems/ apoptosis (53-4792), Signal transduction (52-3791), Tanscription/ replication/ repair (59-3303), Metabolism (51-3223), Heat shock proteins (55-3073), Transposable elements (87-2624), Cytoskeleton/ structural proteins (52-2419), Secreted proteins (59-1317), Transcriptional regulation (51-1315), Cell surface receptors/ CAMs/ ion channels (51-1152), RNA binding (52-1095), Cell cycle (56-405), and unknown function (61-1114).

Fig. 2

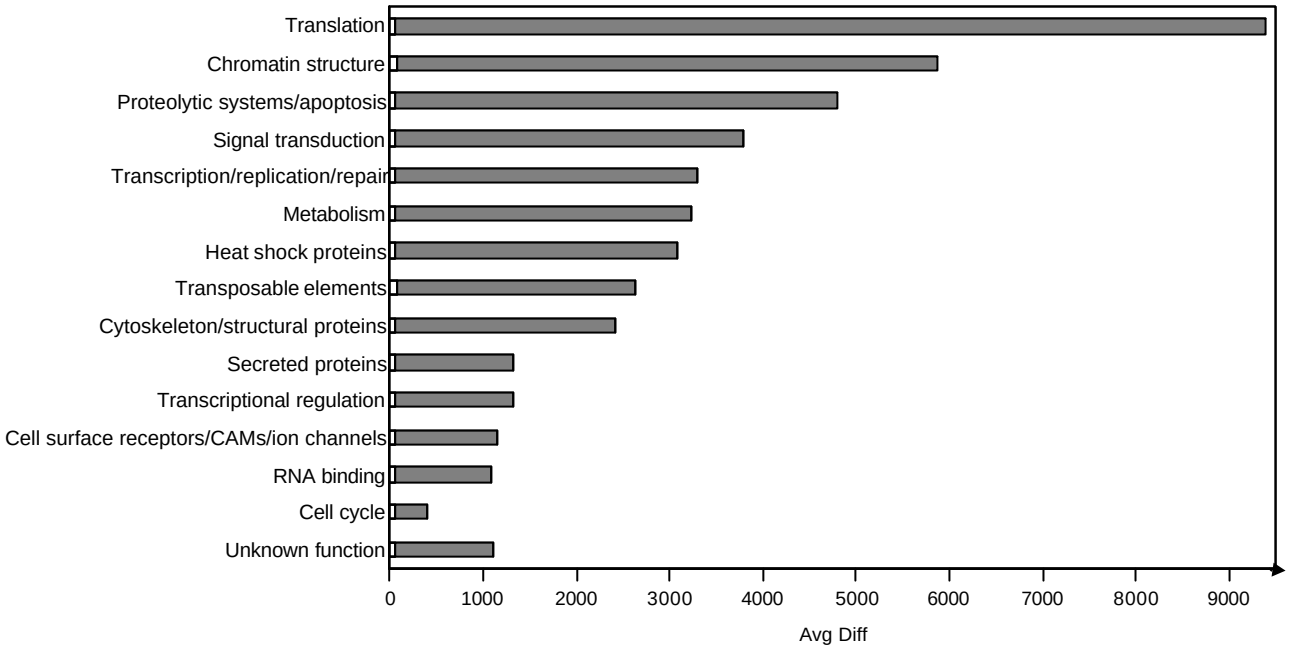

Translation
Chromatin structure
Proteolytic systems/apoptosis
Signal transduction
Transcription/replication/repair
Metabolism
Heat shock proteins
Transposable elements
Cytoskeleton/structural proteins
Secreted proteins
Transcriptional regulation
Cell surface receptors/CAMs/ion channels
RNA binding
Cell cycle
Unknown function
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Avg Diff

Figure 3. Differentially expressed genes observed in heat shocked versus non-heat shocked stage 10-17 wild type embryos, grouped according to functional classes. Bars represent the fold-change of differentially expressed genes in the heat shock versus standard condition. Positive values indicate that the relative expression level of a gene is increased after heat heat shock and negative values indicate a decrease. Avg Diff values are given for the heat shocked condition as follows: white bars represent Avg Diff<100, grey bars represent Avg Diff ranging from 100-1000 and black bars represent Avg Diff>1000.

Fig. 3

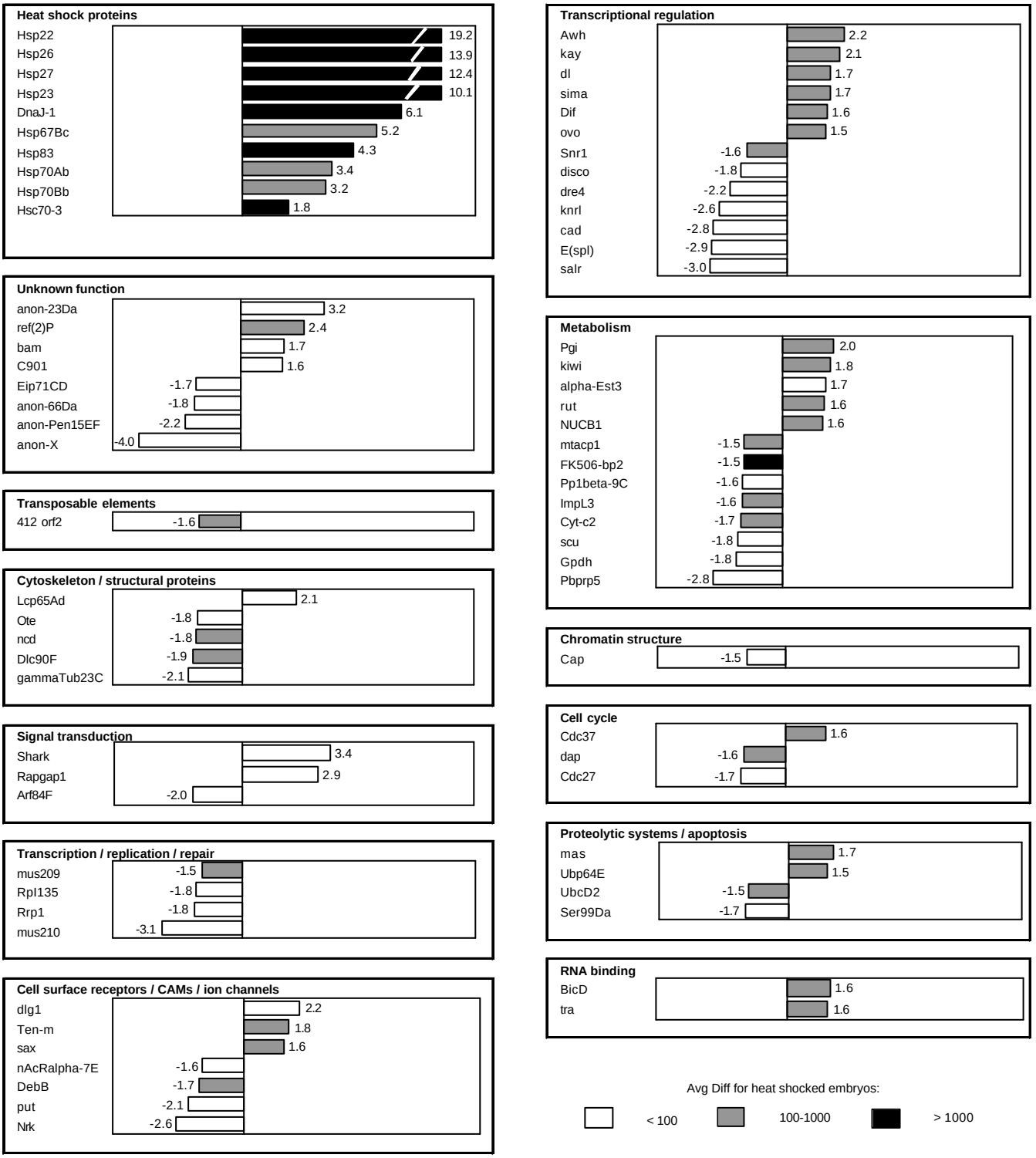

Heat shock proteins
Hsp22 19.2
Hsp26 13.9
Hsp27 12.4
Hsp23 10.1
DnaJ-1 6.1
Hsp67Bc 5.2
Hsp83 4.3
Hsp70Ab 3.4
Hsp70Bb 3.2
Hsc70-3 1.8
Unknown function
anon-23Da 3.2
ref(2)P 2.4
bam 1.7
C901 1.6
Eip71CD -1.7
anon-66Da -1.8
anon-Pen15EF -2.2
anon-X -4.0
Transposable elements
412 orf2 -1.6
Cytoskeleton / structural proteins
Lcp65Ad 2.1
Ote -1.8
ncd -1.8
Dlc90F -1.9
gammaTub23C -2.1
Signal transduction
Shark 3.4
Rapgap1 2.9
Arf84F -2.0
Transcription / replication / repair
mus209 -1.5
RpI135 -1.8
Rrp1 -1.8
mus210 -3.1
Cell surface receptors / CAMs / ion channels
dlg1 2.2
Ten-m 1.8
sax 1.6
nAcRalpha-7E -1.6
DebB -1.7
put -2.1
Nrk -2.6
Transcriptional regulation
Awh 2.2
kay 2.1
dl 1.7
sima 1.7
Dif 1.6
ovo 1.5
Snr1 -1.6
disco -1.8
dre4 -2.2
knrl -2.6
cad -2.8
E(spl) -2.9
salr -3.0
Metabolism
Pgi 2.0
kiwi 1.8
alpha-Est3 1.7
rut 1.6
NUCB1 1.6
mtacp1 -1.5
FK506-bp2 -1.5
Pp1beta-9C -1.6
ImpL3 -1.6
Cyt-c2 -1.7
scu -1.8
Gpdh -1.8
Pbprp5 -2.8
Chromatin structure
Cap -1.5
Cell cycle
Cdc37 1.6
dap -1.6
Cdc27 -1.7
Proteolytic systems / apoptosis
mas 1.7
Ubp64E 1.5
UbcD2 -1.5
Ser99Da -1.7
RNA binding
BicD 1.6
tra 1.6
Avg Diff for heat shocked embryos:
< 100
100-1000
> 1000

Figure 4. Comparison of whole mount in situ hybridizations between non-heat shocked and heat shocked wild type embryos. (*A-F*) lateral views, anterior to the left. (*A, C, E*) non-heat shocked wild type, (*B, D, F*) heat shocked wild type embryos. (*A, B*) at stage 11 *Hsp22* expression is confined to metameric ectodermal patches in non-heat shocked wild type embryos (*A*) whereas *Hsp22* is ubiquitously expressed in the ectoderm of heat shocked wild type embryos (*B*). (*C, D*) at stage 12 there is no expression of *Hsp26* in the ectoderm of non-heat shocked wild type embryos (*C*, gut staining out of focal plane) whereas *Hsp26* is expressed in all ectodermal cells of heat shocked wild type embryos (*D*) (*E;F*) at stage 11 *DnaJ-1* is not detected in non-heat shocked wild type embryos (*E*) whereas heat shocked wild type embryos show strong expression in all ectodermal cells (*F*).

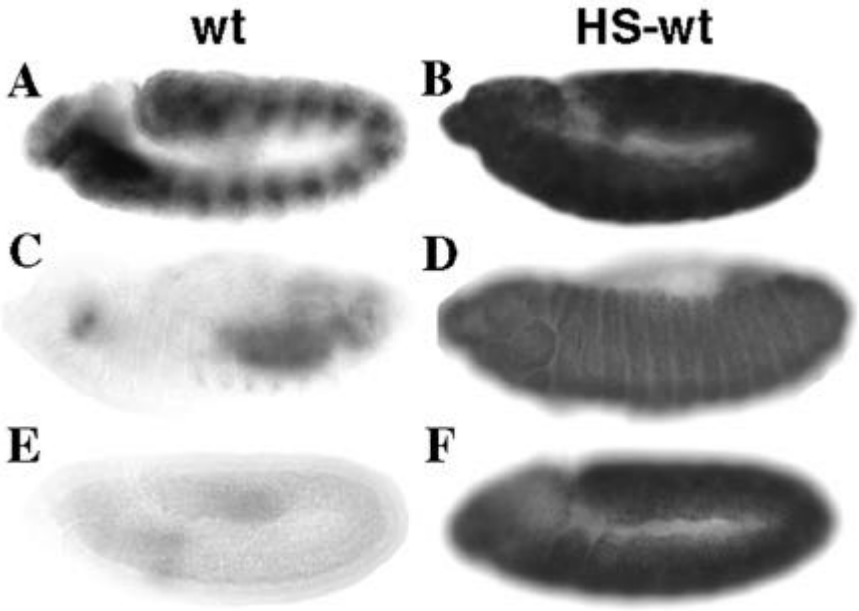

wt
HS-wt
A
B
C
D
E
F

Table 2. RT-PCR was performed on cDNA derived from heat-shocked embryos and embryos raised under standard conditions. Change folds determined by RT-PCR are represented as the mean value of eight independent replicates, derived from two different cDNA preparations. Wt, wild type; HS, heat shock.

Comparison of change folds between oligonucleotide arrays and RT-PCR

Gene	Avg Diff (Array)		Change Fold	
	wt	HS-wt	Array	RT-PCR
Hsp27	347	4646	12.4	20.0
Hsp67Bc	183	944	5.2	8.0
anon-23Da	6	64	3.2	2.6
kay	74	153	2.1	1.4
Ten-m	92	162	1.8	2.1
kiwi	108	199	1.8	4.0
Cdc37	179	286	1.6	3.4
Rac2	424	425	1.0	1.1
FK506-bp2	1918	1248	-1.5	-2.0

4. Identification of candidate downstream genes for the homeodomain transcription factor Labial in *Drosophila* through oligonucleotide array transcript imaging

Ronny Leemans[1], Thomas Loop[1], Boris Egger[1], Haiqiong He[1], Lars Kammermeier[1], Beate Hartmann[1], Ulrich Certa[2], Heinrich Reichert[1] and Frank Hirth[1*]

1) Institute of Zoology, Biocenter/Pharmacenter, University of Basel, Basel, Switzerland
2) PRGT, Hoffmann-La Roche Ltd., Basel, Switzerland

Genome Biology (2001), **2 (5)**: research0015.1-0015.9

SUMMARY

Homeotic genes are key developmental regulators that are highly conserved throughout evolution. Their encoded homeoproteins function as transcription factors to control a wide range of developmental processes. Although much is known about homeodomain-DNA interactions, only a small number of genes acting downstream of homeoproteins have been identified. Here we use a functional genomic approach to identify candidate target genes of the *Drosophila* homeodomain transcription factor Labial (Lab).

High-density oligonucleotide arrays with probe sets representing 1513 identified and sequenced genes were used to analyse differential gene expression following *lab* overexpression in *Drosophila* embryos. We find significant expression level changes for 96 genes belonging to all functional classes represented on the array. In accordance with our experimental procedure, we expect that these genes are either direct or indirect targets of *labial* gene action. Among these genes, 48 were upregulated and 48 were downregulated following *lab* overexpression. This corresponds to 6.3% of the genes represented on the array. For a selection of these genes we show that the data obtained with the oligonucleotide arrays are consistent with data obtained using quantitative RT-PCR.

Our results identify a number of novel candidate downstream genes for Labial, suggesting that this homeoprotein differentially regulates a limited and distinct set of embryonically expressed *Drosophila* genes.

INTRODUCTION

The homeotic/Hox genes encode a network of evolutionary conserved homeodomain transcription factors that are involved in the specification of segmental identity along the anterior-posterior body axis of animals as diverse as insects and vertebrates (McGinnis and Krumlauf, 1992; Krumlauf, 1994; Manak and Scott, 1994; Carroll, 1995; Maconochie et al., 1996; Gellon and McGinnis, 1998). In *Drosophila*, these genes are arranged along the chromosome in two gene clusters known as the Antennapedia and Bithorax complexes. There is a correlation between the relative position of the Hox genes within the cluster and their

spatial and temporal expression pattern in the body, in that genes located towards the 3' end are expressed more anterior and earlier than genes located towards the 5' end (spatial and temporal colinearity) (Duboule and Morata, 1994; Lawrence and Morata, 1994; Graba et al., 1997; Morata and Sanchez-Herrero, 1999; Mann and Morata, 2000).

Given their central role in developmental processes, it has been proposed that the homeoproteins do not act directly to specify morphological differences but rather control a battery of subordinate genes encoding cellular functions directly required in differentiation (Garcia-Bellido, 1975; Pradel and White, 1998). In search of these subordinate genes, various strategies such as enhancer trapping, immunoprecipitation of chromatin fragments, subtractive hybridization, selection for binding sites in yeast, and heat shock induced overexpression have been employed (Andrew and Scott, 1992; Morata, 1993; Gehring et al., 1994a; Mastick et al., 1995; Botas and Auwers, 1996; Graba et al., 1997; Mannervik, 1999; Nasiadka and Krause, 1999; Nasiadka et al., 2000). However, only a small number of target genes of homeoproteins have been identified to date; most of these encode either transcription factors or cell signalling molecules (Graba et al., 1997). In contrast to these results, recent studies suggest that homeoproteins may bind at significant levels to the majority of genes in the *Drosophila* embryo and regulate a large number of downstream genes (Biggin and McGinnis, 1997; Liang and Biggin, 1998).

Here we focus on the homeotic gene *labial (lab)* in the *Drosophila* embryo. *lab* is the most proximal gene located within the *Drosophila* Antennapedia complex; it encodes an antennapedia-like Q50 homeodomain transcription factor and is one of the most anteriorly expressed homeotic genes along the anterior-posterior body axis (Mlodzik et al., 1988; Diederich et al., 1989; Kaufman et al., 1990; Duboule, 1994). Genetic studies have demonstrated that *lab* is required for proper head formation (Merrill et al., 1989), for the specification of cellular identity in the midgut (Hoppler and Bienz, 1994) as well as in the embryonic brain (Hirth et al., 1998). The *lab* gene and its vertebrate *Hox1* orthologs are among the best characterized examples of evolutionary conservation of structure, expression and function of Hox genes in animal development (Pöpperl et al., 1995; Chan and Mann, 1996; Lutz et al., 1996; Mann and Chan, 1996; Hirth and Reichert, 1999).

In order to address the question of which and how many downstream genes are under control of *lab*, we used a combination of *in vivo* overexpression techniques and quantitative transcript imaging with oligonucleotide arrays. By using transgenic flies carrying the *lab* gene under

control of a heat inducible promoter, we ubiquitously overexpressed *lab* following heat shock treatment in *Drosophila* embryos. We then used high density oligonucleotide arrays representing 1513 identified *Drosophila* genes for large scale detection and quantification of induced gene expression (Lockhart et al., 1996; Lipshutz et al., 1999; Leemans et al., 2000; Rubin, 2000). We find significant changes in gene expression for 96 identified genes following *lab* overexpression. Quantitative RT-PCR on a selection of these genes verified the differential expression levels in response to heat shock induced overexpression of *lab*. Our findings identify a number of novel candidate downstream genes for *lab* and, thus, demonstrate that oligonucleotide arrays are powerful tools for analysing, at a genome wide level, the number, identity and quantitative expression level of genes in the *Drosophila* embryo.

MATERIALS AND METHODS

Fly strains, embryo collections and heat shock regime

The wild type was *Drosophila melanogaster* Oregon-R. For ectopic overexpression of *lab*, we used the line *p (w*$^{+}$hs-*lab)* with a heat shock *lab* construct homozygous on the X chromosome (Heuer and Kaufman, 1992). All fly stocks were kept on standard cornmeal/yeast/agar medium at 25°C. Embryos were collected overnight for 12 hours on grape juice plates, further kept for 4 hours at 25°C and then subjected to a 36°C heat shock for 25 min, followed by a recovery period of 25 min at 25°C before RNA isolation. Therefore, at the time of RNA isolation these embryos were at embryonic stages 10-17 stages according to (Campos-Ortega and Hartenstein, 1997)). Embryos younger than embryonic stage 10 were not used, since heat shock in these earlier stages results in lethality (Walter et al., 1990).

Whole mount in situ hybridization and immunocytochemistry.

For in situ hybridization, digoxigenin-labeled sense and antisense *lab* RNA probes were generated *in vitro*, with a DIG labeling kit (Roche Diagnostics) and hybridized to whole mount embryos following standard procedures (Tautz and Pfeifle, 1989). Hybridized transcripts were detected with an alkaline phosphatase conjugated anti-digoxigenin Fab fragment (Roche Diagnostics) using Nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-

indolyl phosphate (BCIP) (Sigma) as chromogenic substrates. For Immunocytochemistry, embryos were dechorionated, fixed and labeled according to (Hirth et al., 1995). The primary antibody was rabbit anti-LAB (Grieder et al., 1997) used 1:100. The histochemical staining was performed using the Vectastain Elite ABC Kit (Vector Laboratories). Embryos were mounted in Canada balsam (Serva) and photographed with a Prog/Res/3008 digital camera (Kontron Electronic) on a Zeiss Axioskop microscope with differential interference contrast optics. Photographs were arranged and labeled using Microsoft PowerPoint, 97.

High-density oligonucleotide arrays

Gene expression analysis was performed as described (Lockhart et al., 1996) using a custom designed *Drosophila* oligonucleotide array (ROEZ003A; Affymetrix Inc). The genes represented on the array and considered in this study correspond to 1513 sequenced *Drosophila* genes encoding open reading frames deposited in SWISS-PROT/TrEMBL databases as of spring 1998. (For a complete list of these genes see supplementary data of (Leemans et al., 2000)). Each gene is represented on the array by a set of 20 oligonucleotide probes (25-mers) matching the gene sequence. To control the specificity of hybridization, the same set of probes, containing a single nucleotide mismatch in a central position, are represented on the array. The difference between the perfect match hybridization signal and the mismatch signal is proportional to the abundance of a given transcript and calculated as its Average Difference value (Avg Diff) (Lipshutz et al., 1999). *Drosophila* genes, which were not unambiguously represented by a probe set of 20 probe pairs on the array, were excluded from further analysis (29 probe sets were not used in this study).

RNA sample preparation and hybridization

Initial experiments designed to determine the sensitivity and reproducibility of hybridization showed that the use of total RNA versus poly $(A)^+$ RNA as a template for cDNA synthesis and subsequent amplification (synthesis of cRNA) gave comparable results, despite the fact that we consistently detected 5S RNA and histone genes present on the array with cRNA derived from total RNA. Based on these findings, all experiments were carried out using a total RNA protocol (Mahadevappa and Warrington, 1999).

Total RNA was isolated from 200 mg of embryonic tissue, using guanidinium isothiocyanate in combination with acidic phenol (pH 4.0) (fast RNA tube green kit from BIO101) in a fast

prep homogenizer FP120 (BIO 101). After precipitation, the RNA was dissolved in DEPC-treated water (Ambion) and spectrophotometrically quantified using a GeneQuant RNA/DNA calculator (Pharmacia Biotech). cDNA was synthesized upon total RNA as a template, using the SuperScript Choice System for cDNA synthesis (Gibco/BRL) with a T7-(T)24 DNA primer.

This primer (5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-(T)24VN-3') was PAGE-gel purified. For first strand cDNA synthesis, a typical 40 µl reaction contained 25 µg RNA, 200 pmoles T7-(T)24 primer, 500 µM of each dNTPs and 800 units reverse transcriptase (AMV Superscript II). The reaction was incubated for one hour at 42°C. Second strand cDNA synthesis was carried out at 18°C for two hours in a total volume of 340 µl, using 20 units *E. coli* DNA ligase, 80 units *E. coli* DNA polymerase I and 4 units RNase H in the presence of 250 µM of each dNTP. After 2nd strand cDNA synthesis, 0.5 µl RNase A (100mg/ml) (Qiagen) was added and the samples were incubated at 37°C for half an hour. Thereafter 7.5 µl proteinase K (10mg/ml) (Sigma) was added and the samples were further incubated at 37°C for another half hour. After cDNA synthesis was completed, samples were phenol-chloroform extracted, using Phase Lock Gel (5 Prime-3 Prime, Inc.) and ethanol precipitated. Biotinylated antisense cRNA was synthesized from the dsDNA template, using T7 RNA polymerase (MEGAscript T7 Kit, Ambion, Inc.). A 20 µl reaction volume contained between 0.3-1.5 µg cDNA, 7.5 mM of both ATP and GTP, 5.6 mM of both UTP and CTP and 1.8 mM of both biotinylated Bio-16-UTP and Bio-11-CTP (ENZO diagnostics) and 2 µl 10x T7 enzyme mix. The reaction was incubated at 37°C for 8 hours. Thereafter the unincorporated NTPs were removed by putting the sample over an RNeasy spin column (Qiagen). Samples were precipitated, taken up in 20 µl DEPC treated water and spectrophotometrically quantified. Thereafter, 40 µg of the biotinylated antisense cRNA was fragmented by heating the sample to 95°C for 35 min in a volume of 25 µl, containing 40 mM tris-acetate (pH 8.1), 100 mM KOAc, 30 mM MgOAc. After the fragmentation, the samples were placed on ice.

Gene Chips were pre-hybridized with 220 µl hybridization buffer (1x MES (pH 6.7), 1 M NaCl, 0.01 % triton, 0.5 µg/µl acetylated BSA, 0.5 µg/µl sonicated herring sperm DNA) for 15 min at 45°C on a rotisserie (Heidolph, Schwabach, Germany) at 60 rpm. Hybridization was done in a final volume of 220 µl hybridization buffer, containing 40 µg fragmented

biotinylated cRNA. The samples were heated to 95°C for 5 min and briefly spun down. Hybridizations were carried out for 16 hours at 45°C with mixing on a rotisserie at 60 rpm. After hybridization, the arrays were briefly rinsed with 6x SSPE-T (0.9 M NaCl, 0.06 M NaH_2PO_4, 6 mM EDTA, 0.01 % triton) and washed on a Fluidics station (Affymetrix Inc.). Hybridized arrays were stained with 220 µl detection solution 1x MES buffer, containing 2.5 µl streptavidin-R phycoerythrin conjugate (1mg/ml) (Molecular Probes)) and 2.0 mg/ml acetylated BSA (Sigma) at 40°C for 15 min and washed again.

Data analysis

Pixel intensities were measured with a commercial confocal laser scanner (Hewlett Packard) and expression signals were analysed with commercial software (Genechip 3.1, Affymetrix Inc.). Detailed data analysis was carried out using Race-A (Roche), Access 97 and Excel 97 (Microsoft) software. For quantification of relative transcript abundance the normalized Average Difference value (Avg Diff) was used. For each of the three experimental conditions (wt, hs-wt, hs-*lab*), four replicates were carried out (for the experimental conditions wt and hs-wt see (Leemans et al., 2000), including supplementary data). For the difference of the means of the Avg Diff values over the 4 replicates between condition 1 (hs-wt) and condition 2 (hs-*lab*) a t-test was performed. Moreover, for downregulation, the mean Avg Diff value of a gene had to be above or equal to 50 in condition 1; for upregulation, the mean Avg Diff value of a gene had to be above or equal to 50 in condition 2. Genes, which had a normalized Avg Diff below 20 obtained automatically an Avg Diff of 20 (Race-A protocol). To obtain a comprehensive analysis of the number and identity of genes differentially regulated by *lab*, candidates that were already differentially expressed in heat shock treated wild type embryos compared to non heat shocked wild type controls, were excluded from further analysis (data not shown; (Leemans et al., 2000)). Previously, we have used quantiative RT-PCR to confirm that relative expression level changes in the 1.5-fold and above range, as detected on this array, accurately reflect differences in mRNA abundance *in vivo* in *Drosophila* embryos (Leemans et al., 2000). In consequence, in this report only relative expression level changes in the 1.5-fold and above range are presented.

Reverse Transcriptase PCR (RT-PCR)

Three hundred ng of poly(A)$^+$ RNA, isolated from heat shocked wild type embryos and heat shocked hs-*lab* embryos (mRNA isolation kit; Roche Diagnostics), was reverse transcribed with AMV-RT and random hexamers (first-strand cDNA synthesis kit for RT-PCR; Roche Diagnostics). PCR was performed with 100 pg of template DNA and gene-specific primers (designed, using SEQ WEB, Wisconsin Package Version 10.0, GCG) on a light cycler (LightCycler, Roche Diagnostics). Continuous fluorescence observation of amplifying DNA was made possible by using SYBR Green I (LightCycler- FastStart DNA master SYBR GreenI; Roche Diagnostics). After cycling, a melting curve was produced by slow denaturation of the PCR end products to check the specificity of amplification. To compare the relative amounts of PCR products, we monitored the amplification profile on a graph, displaying the log of the fluorescence against the number of cycles. Relative change folds for a given gene under both conditions (heat shock wt vs. heat shock hs-*lab*) were calculated by using the fit point method (LightCycler operator's manual, version 3.0, Roche Diagnostics).

Functional classification

The genes represented on the high-density oligonucleotide array were grouped into 14 functional classes according to the function of the gene product and currently available genetic data (Leemans et al., 2000). For this, notations in Flybase, Interactive Fly, and SWISS-PROT/TrEMBL databases were used. A comprehensive presentation of all the genes represented on the oligonucleotide array as well as their attribution to functional classes is given as supplementary data at website (www.pnas.org).

RESULTS

In this study, transgenic fly strains carrying the *lab* coding sequence under control of the heat inducible Hsp70 promotor were used (Heuer and Kaufman, 1992). Stage 10-17 embryos were given a 25 min heat pulse in order to overexpress *lab* and allowed to recover for 25 min (see Materials and methods for heat shock protocol). Ubiquitous overexpression of *lab* was verified by whole mount in situ hybridization with a *lab* specific antisense RNA probe. Ubiquitous overexpression of Labial protein was verified by immunocytochemistry with an

anti-Labial antibody. These experiments demonstrated that both *lab* RNA and Labial protein were strongly overexpressed 50 min after the onset of heat shock in these strains (Fig. 1). Wild type control flies were subjected to the identical heat shock regime.

Following ubiquitous overexpression of *lab*, transcript profiles were analysed using a high-density oligonucleotide array and compared to the transcript profiles of heat shock treated wild type control embryos. For each of the two experimental conditions (hs-wt and hs-*lab*), four replicates were carried out and the data set was analysed with an unpaired *t*-test (see Materials and methods, and (Leemans et al., 2000)). The genes represented on the oligonucleotide array correspond to probe sets that are complementary to 1513 identified and sequenced *Drosophila* genes. Most of these genes can be grouped into 14 functional categories according to the nature of the encoded protein (Leemans et al., 2000).

At a significance level of $p \leq 0.01$, a total of 96 genes were found to be differentially regulated following *lab* overexpression, as compared to heat shocked wild type control embryos. This corresponds to 6.3% of the genes represented on the array. At a significance level of $p \leq 0.05$, 205 genes were found to be differentially regulated following *lab* overexpression as compared to heat shocked wild type control embryos (data not shown). This corresponds to 13.5% of the genes represented on the array. The relative distribution of *lab* regulated genes in particular functional classes as well as the percentage of genes regulated within a given functional class were comparable between the $p \leq 0.01$ group and the $p \leq 0.05$ group. In the following, only genes that were differentially expressed at a significance level of $p \leq 0.01$ are considered further. We posit these genes to be potential direct or indirect downstream genes for the homeodomain transcription factor Labial.

When ubiquitously expressed in the embryo, *lab* caused a significant transcriptional response among a wide variety of genes belonging to all functional classes represented on the array (Table 1). The functional class with the highest absolute number of differentially regulated genes was transcriptional regulation (n=20). Other functional classes with high numbers of differentially regulated genes were metabolism (n=13), proteolytic systems/apoptosis (n=12), cell surface receptors/CAMs/ion channels (n=12), and RNA binding (n=7). Relative to the number of genes represented on the array within a given functional class, the highest relative percentage of differentially regulated genes was found in the functional classes proteolytic systems/apoptosis (19.4%), cell cycle (13.5%), transposable elements (11.4%), chromatin structure (11.1%), RNA binding (11.9%), and transcriptional regulation (7.6%).

Figure 2 shows the *lab* regulated genes and presents a quantitative representation of the change in expression levels for these genes. Of the 96 genes that were differentially regulated, 48 showed increased expression levels and 48 showed decreased expression levels. The gene with the highest increase in expression level (26-fold) was *lab* itself, in accordance with our experimental procedure. Increases in expression levels above 10-fold were also observed for the genes *Bicaudal C* (*BicC*), *swallow* (*swa*) and *oskar* (*osk)*, all encoding proteins involved in RNA binding, as well as for the *wings apart-like* (*wapl*) gene belonging to the functional class chromatin structure. The increased expression levels in *BicC*, *swa*, and *osk* are surprising since all of these genes are known to function as maternal control genes during early embryogenesis (St Johnston and Nusslein-Volhard, 1992; Micklem, 1995). Since *lab* activity is normally only observed from gastrulation onwards (Kaufman et al., 1990), this suggests that high levels of widespread ectopic *lab* expression are able to activate genes which under wild type conditions show non-overlapping spatio-temporal expression domains as compared to that of *lab*. Increases in the 5-10 fold range were seen for six genes. One of these encodes the enzyme Ubiquitin carboxy-terminal hydrolase, whose mammalian homolog has also been found to be differentially upregulated by ectopic overexpression of the *lab* ortholog *Hoxa1* (Shen et al., 2000). Increased expression levels in the 1.5-5 fold range were prominent in several functional classes. For example, in the funtional class proteolytic systems/apoptosis, 12 of 13 differentially regulated genes were upregulated and most of these showed increased expression levels ranging between 1.5 and 5. Strikingly, in the functional class cell cycle and in the functional class transcription/replication/repair all of the differentially regulated genes were upregulated. Thus, differentialy expressed genes such as *twine* (*twe*), *Cyclin B* (*CycB*), and *Cyclin D* (*CycD*), belonging to the functional category cell cycle, were all upregulated following *lab* overexpression. It is noteworthy in this respect that recent experiments carried out on mammalian cell lines demonstrated that ectopic overexpression of the *lab* ortholog *Hoxa1* also causes differential upregulation of cell cycle regulatory proteins (Shen et al., 2000).

Decreases in expression levels in the 10-fold and above range were not observed and decreases in the 5-10 fold range were only seen for the transposable *R2 rDNA element* gene. Decreased expression levels in the 1.5-5 fold range were, however, prominent in the functional class transcriptional regulation and in the functional class cell surface receptors/CAMs/ion channels. Thus, almost 3/4 of the differentially regulated genes encoding

transcription factors showed significant decreases in expression levels following *lab* overexpression. For example, the genes *prospero (pros)*, *Distal-less (Dll)*, *tailup/islet (tup)*, *mirror (mirr)*, *huckebein* (*hkb*), and *abrupt (ab)* were all downregulated. Interestingly, it has been shown that *Distal-less* is a direct target of homeotic gene control (Graba et al., 1997), and recent genetic studies demonstrated that *tailup/islet* expression in the *lab*-specific territory of the embryonic *Drosophila* brain is dependent on *lab* gene action (Hirth et al., 1998). Similar to the situation of the functional class transcriptional regulation, 10 out of 12 genes representing the functional category cell surface receptors/CAMs/ion channels were downregulated, including the genes *derailed* (*drl*), *frizzled 2* (*fz2*), *Neurotactin* (*Nrt*), *Neurexin* (*Nrx*), *rhomboid* (*rho*), and *18 wheeler* (*18w*). As is the case for *tailup/islet*, *Neurotactin* expression in the *lab*-specific territory of the embryonic *Drosophila* brain is dependent on *lab* gene action (Hirth et al., 1998). *18 wheeler* has been identified as a binding site of the homeotic protein UBX in polytene chromosomes (Botas and Auwers, 1996).

To verify the differences in expression level after heat shock induced overexpression of *lab* as compared to heat shocked wild type embryos, quantitative RT-PCR was performed on selected candidate target genes. Changes in expression levels were determined for eight genes that were differentially regulated following *lab* overexpression, namely *labial* (*lab*), *swallow* (*swa*), *Ubiquitin conjugating enzyme 4* (*UbcD4*), *twine* (*twe*), *cyclin B* (*cycB*), *Ubiquitin carboxy-terminal hydrolase* (*Uch*), *scratch* (*scrt*) and *phosphoenolpyruvate carboxykinase* (*Pepck*). The gene *squid* (*sqd*), whose expression level remained unchanged under both experimental conditions, served as a control. As indicated in Table 2, these experiments demonstrated that the changes in relative expression level as measured by RT-PCR, are consistent with the data obtained with the oligonucleotide arrays.

DISCUSSION

In this report we have used a novel combination of manipulative genetics and functional genomics to gain further insight into homeotic gene action in *Drosophila* from a genomic perspective. Using inducible overexpression and quantitative transcript imaging through oligonucleotide arrays, we have identified 96 genes whose expression levels change significantly following *lab* overexpression. Accordingly, of the 1513 identified genes

represented on the oligonucleotide array, only 6.3% showed significant differential regulation following overexpression of the homeoprotein *lab*.

These findings suggest a specific differential regulation of a limited and distinct set of candidate downstream genes for *lab*. As such, this appears to contrast with previous reports indicating that in late embryogenesis the majority of *Drosophila* genes are under control of homeoproteins (Liang and Biggin, 1998; Carr and Biggin, 1999). However, it should be stressed, that there are a number of features of our functional genomic analysis that impede a direct comparison with these reports, which are based on DNA binding studies. First, although our analysis can quantify gene expression accurately and simultaneously for many identified genes, the temporal and spatial resolution of our analysis is low. This is because our experimental design averages gene expression throughout the embryo and during several embryonic stages. In consequence, our analysis may fail to detect genes that are only expressed in a small subset of cells or during a very restricted time period in embryogenesis. Second, our overexpression protocol makes it difficult to control the level of Lab protein as well as the temporal dynamics and stability of this protein. Since different levels of a given homeoprotein can have different functional consequences in terms of developmental specificity (Hoppler and Bienz, 1994; Cribbs et al., 1995), the high level of Lab protein may bias the set of candidate downstream target genes identified. Third, in our studies *lab* overexpression is not accompanied by concomitant overexpression of cofactors, which are thought to act together with homeotic proteins to determine their *in vivo* target specificity (Mann and Chan, 1996; Mann and Affolter, 1998). It is noteworthy in this respect, that the gene *mirror*, which has been proposed to be an additional cofactor for homeoprotein specificity (Mann, 1995), was detected as downregulated following *lab* overexpression.

Although the question of the total number of target genes that are regulated by homeoproteins *in vivo* must await further analysis, our genomic perspective of *lab* gene targets does reveal several specific features of homeoprotein action. First, our results demonstrate that the homeodomain transcription factor *lab* acts on numerous candidate target genes that also encode transcription factors. The category transcriptional regulation comprises one of the largest sets of differentially regulated genes following *lab* overexpression. This is consistent with the idea that homeobox genes establish developmental patterns by acting through a cascade of transcription factors, which regulate the expression of their own subset of downstream genes (McGinnis and Krumlauf, 1992; Morata, 1993; Manak and Scott, 1994;

Graba et al., 1997). Second, our data indicate that upregulation of gene expression is prominent in several functional classes. Thus, virtually all of the *lab* regulated genes in the functional classes cell cycle, transcription/replication/repair, and proteolytic systems/apoptosis show increased expression values. Third, our results show that *lab* overexpression causes not only widespread activation but also widespread repression of gene expression. Thus, of the 96 genes that are potential targets of *lab,* one half is downregulated by overexpression of this homeobox gene. This widespread repression is especially pronounced in the functional classes of transcriptional regulation and cell surface receptors/CAMs/ion channels. For example, following *lab* overexpression over 80% of the differentially regulated genes encoding cell surface receptors/CAMs/ion channels showed decrease in expression level.

CONCLUSION

Taken together, our results identify a large number of novel candidate downstream genes of the homeodomain transcription factor Labial. To our knowledge, most of these 96 identified and sequenced genes have not been previously shown to be *lab* targets. At present, we do not know which genes are direct targets (regulated directly by Labial protein binding to DNA regulatory sequences) or indirect targets of *lab* gene action. Furthermore, our results demonstrate that oligonucleotide arrays are useful tools for analysing, at a genome wide level, the number, identity and quantitative expression levels of candidate downstream genes differentially regulated *in vivo* by developmental control genes. This confirms the general utility of microarrays for studying diverse molecular and cellular processes in *Drosophila* (Bryant et al., 1999b; White et al., 1999; Andrews et al., 2000). Considering the evolutionary conservation of gene structure, expression and function (McGinnis and Krumlauf, 1992; Hirth and Reichert, 1999), we posit that these results obtained in *Drosophila* will also be valid for *lab* gene orthologs in other animal species including vertebrates. It will now be important to determine which of the detected candidate downstream genes in *Drosophila* are direct targets and how they exert the developmental genetic programs imposed by *lab* gene action.

Acknowledgements:

We thank Jan Mous, Adrian Roth, Michel Tessier, Monika Seiler, and Reto Brem for essential contributions and helpful advice. We are particularly grateful to Clemens Broger, Martin Strahm, and Martin Neeb (F. Hoffman-La Roche) for allowing us to use their RACE-A CHIP analysis software and to Volker Schmid and Natalie Yanze for help with the light cycler. This research was supported by grants from the SNSF and the EU-BIOTECH program (to H.R.) and by F. Hoffmann-La Roche.

FIGURES

Figure 1. Heat shock driven ubiquitous overexpression of *lab* monitored by in situ hybridization and immunocytochemistry. (A, B, C, D) RNA in situ hybridization. (E, F, G, H) immunocytochemical staining. *lab* gene expression is shown in heat shocked wild-type embryos (A, C, E, G) and in heat shocked embryos carrying a hs-*lab* construct (B, D, F, H). (A, B, E, F) Overview of stage 10-17 embryos. (C, D) Higher magnification of a single stage 15 embryo and (G, H) a single stage 13 embryo; lateral view and anterior is to the left. Embryos were exposed to a heat shock at 36° C for 25 min and were allowed to recover for another 25 min before fixation.

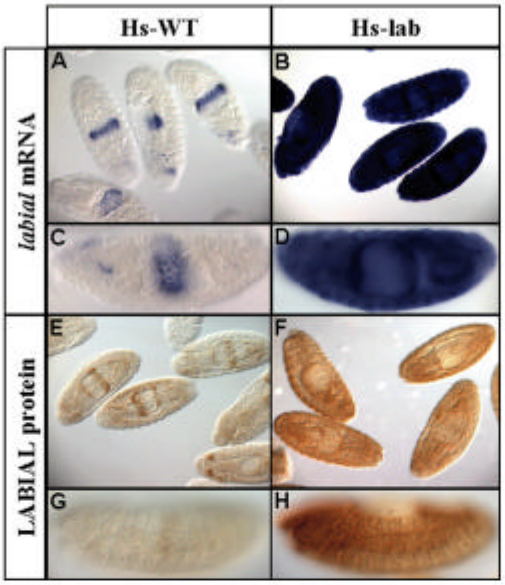

Hs-WT
Hs-lab
labial mRNA
LABIAL protein
A
B
C
D
E
F
G
H

Table 1. Genes that are differentially expressed following heat induced ubiquitous overexpression of *lab* in stage 10-17 hs-*lab* embryos, grouped according to functional classes. (N) Number of genes within a functional group present on the chip. (n) Number of genes differentially expressed within a functional group following *lab* overexpression. (n/N x 100) Number of differentially expressed genes within a functional class following *lab* overexpression, given as % of the total number of genes in this class present on the array. (down-regulated) Total number of genes within each functional class differentially down-regulated following *lab* overexpression. (up-regulated) Total number of genes within each functional class differentially up-regulated following *lab* overexpression. (*) The functional class Heat shock proteins was excluded from the analysis (see Materials and methods).

Table 1. Genes differentially expressed in response to *lab* overexpression

Functional class	Genes on the array (N)	Differentially expressed transcripts (n)	n/N x 100 (%)	down-regulated	up-regulated
Signal transduction	107	5	4.7	2	3
Transcriptional regulation	263	20	7.6	14	6
Cell cycle	37	5	13.5	0	5
Cytoskeleton/structural proteins	149	5	3.4	4	1
Metabolism	315	13	4.1	6	7
Translation	59	1	1.7	1	0
Heat shock proteins	18	*	*	*	*
Transcription/replication/repair	73	4	5.5	0	4
Proteolytic systems/apoptosis	62	12	19.4	1	11
Cell surface receptors/CAMs/ion channels	181	12	6.6	10	2
Transposable elements	35	4	11.4	3	1
Chromatin structure	36	4	11.1	2	2
RNA binding	59	7	11.9	2	5
Secreted proteins	34	2	5.9	2	0
Unknown function	85	2	2.4	1	1
	$\Sigma N = 1513$	$\Sigma n = 96$		48	48

Figure 2. Genes differentially expressed in response to heat shock induced overexpression of *lab*, grouped according to functional classes. Bars represent the fold-change between differentially expressed genes in heat shock treated wild type embryos and heat shocked hs-*lab* embryos. Positive values indicate that the relative expression level of a gene is increased (up-regulated) following *lab* overexpression and negative values indicate a decrease (down-regulated). Absolute Average Difference (Avg Diff) values are given for the *lab* overexpression condition as follows: white bars represent Avg Diff < 100, grey bars represent Avg Diff ranging from 100-1000, and black bars represent Avg Diff >1000.


Transcriptional regulation
lab labial 26.3
CG11971 CG11971 4.1
Gnf1 Germ line transcription factor 1 3.7
stc shuttle craft 2.7
wdn wings down 2.0
gro groucho 1.9
Eip75B Ecdysone-induced protein 75B -1.6
HLHm7 E(spl) region transcript m7 -1.7
pros prospero -1.8
Pdp1 PAR-domain protein 1 -1.9
Dll Distal-less -2.0
tup tailup -2.2
CrebA Cyclic-AMP response element binding protein A -2.3
HLHm3 E(spl) region transcript m3 -2.5
mirr mirror -2.6
Mef2 Myocyte enhancing factor 2 -2.7
vvl ventral veins lacking -2.8
scrt scratch -2.9
hkb huckebein -2.9
ab abrupt -4.1
Metabolism
SamDC S-adenosylmethionine decarboxylase 5.5
Cyp4e2 Cytochrome P450-4e2 2.9
awd abnormal wing discs 2.5
Tpi Triose phosphate isomerase 2.0
Pdk Pyruvate dehydrogenase kinase 1.9
Eno Enolase 1.9
Pp2A-29E Protein phosphatase 2A at 29B 1.5
Crc Calreticulin -1.7
Pi3K59F Phosphotidylinositol 3 kinase 59F -1.7
Atpα Na pump α subunit -1.8
Bc Black cells -1.9
Mlc-k Myosin light chain kinase -3.1
Pepck Phosphoenolpyruvate carboxykinase -3.6
Proteolytic systems/apoptosis
UbcD4 Ubiquitin conjugating enzyme 4 9.6
Uch Ubiquitin carboxy-terminal hydrolase 5.1
Pros28 Proteasome 28kD subunit 1 3.9
αTry αTrypsin 3.7
Pros35 Proteasome 35kD subunit 3.3
Prosb2 Proteasome b2 subunit 3.1
Fur2 Furin 2 2.9
ben bendless 2.6
Prosα7 Proteasome α7 subunit 2.0
faf fat facets 1.8
Dcp-1 Death caspase-1 1.5
Ubi-p63E Ubiquitin-63E -2.1
Cell surface receptors/CAMs/ion channels
tor torso 6.9
smo smoothened 1.7
drl derailed -1.7
fz2 frizzled 2 -1.7
Nrt Neurotactin -1.8
Nrx Neurexin -1.9
Gp150 Gp150 -1.9
Rya-r44F Ryanodine receptor 44F -2.0
Hem HEM-protein -2.1
rho rhomboid -2.3
18w 18 wheeler -3.1
trn tartan -4.8
RNA binding
BicC Bicaudal C 20.7
swa swallow 20.3
osk oskar 10.1
vas vasa 4.4
stau staufen 1.5
how held out wings -1.7
elav embryonic lethal abnormal vision -2.4
Signal transduction
pll pelle 3.9
Ras85D Ras oncogene at 85D 1.9
Btk29A Btk family kinase at 29A 1.6
Gtp-bp GTP-binding protein -1.6
Src42A Src oncogene at 42A -1.8
Cell cycle
twe twine 6.6
CycB Cyclin B 5.5
Pen Pendulin 4.2
polo polo 2.5
CycD Cyclin D 2.0
Cytoskeleton/structural proteins
AnnX Annexin X 2.7
fax failed axon connections -1.5
Klp64D Kinesin-like protein at 64D -1.8
LanB1 Laminin B1 -1.8
vkg viking -2.5
Transcription/replication/repair
mus309 mutagen-sensitive 309 3.5
TfIIFB Transcription factor IIFB 2.9
Mcm5 Minichromosome maintenance 5 2.5
TfIIFA Transcription factor IIFA 2.3
Transposable elements
R2-element R2-element 2.6
jockey jockey element ORF2 -1.6
3S18 3S18 element -3.4
R2-element R2 rDNA element -7.8
Chromatin structure
wapl wings apart-like 10.3
thr three rows 2.6
corto corto (CP-1) -2.3
corto corto (ccf) -4.4
Secreted proteins
gbb glass bottom boat -1.8
mspo M-spondin -3.2
Unknown function
Atu Another transcription unit
CG4844 CG4844
Translation
Nmda1 Aspartate receptor ass. protein -1.9
Avg Diff range:
< 100
100-1000
> 1000

Table 2. Comparison of change folds between oligonucleotide arrays and RT-PCR. RT-PCR was performed on cDNA derived from heat shocked wild type embryos and heat shocked hs-*lab* embryos. Change folds determined by RT-PCR are represented as the mean value of eight independent replicates, derived from two different cDNA preparations.

Table 2 - Comparison of change folds between oligonucleotide arrays and RT-PCR

	Avg Diff (array)		Change fold	
Gene	HS-wt	HS-lab	Array	RT-PCR
lab	41	1078	26.3	55.7
swa	20	406	20.3	18.4
UbcD4	44	423	9.6	6.5
twe	20	132	6.6	4.9
cycB	243	1344	5.5	4.6
Uch	61	312	5.1	12.1
sqd	373	370	1.0	1.1
scrt	225	79	-2.9	-3.7
Pepck	610	171	-3.6	-4.6

5. Evolutionary conservation of *otd/Otx2* transcription factor action: a genome-wide microarray analysis in *Drosophila*

Haiqiong Montalta-He, Ronny Leemans, Thomas Loop, Martin Strahm[1], Ulrich Certa[2], Dario Acampora[3], Antonio Simeone[3] and Heinrich Reichert.

Institute of Zoology, Biocenter/Pharmacenter, University of Basel, Basel, Switzerland.
[1]Roche Bioinformatics and [2]Genetics Pharmaceuticals Division, F.Hoffmann-La Roche Ltd., Basel, Switzerland
[3]MRC Center for Developmental Neurobiology, King's College London, UK

Genome Biology (2003) 3(4): RESEARCH0015.1-0015.15.

SUMMARY

The homeobox transcription factors encoded by the genes of the *orthodenticle (otd)/Otx* family play evolutionarily conserved roles in embryogenesis of the head and brain. Moreover, cross-phylum gene replacement experiments show that the *Drosophila otd* gene and orthologous mammalian *Otx* genes are functionally equivalent, in that overexpression of either gene in null mutants of *Drosophila* or mouse can lead to the restoration of defects in cephalic and brain development. These experiments suggest that *otd* and *Otx* genes can control a comparable subset of downstream target genes in either organism. Here we use quantitative transcript imaging to analyze this equivalence of *Drosophila otd* and human *Otx* gene action at a genomic level. Oligonucleotide arrays representing 13,400 annotated *Drosophila* genes were used for detection and quantification of differential gene expression in transgenic flies in which either the *Drosophila otd* gene or the human *Otx2* gene was overexpressed. 287 identified transcripts showed highly significant changes in expression levels in response to *otd* overexpression, and 682 identified transcripts showed highly significant changes in expression levels in response to *Otx2* overexpression. Among these, 93 transcripts were differentially expressed following overexpression of either *otd* or *Otx2*, and most of these were influenced in the same direction, either upregulated or downregulated. We postulate that these 93 transcripts are common downstream targets for the fly *otd* gene and for the human *Otx2* gene in *Drosophila*. Our experiments indicate that approximately one third of the *otd*-regulated transcripts in *Drosophila* can also be controlled by the human *Otx2* gene homolog. These common *otd*/*Otx2* downstream genes are likely to represent the molecular basis for the functional equivalence of *otd* and *Otx2* gene action in *Drosophila*.

Introduction

Studies on developmental control genes involved in anterior patterning have revealed a set of homologous genes encoding transcription factors that are required for the development of the head and brain in diverse animal phyla(Finkelstein and Boncinelli, 1994; Thor, 1995; Sharman and Brand, 1998; Holland and Holland, 1999; Galliot and Miller, 2000). A striking example for the evolutionary conservation of expression and function of such genes between invertebrates and vertebrates are the homeobox genes of the *orthodenticle* gene family, which includes the *Drosophila orthodenticle (otd)* and the murine *Otx1* and *Otx2* genes(Finkelstein and Perrimon, 1990; Simeone et al., 1992b; Simeone et al., 1993; Acampora et al., 1999). The *Drosophila otd* gene is expressed in the anterior region of the early embryo in a domain that includes the precursors of the procephalic regions of the head, and it is also expressed in anterior brain regions and in midline CNS structures(Cohen and Jurgens, 1990; Finkelstein and Perrimon, 1990; Wieschaus et al., 1992; Grossniklaus et al., 1994; Schmidt-Ott et al., 1994; Gao et al., 1996). Mutational inactivation of *otd* in *Drosophila* results in defects in head structures and in deletions in anterior parts of the brain as well as in ventral nerve cord defects(Finkelstein and Perrimon, 1990; Klambt et al., 1991; Wieschaus et al., 1992). The two *otd*-related genes in the mouse, *Otx1* and *Otx2*, are also expressed anteriorly in the embryo in nested domains that include the embryonic forebrain and midbrain(Simeone et al., 1992a). Mutational inactivation of these genes result in specific defects in the head and anterior CNS; *Otx2* null mice die early in development and fail in specification of the rostral neuroectoderm and proper gastrulation(Acampora et al., 1995; Matsuo et al., 1995; Ang et al., 1996; Suda et al., 1996). *Otx1* null mice are viable but show spontaneous epileptic seizures and abnormalities affecting the dorsal telencephalic cortex (Acampora et al., 1996).

In addition to the remarkable similarities in expression patterns and mutant phenotypes of the *otd*/*Otx* gene family, *in vivo* gene replacement experiments provide further evidence for conservation of functional properties (Acampora et al., 1998; Leuzinger et al., 1998; Nagao et al., 1998; Sharman and Brand, 1998). In these cross-phylum rescue experiments, human *Otx1* or *Otx2* genes were overexpressed in *Drosophila otd* mutants and, conversely, murine *Otx1* or *Otx2* genes were replaced with the *Drosophila otd* gene in the mouse. Human *Otx1* and *Otx2* genes were able to rescue the brain and cephalic defects in *Drosophila* although *Otx2* rescues

at a lower frequency than *otd*, and *Otx1* rescues less efficiently still(Leuzinger et al., 1998; Nagao et al., 1998). Similarly, the *Drosophila otd* gene coding sequence introduced into the mice *Otx1* locus was able to rescue most of the brain pattering defects in *Otx1* mouse mutants and, when provided with the appropriate *Otx2* posttranslational control elements, also in *Otx2* mouse mutants(Acampora et al., 1998; Boyl et al., 2001a).

Drosophila and vertebrate *otd*/*Otx* gene products share structural homology that is confined mainly to the homeodomain. The 60 amino acid residues of the fly *otd* homeodomain differ from the homeodomains of the human *Otx1* and *Otx2* protein in only three and two amino acids, respectively. It, thus, seems likely that most of the conserved functional action of the *otd*/*Otx* genes is mediated by the evolutionarily highly conserved homeodomain of the encoded transcription factor protein(Nagao et al., 1998; Acampora et al., 2001b). Given this highly conserved homeodomain, one might predict that the *in vivo* functional equivalence of *otd*/*Otx* genes, as demonstrated in the cross-phylum rescue experiments, is due to the fact that both *otd* and *Otx* genes can control a comparable set of downstream target genes irrespective of whether the *otd*/*Otx* genes are expressed in flies or in mammals(Acampora et al., 2001b). However, currently, little is known about the downstream targets of either *otd* or *Otx* genes in flies or in mammals, and no information on common targets of *otd* and *Otx* genes is available in any species context(Acampora et al., 2001b; Boncinelli and Morgan, 2001).

In order to address this issue in a comprehensive manner we have combined cross-phylum overexpression experiments with genome-wide expression analysis based on oligonucleotide arrays in order to identify the number and identity of downstream target genes that are affected by overexpression of *otd* and of human *Otx2* in *Drosophila*. For this we used transgenic flies, which carried either the fly *otd* gene or the human *Otx2* gene under the control of a heat-inducible promoter, to ubiquitously overexpress these transgenes in *Drosophila* embryos and then employed high-density oligonucleotide arrays representing the entire annotated fly genome for large-scale detection and quantification of induced gene expression(Lockhart et al., 1996; Lipshutz et al., 1999; Adams et al., 2000; Leemans et al., 2000; Leemans et al., 2001). These experiments identified 287 annotated genes that showed highly significant ($p=0.001$) changes in expression levels in response to *otd* overexpression in *Drosophila*. Among these genes, 93 also showed highly significant differential expression

changes in response to *Otx2* overexpression. Moreover the expression levels of 90 of these 93 genes were influenced in the same direction, either upregulated or downregulated, by *otd* and by *Otx2* overexpression. Thus, approximately one third of the candidate *otd* downstream target genes in *Drosophila* can be controlled in a comparable manner by the human *Otx2* gene homolog. From a genome-wide perspective, it is likely that the conserved genetic control of these common *otd*/*Otx2* downstream genes forms the molecular genetic basis for the striking *in vivo* functional equivalence of *otd* and *Otx* gene action in *Drosophila*.

Results

in vivo overexpression and microarray analysis

In this study, transgenic fly strains carrying the *otd* coding sequence or the human *Otx2* coding sequence under control of the heat inducible Hsp70 promoter were used(Leuzinger et al., 1998). Stage 10-17 embryos were given a 25-min heat pulse in order to overexpress the *otd* or *Otx2* genes and allowed to recover for 25 min. Ubiquitous overexpression of *otd* and *Otx2* was verified by whole mount in situ hybridization with *otd*- or *Otx2*-specific antisense RNA probes. These experiments demonstrated that RNA was strongly overexpressed 50 min after the onset of heat shock in these strains. Wild type control flies were subjected to the identical heat shock regime.

Following ubiquitous overexpression of *otd* or *Otx2,* transcript profiles were analyzed using a genome-wide high-density oligonucleotide array and compared to the transcript profiles of heat shock treated wild type control embryos. The transcripts represented on the oligonucleotide array correspond to probe sets that are complementary to approximately 13,400 annotated *Drosophila* genes according to Release 1.0 of the *Drosophila* genome(Adams et al., 2000). For each experimental condition, several replicates were carried out (see Materials and methods). An example of the primary data obtained in experimental replicates is shown in scatter plots for four experimental conditions in figure 1. A complete description of the microarray content as well as all primary data for obtained in each individual microarray experiment are given as supplementary data *(linked according to genome-biology)*

Overview of differentially expressed transcripts

An overview of the total number of transcripts that were differentially regulated following *otd* or *Otx2* overexpression is given in Table 1. Two levels of significance for the experimental data are considered in this overview. At a significance level of $p=0.001$, a total of 287 genes were found to be differentially regulated following *otd* overexpression, as compared to heat shocked wild type control embryos. This corresponds to 2.1% of the genes represented on the array. At a significance level of $p=0.01$, a total of 762 genes were found to be differentially regulated following *otd* overexpression as compared to heat shocked wild type control embryos. This corresponds to 5.7% of the genes represented on the array. In both cases,

approximately one fourth of the differentially regulated transcripts corresponded to known genes, and the rest corresponded to genes that are currently characterized only by sequence information and predicted function, CG-transcripts (CG: Celera Genomics)

Overexpression of the human *Otx2* gene in *Drosophila* embryos resulted in a larger number of differentially expressed transcripts than did overexpression of the *Drosophila otd* gene. Thus, at a significance level of p=0.001, a total of 682 genes were found to be differentially expressed following *Otx2* overexpression, as compared to heat shocked wild type control embryos. This corresponds to 5.1% of the genes represented on the array. At a significance level of p=0.01, 1395 genes were found to be differentially expressed following *Otx2* overexpression as compared to heat shocked wild type control embryos. This corresponds to 10.4% of the genes represented on the array. Again, in both cases, approximately one fourth of the differentially regulated transcripts corresponded to known genes, and the rest were CG-transcripts.

A subset of the transcripts found to be differentially regulated following *otd* overexpression were also differentially regulated following *Otx2* overexpression. Among the transcripts that were differentially expressed at the significance level of p=0.001, 93 transcripts were found to be differentially regulated following overexpression of either gene. This implies that 32% of the *otd*-regulated transcripts were also regulated by *Otx2*. Among the transcripts that were differentially expressed at the significance level of p=0.01, 351 transcripts were found to be differentially regulated following overexpression of either gene. This implies that 46% of the *otd*-regulated transcripts were also regulated by *Otx2*. In the following, only genes that were differentially expressed at the significance level of p=0.001 are considered further. We propose these genes to be potential direct or indirect downstream targets for the homeodomain transcription factors *otd* and *Otx2*.

Functional classification of differentially expressed transcripts

When ubiquitously expressed in the embryo, both *otd* and *Otx2* caused a significant transcriptional response of genes encoding a wide variety of functionally different gene products. A detailed classification of the *otd*- and *Otx2*-regulated transcripts into different functional classes was carried out according to Gene Ontology (GO) and is presented in table

2. (In the GO classification scheme, a given gene can be grouped into more than one functional class; (Ashburner et al., 2000)) The *otd*- and *Otx2*-regulated transcripts fall into 92 GO classes, but only about half of these classes are characterized by more than one regulated transcript.

In terms of known function, the two classes with the highest absolute and relative numbers of regulated transcripts were 'enzymes' and 'transcription factors'; this was the case for both *otd*-regulated and *Otx2*-regulated transcripts. Other functional classes with high numbers of differentially regulated genes were 'signal transduction', 'DNA binding', 'transporter', 'protein kinase', 'motor', 'ligand binding or carrier', and 'endopeptidase'; again this was the case for both *otd*- and *Otx2*-regulated transcripts. Indeed, in most cases in which a functional class was characterized by both *otd*- and *Otx2*-regulated transcripts, the relative number (n/M; see table 2) of *otd*-regulated transcripts was similar to that of *Otx2*-regulated transcripts. For example, 2.79% of the *otd*-regulated transcripts versus 2.20% of the *Otx2*-regulated transcripts were classified under 'cell adhesion', and 3.48% of the *otd*-regulated transcripts versus 3.67% of the *Otx2*-regulated transcripts were classified under 'signal transduction'. Approximately half of both the *otd*-regulated and the *Otx2*-regulated transcripts belong to the class 'function unknown'.

Quantitative profiling of differentially expressed transcripts

Figure 1 shows the *otd*-regulated transcripts that correspond to known *Drosophila* transcripts and presents a quantitative representation of the change in expression levels for these transcripts. For clarity, these transcripts are only grouped into mother classes and not into the detailed GO classes. Most of the 63 known transcripts that were differentially expressed following *otd* overexpression showed increased expression levels; less than 20% of these transcripts were downregulated. The gene with the highest increase in expression level (78-fold) was *otd* itself, in accordance with our experimental overexpression protocol. Increases in expression levels above 10-fold were also observed for *fc96Cb* which encodes a nuclear binding protein, for *ptc* which encodes a protein involved in signal transduction, for *picot*, which encodes a transporter, and for *cortactin* and *Rca1*, which encode gene products of currently unknown function. Only two transcripts showed increases in the 5-10-fold range, namely *sut1* encoding a protein involved in sugar transportation, and *scra* encoding an actin

binding protein. The majority of the upregulated transcripts had increases in the 2-5-fold range. The transcript with the most marked decrease in expression was *eyg*, encoding a transcription factor known to be involved in eye development.

Figure 2 shows the *Otx2*-regulated transcripts that correspond to known *Drosophila* genes and presents a quantitative representation of their expression level changes. Again, these transcripts are grouped into mother classes and not into detailed GO classes. As was the case for *otd* overexpression, most of the known transcripts that were differentially expressed following *Otx2* overexpression showed increased expression levels. For example, in the functional class of 'enzyme', 45 out of 49 transcripts were upregulated. In total less than 13% of the 184 *Otx2*-regulated known transcripts were downregulated. Increases in expression levels above 10-fold were observed for 23 genes and for 6 of these genes, *retn, SMC2, lic, Rtc1, H* and *dhd*, the increases were greater than 50-fold. 22 transcripts showed increases in the 5-10-fold range, and, similar to the *otd* overexpression situation, increases of 2-5-fold dominated in most of the functional classes. The transcript with the most marked decrease in expression was once again *eyg*.

Common candidate downstream genes of *otd* and *Otx2*

93 transcripts were differentially expressed both in response to *otd* overexpression and in response to *Otx2* overexpression. This indicates that approximately one third of the *otd*-regulated genes in *Drosophila* can also be controlled by the human *Otx2* gene homolog. Figure 3 shows the expression levels for these transcripts which are, thus, likely to represent the common downstream target genes for *otd* and *Otx2*. 21 of these transcripts correspond to known *Drosophila* genes and 72 correspond to annotated CG-transcripts. The expression levels of all of the known transcripts were influenced in the same manner by overexpression of *otd* and *Otx2*, in that a given downstream target gene was either upregulated in both cases or downregulated in both cases. Moreover, for most of these transcripts the absolute expression levels were similar in response both to *otd* and to *Otx2*. Two marked exceptions were *pim*, which was upregulated 12.4-fold following *Otx2* overexpression and 2.1-fold following *otd* overexpression, and *eyg*, which was downregulated 77.6-fold following *Otx2* overexpression (but see PCR data below) and downregulated 6.8 fold following *otd* overexpression. Similarly, the expression levels of 68 of the CG transcripts were influenced

in the same manner by overexpression of *otd* and *Otx2*. Only in the three remaining cases were transcripts upregulated by overexpression of one of the *otd*/*Otx* transgenes and downregulated by overexpression of the other. Thus, approximately one third of the candidate *otd* downstream target genes in *Drosophila* can be controlled in a comparable manner by the human *Otx2* gene homolog.

The four known transcripts in class 'ligand binding or carrier', *scra, Klp61F, alpha-Spec* and *Cen190*, are all involved in actin or microtubule (Heck et al., 1993; Lee et al., 1993; Field and Alberts, 1995; Barbosa et al., 2000). This finding is notable since one of the *Otx2* downstream genes identified in the mouse is a tropomyosin gene, which also encodes actin binding protein(Zakin et al., 2000). Among the four known transcripts in the class 'nucleic acid binding' are the genes *Mcm7* and *Su(var)205* (Wakimoto, 1998; Feger, 1999) which encode chromatin binding proteins and the genes *eyg* and *HLH54F* which encode transcription factors(Treisman, 1999; Ledent and Vervoort, 2001). The four known transcripts in the functional class 'enzymes' are *LysD, cdc2, Rpd3,* and *BcDNA:LD08534(Regel et al., 1998; Su and O'Farrell, 1998; Wallrath, 1998; 1999).* Although the *cdc2* gene product is classified as 'enzyme', it also acts at the G2/M transition of mitotic cell cycle (Su and O'Farrell, 1998). Moreover, *Rpd3* encodes a histone deacetylase which is involved in the chromatin structure(Wallrath, 1998). In the class "transporter" the SNAP receptor encoding *n-syb* gene is involved in synaptic vesicle docking and fusion and is expressed in the embryonic CNS(Saitoe et al., 2001). In the class 'signal transducer', the gene *EG: 30B8.6* encodes a putative GABA-B receptor(Benos et al., 2001). Finally, the gene *Sd* classified as 'enzyme regulator' encodes a RAN GTPase activator(Crow, 1999). Among the transcripts of known genes are several genes, whose precise functional role is not well defined. These are the *Bx34* and *MRG15* genes(Zimowska et al., 1997; Bertram and Pereira-Smith, 2001) which encode components of the nucleus and the *gluon*, *Bub3* and *pim* genes which are all involved in mitosis. *gluon* encodes a putative component of the condensin, and *gluon* mutants show PNS defects during embryogenesis (Prokopenko et al., 2000). The gene product of *Bub3* is localized to the kinetochore and may function in the mitotic check point(Dobie et al., 1999). *pim (pimples)* is expressed in the embryonic CNS and encodes a protein implicated in mitotic sister chromatid separation(Philip, 1998).

Verification of microarray expression data with RT-PCR

To confirm the differences in gene expression levels after heat-shock induced overexpression of *otd* and human *Otx2* as compared to heat-shocked wild type embryos, quantitative RT-PCR was performed on selected candidate target genes. Changes in expression levels were determined for eight genes that were differentially regulated by *otd* or human *Otx2*, namely *scra, LysD, glu, Rpd3, pim, n-syb, eyg* and *otd*. The genes *wun* and *Scc1*, whose expression levels remained unchanged in response to *otd* or *Otx2* overexpression, served as controls. As indicated in Table 3, these experiments showed that the changes in relative expression level, as measured by RT-PCR, are generally consistent with the data obtained with the oligonucleotide arrays. An exception is data on the response of the *eyg* gene to *Otx2* overexpression; RT-PCR data indicate a weak downregulation (-1.62) whereas array data indicate a strong downregulation (-77.6).

Discussion

Common downstream target genes for *otd* and *Otx*

Cross-phylum gene replacement experiments have shown that the fly *otd* gene and the homologous human *Otx* genes are functionally equivalent *in vivo*, in that overexpression of either gene in *Drosophila otd* null mutants can lead to the restoration of defects in cephalic and brain development(Acampora et al., 1998; Leuzinger et al., 1998; Nagao et al., 1998; Boyl et al., 2001a). We have used a combination of transgenic overexpression genetics and functional genomics to gain insight into the equivalence of *otd* and *Otx* gene expression in *Drosophila* at a comprehensive, genome-wide level. Using inducible overexpression and quantitative transcript imaging through oligonucleotide arrays representing the total number of 13,400 annotated *Drosophila* genes, we have identified hundreds of candidate downstream genes both for the fly *otd* gene and for the human *Otx2* gene. A comparison of these candidate downstream genes reveals that both *otd* and *Otx* genes can control an overlapping set of genes; we refer to these genes as common downstream genes. The number of identified common downstream genes for *otd* and *Otx2* depends on the statistical level of significance used to determine if a given gene showed differential expression in response to transgene overexpression. If the analysis is restricted to highly significant (p=0.001) data sets, we find 93 common downstream genes, equivalent to 32% of the candidate *otd* downstream genes or

approximately 1% of transcripts in the annotated fly genome. If, in contrast, the analysis is based on significant (p=0.01) data sets, we find 351 common downstream genes, equivalent to 46% of the candidate *otd* downstream genes or approximately 3% of transcripts in the annotated fly genome. In either case, a substantial, but far from complete, set of the *otd* regulated genes are common downstream targets of both fly and human transgenes.

On one hand, it is interesting that, at the genome-wide transcript level, the *Otx2* gene does not appear to be able to replace *otd* action in full; over half of the transcripts that are influenced by *otd* overexpression are not influenced by *Otx2* overexpression. Given the pronounced differences in molecular sequence of the *otd* and *Otx2* gene products, this may not be altogether surprising. The OTD protein is 548 amino acids long, the OTX2 protein is 289 amino acids long; shared homology is restricted to the homeodomain and to a short domain immediately upstream of the homeodomain as well as a tripeptide at the amino terminus(Nagao et al., 1998). Moreover, since *Otx* genes cannot completely replace the *otd* genes in cross-phylum rescue experiments *in vivo*, a complete correspondence of *otd* downstream genes and common *otd*/*Otx* downstream genes might not be expected (Leuzinger et al., 1998; Nagao et al., 1998; Sharman and Brand, 1998). On the other hand, it is equally interesting that one third (to one half) of the *otd*-regulated genes can indeed be regulated by *Otx2* overexpression. This is because these common downstream genes are likely to represent the molecular substrate for the functional equivalence of the *otd*/*Otx* genes in cross-phylum rescue experiments *in vivo*. In this sense these genes are a direct manifestation of the evolutionarily conserved action of the members of the *otd*/*Otx* gene family. Although this analysis does identify common downstream targets of *otd* and *Otx2* in *Drosophila*, it does not demonstrate that the genetic circuitry that is downstream of *otd* and *Otx2* genes is evolutionarily conserved. To investigate this, it will now be important to carry out similar functional genomic analyses of *otd* and *Otx* gene action in a mammalian system such as the mouse (Acampora et al., 2001b).

otd overexpression: a genomic perspective of candidate downstream genes

The experiments reported here identify approximately 300 genes that showed highly significant (p=0.001) changes in expression levels in response to *otd* overexpression in *Drosophila*. The genomic perspective of these identified *otd* downstream target genes reveals

several features of *otd* action at a novel level of insight. First, this finding indicates that the *otd* gene product, a homeodomain transcription factor, regulates a limited and distinct set of candidate downstream genes. At a significance level of $p<0.001$, 287 genes were found to be differentially regulated corresponding to approximately 2.1% of the transcripts in the annotated fly genome. At a significance level of $p<0.01$, 762 genes were found to be differentially regulated corresponding to approximately 5.7% of the transcripts in the annotated fly genome. This is further evidence for the notion, that homeoproteins in *Drosophila* control only a subset and not the majority of genes in the genome (Leemans et al., 2001). Indeed, in similar experiments in which the homeobox gene *labial* was overexpressed using the same heat shock regime as described here, 6.4% of the genes represented on the array used were shown to be differentially regulated at a significance level of $p<0.01$ (**Leemans et al., 2000).** (It should however, be noted that the array used in these *labial* overexpression experiments represent only 10% of the genes in the fly genome.) Thus the number of putative *otd* targets is in the same range as the number of putative *labial* targets.

Second, these experiments show that the OTD homeodomain transcription factor acts on numerous candidate target genes that also encode transcription factors, consistent with the idea that homeodomain proteins act through a cascade of transcription factors which regulate the expression of their own subset of downstream genes (Kablar et al., 1996). Currently, we do not know which of the downstream target genes are direct OTD targets and are, thus regulated directly by OTD protein binding to DNA regulatory sequences, and which are indirect targets. At present, little is known about temporal response of putative target genes following pulsed expression of a transcription factor. Some studies have been carried out, based on the assumption that direct targets respond immediately while indirect targets respond with a delay due to the time required for intermediary gene expression. Nasiadka and Krause used a kinetic approach to identify direct and indirect targets of the ectopically expressed homeodomain transcription factor Fushi tarazu (Ftz)(Nasiadka et al., 2000). Their results show that target genes respond to pulses of Ftz expression within two distinct temporal windows. Direct responses (no intermediary gene transcription is required) are 50% complete within about 18 minutes post heat shock. Indirect responses do not reach the same level of response until 26 min post heat shock. Under the assumption that OTD expression follows a similar kinetic profile as Ftz, that would mean that , under our experimental conditions, we

would detect primary targets as well as later responding genes (requiring intermediate transcription).

Third, these results show that the principle type of action of *otd* overexpression in *Drosophila* is upregulation of downstream target genes. Indeed more than 80% of the genes that were differentially expressed following *otd* overexpression showed increased expression levels. This contrasts with the action of the homeotic gene *lab*; overexpression of *lab* under comparable conditions resulted in approximately equal number of upregulated and downregulated target genes (Leemans et al., 2001).

The majority of potential downstream target genes of otd are annotated CG transcripts and, hence, correspond to predicted genes which have not yet been studied in detail in *in vivo* context. This is surprising given the fact that numerous classical genetic screens for genes involved in cephalic and CNS embryogenesis have been carried out **(see Bate and Martinez-Arias, 1993)**. This may indicate that many of the genes involved in those aspects of cephalic and CNS embryogenesis that are under the control of *otd* in *Drosophila* have not yet been studied adequately. Alternatively, this finding may reflect specific constraints of the overexpression experiment. For example, the overexpression protocol used makes it difficult to control the level of the OTD protein as well as the dynamic stability of this protein. As different levels of a homeoprotein can have different developmental consequences, the relatively high level of OTD protein attained may influence target genes that are not affected by the endogenously attained protein level (Hoppler and Bienz, 1994; Cribbs et al., 1995). Moreover, the fact that *otd* overexpression is not accompanied by simultaneous overexpression of cofactors, which can act together with homeodomain transcription factors to determine their *in vivo* target specificity, may also lead to unspecific activation of target genes(Hartmann and Reichert, 1998).

Functional genomics of a human transgene overexpressed in *Drosophila*

In several cases, human transgenes have been overexpressed in *Drosophila* in order to gain insight into the evolutionary conservation of developmental control gene action(Leuzinger et al., 1998; Nagao et al., 1998; Parkes et al., 1998; Hartmann et al., 2000; Gunawardena and

Goldstein, 2001; Yang et al., 2001). This has also been the primary goal of the overexpression of human *Otx2* in *Drosophila* carried out in this report. In addition to the identification of common *otd*/*Otx* downstream genes, the genomic level of analysis carried out here has uncovered remarkable similarities and differences in the action of the human transgene in the fly versus that of its fly homolog. In terms of similarities, most of the genes affected by *Otx2* overexpression were upregulated, as was the case for *otd* overexpression. Moreover the same relative distribution of classes of downstream genes was observed in both cases. For example, the classes 'enzymes' and 'transcription factors' had the highest absolute and relative transcript number. The striking difference in the action of the two transgenes is that overexpression of human *Otx2* causes expression changes in much more downstream genes than does overexpression of the fly *otd* gene.

The experiments reported here identify approximately 700 genes that showed highly significant (p=0.001) changes in expression levels in response to *Otx2*; this is over two times more than that observed in response to *otd*. It is unlikely that this difference is due to corresponding differences in the expression levels attained for *Otx2* versus *otd* transcripts. Indeed the transcript abundance of *otd* was higher than that of *Otx2* in these experiments (see methods). Nevertheless, these data should be interpreted with caution for several reasons. First, only one single transgenic strain of *otd* and and only one single transgenic strain of *Otx2* were used. Thus, strain differences or positional dependent transgene insertional effects might account for the fact that *Otx2* controls more downstream genes than *otd* in *Drosophila*. Second, it is conceivable, that overexpression of the *Otx2* gene affects many more downstream genes in *Drosophila* than *otd* because the OTX2 transcription factor binds to many more DNA regulatory regions than does OTD. The smaller OTX2 protein might, therefore, have a lower specificity for target gene regulatory regions. Indeed, in vitro studies suggest that the DNA binding specificity of homeoproteins may be low (publication). Third, the OTX2 protein might be more promiscuous than OTD in its interactions with the numerous cofactors that determine target specificity. Fourth, the OTX2 product could influence the expression of a small number of transcription factors, which are not affected by OTD and which then regulate the expression of their own subset of downstream genes. Whatever the molecular basis for this unexpected difference in the result of *Otx2* versus *otd* overexpression

may be, its discovery is a further demonstration of the novel level of insight that can be attained from a genome-wide functional perspective.

Materials and methods

Embryos

The wild type was *Drosophila melanogaster* Oregon-R. For overexpression of *otd*, we used the *hsp-otd* line 5A generated by (Royet and Finkelstein, 1995). For overexpression of human *Otx2*, we used the *hsp-Otx2* line generated by (Leuzinger et al., 1998). All fly stocks were kept on standard cornmeal/yeast/agar medium at 25°C. Embryos were collected overnight for 12 hours on grape juice plates, further kept for 4 hours at 25 °C and then subjected to a 37 °C heat shock for 25 min, followed by a recovery period of 25 min at 25 °C before RNA isolation. Therefore, at the time of RNA isolation these embryos were at embryonic stages 10-17(Leemans et al., 2000). Embryos younger than embryonic stage 10 were not used, since heat shock in these earlier stages results in lethality(Walter et al., 1990). Embryos used for in situ hybridization studies were collected and heat shock treated in the same way.

Whole mount in situ hybridization

For in situ hybridization, digoxigenin-labeled sense and antisense *lab* RNA probes were generated *in vitro*, with a DIG labeling kit (Roche Diagnostics) and hybridized to whole mount embryos following standard procedures (Tautz and Pfeifle, 1989). Hybridized transcripts were detected with an alkaline phosphatase conjugated anti-digoxigenin Fab fragment (Roche Diagnostics) using Nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (Sigma) as chromogenic substrates.

High density oligonucleotide arrays and hybridization

In this study, a custom-designed *Drosophila* oligonucleotide array (roDROMEGAa, Affymetrix, Santa Clara, CA) was used. It contains 14,090 sequences representing *Drosophila* specific transcripts, prokaryotic control sequences and custom chosen sequences for transgenes such as *gal4*, *gfp, and lacZ.* 13,998 sequences correspond to *Drosophila* specific transcripts that were annotated by Celera Genome Release NO 1 (Adams et al., 2000) and deposited in SWISS-PROT/TrEMBL databases. These 13,998 sequences represent approximately 13,400 genes in the *Drosophila* genome and therefore some genes are

represented by more than one probe set. Each sequence is represented on the array by a set of 14 oligonucleotide probes (25-mers) matching the sequence. To control the specificity of hybridization, the same probes are represented on the array with a single nucleotide mismatch in a central position. As such, each sequence is represented by 14 perfect match and 14 mismatch probes. The Average Difference (Avg Diff) between the perfect match hybridization signal and the mismatch signal is proportional to the abundance of a given transcript (Lipshutz et al., 1999). RNA was isolated, labeled and hybridized to the arrays as described (Leemans et al., 2000; Leemans et al., 2001) with minor modifications.

Data Analysis

Probe arrays were scanned with a commercial confocal laser scanner (Hewlett-Packard). Pixel intensities were measured, and expression signals were analyzed with commercial software (GENECHIP 3.1, Affymetrix). Data processing was carried out using RACE-A (F. Hoffmann-La Roche), Access 97 and Excel 97 (Microsoft) software. Scatter plots were prepared using GeneSpring TM software (version 4.1; Silicon Genetics, CA.). For quantification of relative transcript abundance, the Average Difference value (Avg Diff) was used(Lipshutz et al., 1999). Four replicates were performed for hs*otd* and hs*Otx2*. Three and five replicates were performed for hs*wt* and *wt* respectively. All arrays were normalized against the mean of the total sums of Avg Diff values across all 16 arrays. In order to avoid huge fold changes (FC), genes with a normalized Avg Diff below 20 were automatically assigned an Avg Diff of 20 (RACE-A protocol). An unpaired t-test for each individual gene was performed for the following pairwise comparisons: hswt vs. wt, hswt vs. hsotd, and hswt vs.hsOTX2. For differential transcript imaging, only transcripts that had highly significant or significant changes in Avg Diff ($p \leq 0.001/0.01$) and whose changes were in the 2-fold and above range are presented. Additionally, the higher mean Avg Diff of a pairwise comparison for a given transcript had to be above or equal to 50. For a comprehensive list of all genes with fold changes and significance level, see supplements (Details depend on the editorial preferences).

Reverse Transcription Polymerase Chain Reaction

300 ng poly(A)$^+$ RNA was isolated (mRNA isolation kit; Roche Diagnostics) and reverse transcribed with AMV-RT and random hexamers (RT-PCR kit; Roche Diagnostics). PCR

was performed with 100 pg template DNA and gene specific primers (Seq Web, Winsconsin Package Version 10.0, GCG) on a light cycler (LightCycler, Roche Diagnostics). Continuous fluorescence observation of amplifying DNA was possible using SYBR Green I (Roche Diagnostics). After cycling, a melting curve was produced by slow denaturation of the PCR end products, to validate the specificity of amplification. To compare the relative amounts of PCR products we monitored the amplification profile on a graph, displaying the log of the fluorescence against the number of cycles. Relative change folds for a given gene under both conditions (hs*otd* vs. hs*wt* or hs*Otx2* vs. hs*wt*) were calculated using the fit point method (Light Cycler Manufacturer, Roche).

Quantification of *otd* and human *Otx2* transcripts by RT-PCR

Plasmids containing fly *otd* or human *Otx2* cDNA were linearized with appropriate restriction enzymes and purified. The concentrations of the linearized plasmids were spectrophotometrically quantified using a GeneQuant RNA/DNA calculator (Pharmacia Biotech) and serial dilutions were made. To quantify the concentration of the *otd* and *Otx2* transcripts from heat shocked hs*otd* and hs*Otx2* embryos, standard curve was established using the serial dilution of the corresponding linearized plasmid on a light cycler (LightCycler, Roche Diagnostics). RT-PCR was performed when the standard curve was established. Thereafter, the steady state concentrations of the *otd* and human *Otx2* were calculated in relation to their standard curves, using the second derivative maximum method (Light Cycler Manufacturer, Roche). This showed that the concentrations of *otd* and *Otx2* transcripts were 1.5×10^{-6} µg/µl and 3.6×10^{-7} µg/µl, respectively.

Acknowledgements

We thank Jan Mous, Adrian Roth, Michel Tessier, Monika Seiler, and Reto Brem for essential contributions and helpful advice. We are particularly grateful to Clemens Broger and Martin Neeb (F. Hoffman-La Roche) for allowing us to use their RACE-A CHIP analysis software and to Volker Schmid and Natalie Yanze for help with the light cycler. We thank all members of our laboratory, especially Boris Egger, Frank Hirth and Martin Mueller for constant and encouraging discussions, and Lars Kammermeier for help with the photography. This research was supported by grants from the SNSF and by F. Hoffmann-La Roche.

Figure 1. Normalized Avg Diff of one pair of replicate arrays of each experimental condition in a log scale. A. hs*Otx2*; B. hs*otd*; C. hs*wt*; D. *wt*. Only probe sets with positive values in both arrays are used. The central line is $y = x$, and the flanking lines indicate the difference of a factor of two.

Fig.1 Log(base 10) expression indexes (normalized Avg Diff) of one pair of replicate arrays of each experimental condition

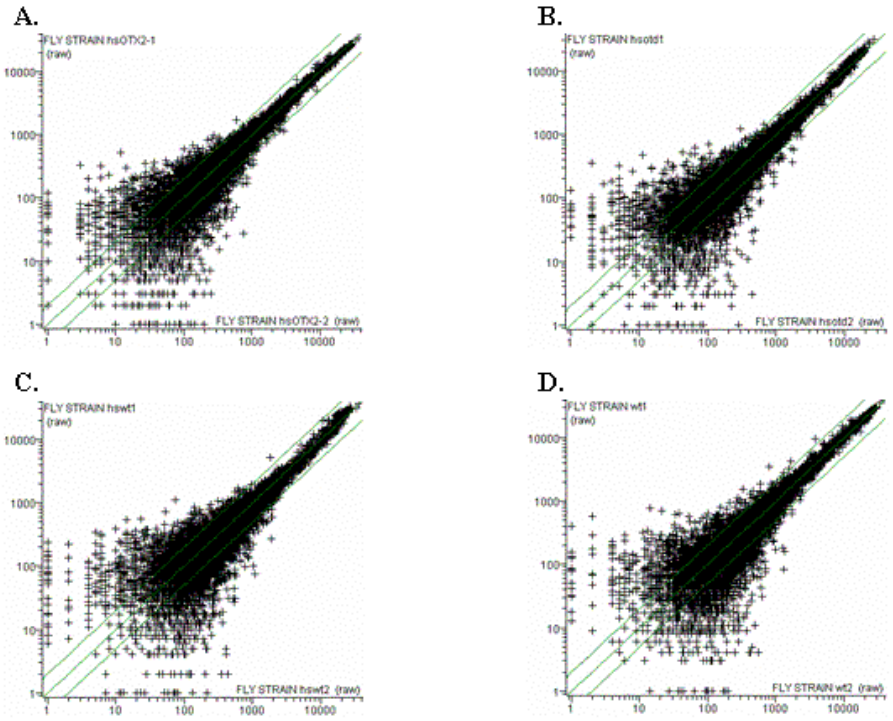


Table 1. Numbers of transcripts differentially regulated by HS-*otd* or HS-*Otx2*

Overview of the numbers of transcripts that were differentially expressed following overexpression of *otd* or human *Otx2*. A. Number of transcripts that were differentially expressed at the significance level of p=0.001. B. Number of transcripts that were differentially expressed at the significance level of p=0.01.

Table 1. Numbers of transcripts differentially regulated by HS-*otd* or HS-*Otx2*

A.

Differentially expression in response to	Total	Named transcripts	CG transcripts
HS-otd	287	63	224
HS-*Otx2*	682	184	498
HS-otd and HS-*Otx2*	93	21	72

B.

Differentially expression in response to	Total	Named transcripts	CG transcripts
HS-otd	762	165	597
HS-*Otx2*	1395	331	1064
HS-otd and HS-*Otx2*	351	69	282

Table 2. Classification of Transcripts differentially expressed in response to *Otx2* and *otd* overexpression

Genes that were differentially expressed following ubiquitous overexpression of *otd* or human *Otx2*, grouped according to Gene Ontology (GO) functional classes. (n) Number of transcripts detected that belong to an individual class. (N) Number of the transcripts represented on the chip for each functional class; the value of N for each for each functional class is given in the parenthesis following the class name. (n/N x 100) Percentage of transcripts that were differentially regulated for each functional class relative to the total number of transcripts in that class represented on the chip. (M) Total number of differentially expressed transcripts (of all classes) following overexpression of *otd* or human *Otx2* ($p = 0.001$); for *otd* and *Otx2*, M is 287 and 682 respectively. (n/M x 100) Percentage of transcripts that were differentially regulated in each functional class relative to the the total number of differentially regulated transcripts for *otd* and *Otx2*.

Functional class	n^{stu}	n^{stu}/N (%)	n^{stu}/M (%)	n^{Dlx2}	n^{Dlx2}/N (%)	n^{Dlx2}/M (%)
Function unknown (7108)	143	2.01	49.83	311	4.38	45.60
Enzyme (1872)	34	1.82	11.85	88	4.70	12.90
Transcription factor (940)	23	2.45	8.01	69	7.34	10.12
Signal transduction (462)	17	3.68	5.92	24	5.19	3.52
DNA binding (306)	14	4.58	4.88	27	8.82	3.96
Transporter (498)	12	2.41	4.18	19	3.82	2.79
Motor (406)	11	2.71	3.83	22	5.42	3.23
Protein kinase (365)	10	2.74	3.48	25	6.85	3.67
Ligand binding or carrier (581)	9	1.55	3.14	28	4.82	4.11
Endopeptidase (413)	8	1.94	2.79	25	6.05	3.67
Nucleic acid binding (369)	8	2.17	2.79	21	5.69	3.08
Cell adhesion (328)	8	2.44	2.79	15	4.57	2.20
Structural protein (335)	7	2.09	2.44	18	5.37	2.64
Actin binding (157)	6	3.82	2.09	10	6.37	1.47
RNA binding (292)	4	1.37	1.39	13	4.45	1.91
Transmembrane receptor (251)	4	1.59	1.39	9	3.59	1.32
Chaperone (195)	3	1.54	1.05	14	7.18	2.05
Cell cycle regulator (190)	3	1.58	1.05	12	6.32	1.76
Ion channel (214)	3	1.40	1.05	7	3.27	1.03
Protein phosphatase (91)	3	3.30	1.05	6	6.59	0.88
DNA repair protein (65)	3	4.62	1.05	4	6.15	0.59
Transcription factor binding (64)	2	3.13	0.70	11	17.19	1.61
Cytoskeletal structural protein (121)	2	1.65	0.70	6	4.96	0.88
DNA replication factor (42)	2	4.76	0.70	5	11.90	0.73
Defense/immunity protein (64)	2	3.13	0.70	4	6.25	0.59
G-protein linked receptor (103)	2	1.94	0.70	3	2.91	0.44
Receptor (97)	2	2.06	0.70	2	2.06	0.29
Cytochrome P450	2	14.29	0.70	0	0	0
Storage protein (25)	1	4.00	0.35	3	12.00	0.44
Peptidase (97)	1	1.03	0.35	3	3.09	0.44
Lysozyme (8)	1	12.50	0.35	2	25.00	0.29
Cyclin-dependent protein kinase (11)	1	9.09	0.35	2	18.18	0.29
GABA-B receptor (1)	1	100.00	0.35	1	100.00	0.15
Enzyme inhibitor (121)	1	0.83	0.35	1	0.83	0.15
Ecdysteroid hormone receptor (2)	1	50.00	0.35	0	0	0
3',5'-cyclic-nucleotide phosphodiesterase (1)	1	100.00	0.35	0	0	0
FK506 binding (2)	1	50.00	0.35	0	0	0
Peptidylprolyl isomerase (3)	1	33.33	0.35	0	0	0
Neurotransmitter transporter (29)	1	3.45	0.35	0	0	0
Steroid hormone receptor (16)	1	6.25	0.35	0	0	0
Acid phosphatase (5)	1	20.00	0.35	0	0	0
Arginine-tRNA ligase (2)	1	50.00	0.35	0	0	0
Carboxypeptidase (1)	1	100.00	0.35	0	0	0
Caspase activator(1)	1	100.00	0.35	0	0	0
Protein tyrosine phosphatase (9)	0	0.00	0.00	4	44.44	0.59
Protein serine/threonine kinase (43)	0	0.00	0.00	4	9.30	0.59
Chromatin binding (16)	0	0.00	0.00	4	25.00	0.59
Ubiquitin conjugating enzyme (12)	0	0.00	0.00	3	25.00	0.44
Structural protein of ribosome (136)	0	0.00	0.00	3	2.21	0.44
Casein kinase I (6)	0	0.00	0.00	3	50.00	0.44
Calcium binding (18)	0	0.00	0.00	3	16.67	0.44
Ubiquitin (14)	0	0.00	0.00	2	14.29	0.29
Translation factor (70)	0	0.00	0.00	2	2.86	0.29
Transcription co-repressor (3)	0	0.00	0.00	2	66.67	0.29
GTP binding (14)	0	0.00	0.00	2	14.29	0.29
Glutathione transferase (7)	0	0.00	0.00	2	28.57	0.29
Furin (2)	0	0.00	0.00	2	100.00	0.29
Electron transfer (35)	0	0.00	0.00	2	5.71	0.29

Functional class	n^{stu}	n^{stu}/N (%)	n^{stu}/M (%)	n^{Dlx2}	n^{Dlx2}/N (%)	n^{Dlx2}/M (%)
Ubiquitinyl hydrolase 1 (2)	0	0.00	0.00	1	50.00	0.15
Ubiquitin-specific protease (5)	0	0.00	0.00	1	20.00	0.15
Ubiquitin-like conjugating enzyme (1)	0	0.00	0.00	1	100.00	0.15
Tubulin-tyrosine ligase (7)	0	0.00	0.00	1	14.29	0.15
Transmembrane receptor protein tyrosine phosphatase (4)	0	0.00	0.00	1	25.00	0.15
Transmembrane receptor protein tyrosine kinase (7)	0	0.00	0.00	1	14.29	0.15
Transcription factor, cytoplasmic sequestering (1)	0	0.00	0.00	1	100.00	0.15
Transcription co-activator (2)	0	0.00	0.00	1	50.00	0.15
Thioredoxin (4)	0	0.00	0.00	1	25.00	0.15
Spermidine synthase (1)	0	0.00	0.00	1	100.00	0.15
SNF1A/AMP-activated protein kinase (1)	0	0.00	0.00	1	100.00	0.15
SH3/SH2 adaptor protein (2)	0	0.00	0.00	1	50.00	0.15
Sarcosine oxidase (2)	0	0.00	0.00	1	50.00	0.15
Ribulose-phosphate 3-epimerase (1)	0	0.00	0.00	1	100.00	0.15
Receptor signalling protein tyrosine phosphatase (1)	0	0.00	0.00	1	100.00	0.15
Protein tagging (2)	0	0.00	0.00	1	50.00	0.15
Prenylated protein tyrosine phosphatase (1)	0	0.00	0.00	1	100.00	0.15
Phosphoserine phosphatase (1)	0	0.00	0.00	1	100.00	0.15
Multicatalytic endopeptidase (4)	0	0.00	0.00	1	25.00	0.15
mRNA (guanine-N7)-methyltransferase (1)	0	0.00	0.00	1	100.00	0.15
Mitochondrial processing peptidase(1)	0	0.00	0.00	1	100.00	0.15
MAP kinase kinase (3)	0	0.00	0.00	1	33.33	0.15
Inositol-1,4,5-triphosphate receptor (1)	0	0.00	0.00	1	100.00	0.15
Electron transfer flavoprotein (1)	0	0.00	0.00	1	100.00	0.15
Effector caspase (3)	0	0.00	0.00	1	33.33	0.15
DNA-directed RNA polymerase III (7)	0	0.00	0.00	1	14.29	0.15
Cyclin (5)	0	0.00	0.00	1	20.00	0.15
CDP-diacylglycerol-serine O-phosphatidyltransferase (1)	0	0.00	0.00	1	100.00	0.15
Caspase (5)	0	0.00	0.00	1	20.00	0.15
cAMP-dependent protein kinase regulator (1)	0	0.00	0.00	1	100.00	0.15
cAMP-dependent protein kinase catalyst (3)	0	0.00	0.00	1	33.33	0.15
cAMP-dependent protein kinase (1)	0	0.00	0.00	1	100.00	0.15
Amine oxidase (flavin-containing) (7)	0	0.00	0.00	1	14.29	0.15
3-oxo-5-alpha-steroid 4-dehydrogenase (1)	0	0.00	0.00	1	100.00	0.15

Figure 2. Known transcripts differentially expressed in response to overexpression of *Otx2*, grouped according to functional classes. Bars represent the fold change between differentially expressed transcripts in hs*wt* embryos and hs*Otx2* embryos. Positive values indicate that the relative expression level of a gene is increased (up-regulated) following *Otx2* overexpression and negative values indicate a decrease (down-regulated). Absolute Average Difference (Avg Diff) values are given for the *Otx2* overexpression condition as follows: white bars represent Avg Diff < 100, gray bars represent Avg Diff ranging from 100-1000, and black bars represent Avg Diff >1000.

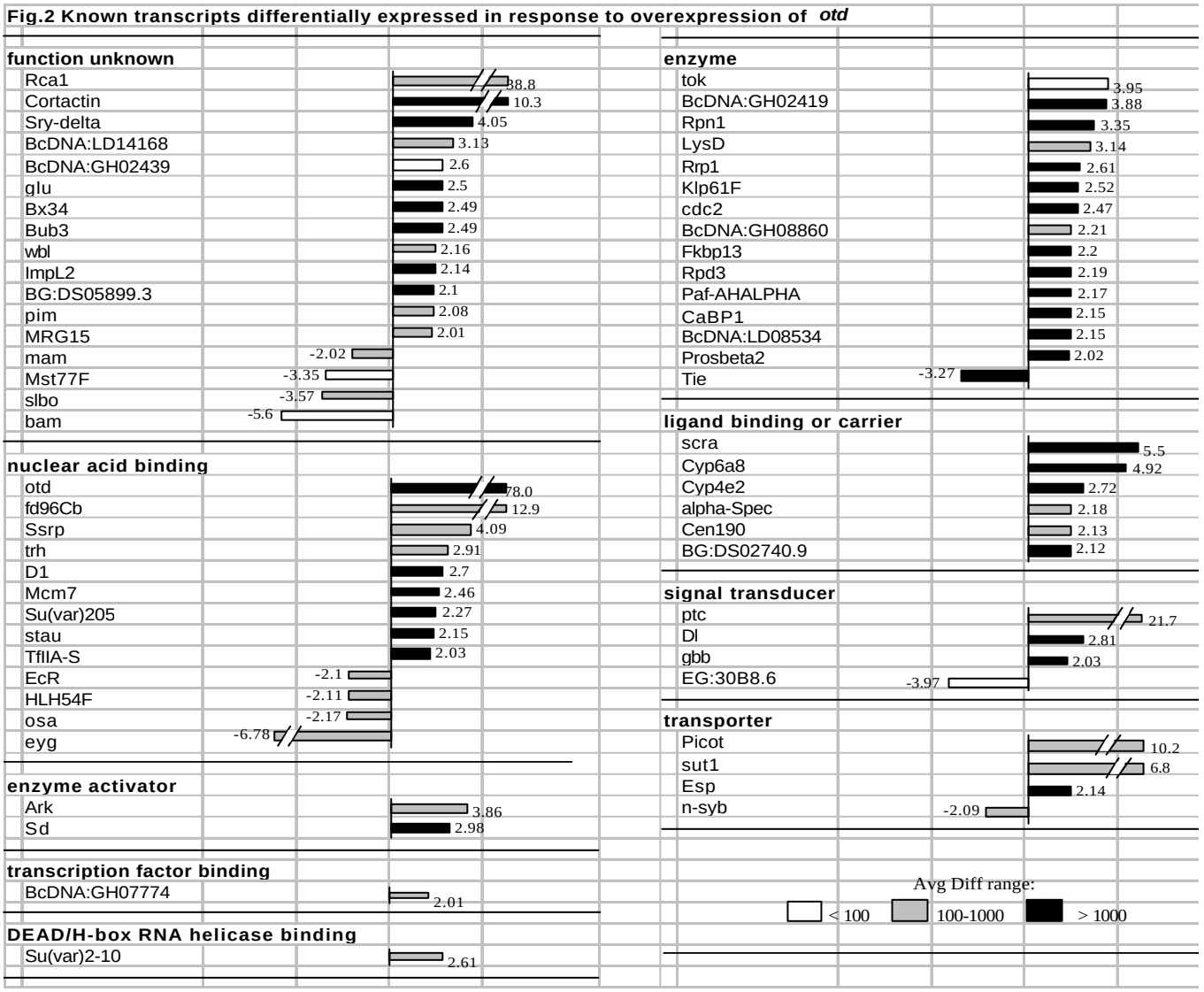


Figure 3 Transcripts differentially expressed in response to overexpression of *otd* and in response to overexpression of human *Otx2*, grouped according to functional classes. Bars represent the fold change between differentially expressed transcripts in hs*wt* embryos and hs*otd* or hs*Otx2* embryos. The upper bars represent the fold change of differentially expressed transcripts following overexpression of *Otx2* and the lower bars represent the fold change of differentially expressed transcripts following overexpression of *otd*. Positive values indicate that the relative expression level of a gene is increased (up-regulated) following *otd* overexpression and negative values indicate a decrease (down-regulated). Absolute Average Difference (Avg Diff) values are given for the *otd* overexpression condition as follows: white bars represent Avg Diff < 100, gray bars represent Avg Diff ranging from 100-1000, and black bars represent Avg Diff >1000.

Fig.3 Known transcripts differentially expressed in response to overexpression of human *Otx2*

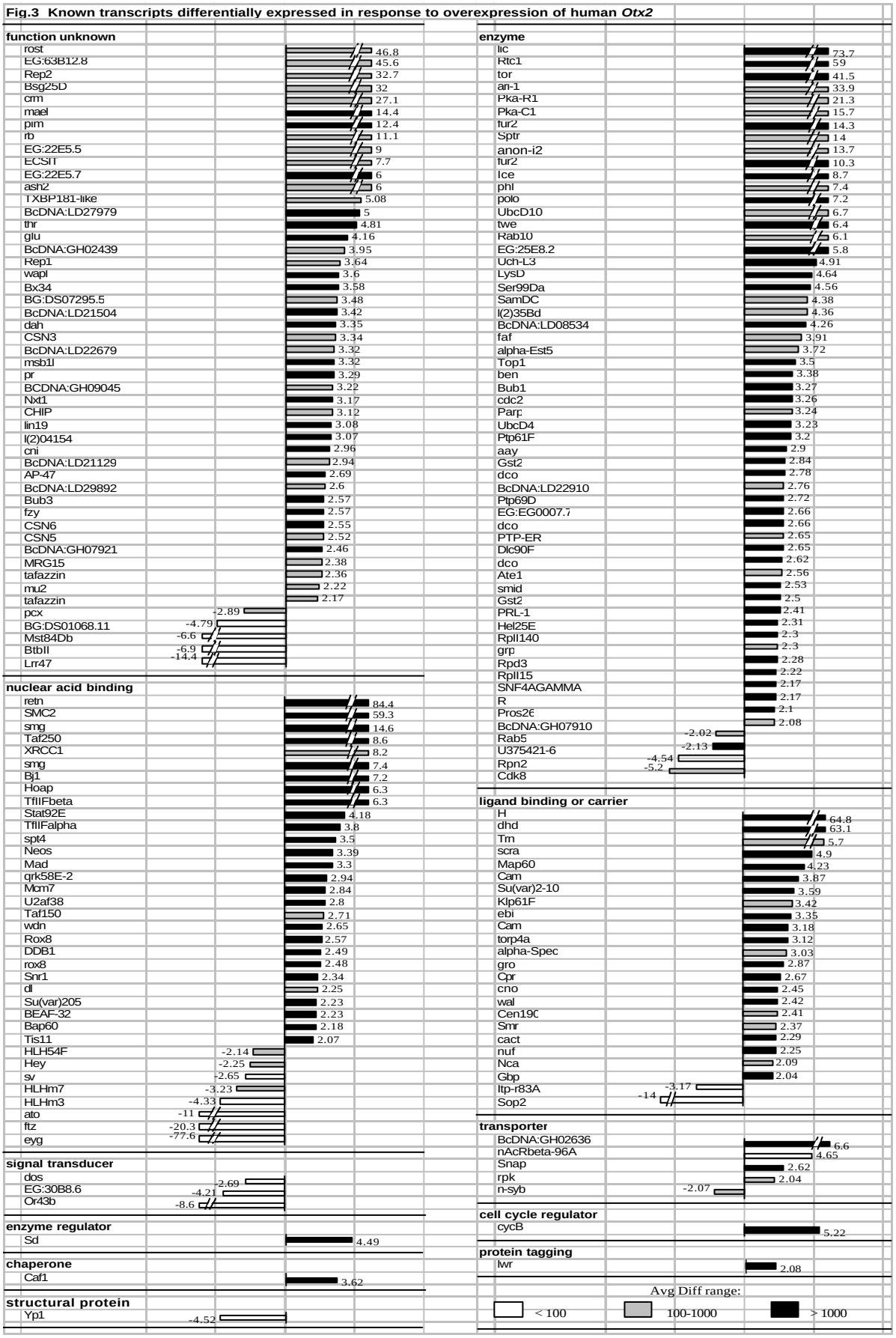


Figure 4 Transcripts differentially expressed in response to overexpression of *otd* and in response to overexpression of human *Otx2*, grouped according to functional classes. Bars represent the fold change between differentially expressed transcripts in hs*wt* embryos and hs*otd* or hs*Otx2* embryos. The upper bars represent the fold change of differentially expressed transcripts following overexpression of *Otx2* and the lower bars represent the fold change of differentially expressed transcripts following overexpression of *otd*. Positive values indicate that the relative expression level of a gene is increased (up-regulated) following *otd* overexpression and negative values indicate a decrease (down-regulated). Absolute Average Difference (Avg Diff) values are given for the *otd* overexpression condition as follows: white bars represent Avg Diff < 100, gray bars represent Avg Diff ranging from 100-1000, and black bars represent Avg Diff >1000.

Fig.4 Transcripts differentially expressed in response to overexpression of *otd* and human *Otx2*

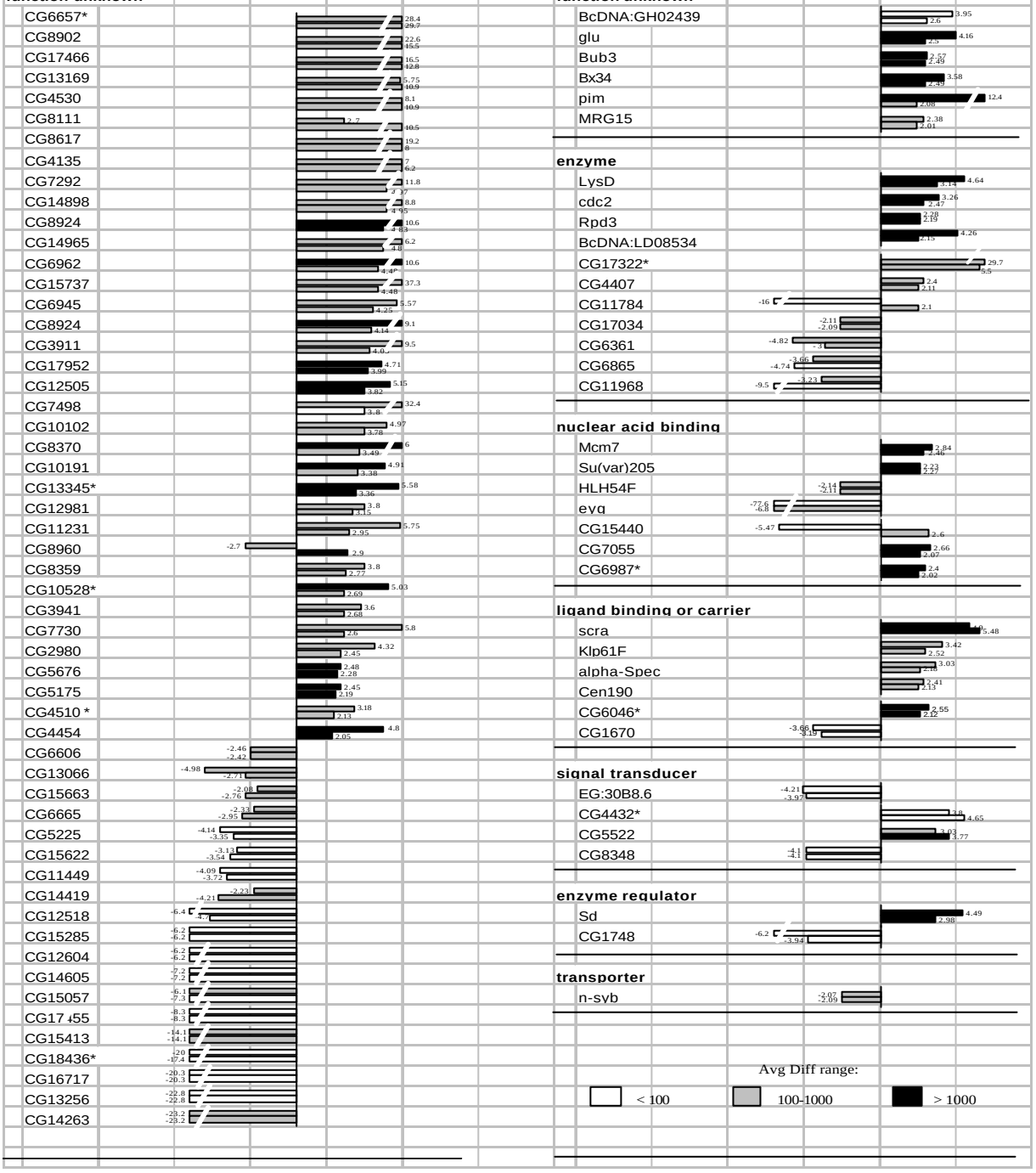


Table 3. Comparison of change folds between oligonucleotide arrays and RT-PCR

RT-PCR was performed on cDNA derived from hs*wt*, hs*otd* or hs*Otx2* embryos. Change folds determined by RT-PCR are represented as the mean value of eight independent replicates, derived from two different cDNA preparations.

	Avg diff			Change fold			
				HS-otd		HS-*Otx2*	
Transcript	**HS-wt**	HS-otd	**HS-*Otx2***	**Array**	**RT-PCR**	**Array**	**RT-PCR**
scra	251	1375	1229	5.5	1.3	4.9	1.6
LysD	525	1646	2436	3.1	1.6	4.6	4.0
glu	479	1196	1991	2.5	1.8	4.2	10.9
Rpd3	1170	2562	2673	2.2	2.0	2.3	2.5
pim	118	246	1467	2.1	1.4	12.4	8.0
n-syb	612	293	296	-2.1	-1.5	-2.1	-1.5
eyg	1552	229	10	-6.7	-1.4	-77.6	-1.6
wun	885	/	884	/	/	1	1.00
Scc1	724	723	/	1	1.0	/	/
otd	84	6555	108	78.0	119.4	1.3	1.5

6. Isolation of genetically labeled cells from the neuroectoderm of *Drosophila* embryos for genome-wide microarray analysis of downstream genes of *glial cells missing*

Haiqiong Montalta-He*, Yun Fan*, Boris Egger*, Ulrich Certa[1] and Heinrich Reichert*

*Institute of Zoology, Biocenter/Pharmacenter, Klingelbergstrasse 50, University of Basel, CH-4056 Basel, Switzerland

[1] Roche Genetics Pharmaceuticals Division, F.Hoffmann-La Roche, Ltd., CH-4070 Basel, Switzerland

Submitted

Summary

A combination of genetic labeling and magnetic cell sorting was used for isolating neuroectodermal cells from *Drosophila* embryos for microarray analysis of *gcm* action in neurogliogenesis. The *GAL4-UAS* system was used to direct expression of mCD8-GFP, a molecular label suitable for magnetic cell isolation, exclusively to the neuroectoderm of stage 11 embryos. Labeled cells were then dissociated and separated using magnetic cell sorting techniques, which permitted a high rate of purification of viable cells from the neuroectoderm as assayed by both cellular and molecular methods. Using this cell separation technique in combination with full genome microarrays, differential gene expression was analysed in wildtype embryos versus embryos in which *gcm* was misexpressed throughout the neuroectoderm. For comparison, we used the same microarrays to analyse differential gene expression in the same two sets of embryos as wholemounts. In microarray experiments involving sorted cells, 76 genes were identified as differentially expressed following *gcm* misexpression. This contrasted with the results of the wholemount-based experiments, in which 242 genes were judged as differentially expressed following *gcm* misexpression. Moreover, validation studies of genes identified as differentially expressed by in situ hybridization revealed a rate of confirmation for the sorted cell-based microarray experiments of more than 80%. These experiments imply that reduction of cell heterogeneity through cell sorting techniques leads to a marked increase in the ability of microarrays to reveal differential gene expression in the developing nervous system.

1. Introduction

The formation of a functional nervous system requires the correct specification of a large number of different cell types. These cell types fall into two major categories, neurons and glial cells (Jones, 2001). Accordingly, an important issue in developmental neurobiology is to understand how this diversity is generated in the nervous system. *Drosophila* has proved to be an excellent genetic model to study the mechanisms involved in neurogliogenesis, and recently significant progress has been made in understanding the mechanisms underlying neuron-glia fate switch, symmetric-asymmetric division of the multipotent precursors, and sublineage specification (Bossing et al., 1996; Schmidt et al., 1997; Schmid et al., 1999; Anderson, 2001).

In *Drosophila*, the gene *glial cell deficient/glial cells missing* (*gcm*) is the master regulator of glial cell fate determination. It encodes a transcription factor that is transiently expressed in glial precursors in the neuroectoderm (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996). In the *gcm* mutant, cells that normally develop into glia enter a neuronal differentiation pathway leading to a loss of glia and a gain of neurons. By contrast, misexpression of *gcm* in neural progenitors results in an increase of glial cells at the expense of neurons (Akiyama-Oda et al., 1998; Bernardoni et al., 1998; Jones, 2001). The molecular mechanisms of *gcm* action in *Drosophila* are thought to be mediated through the regulation of *gcm* downstream target genes. However, until recently, molecular genetic analyses have identified only few genes as *gcm* targets that are involved in gliogenesis (Klämbt, 1993; Campbell et al., 1994; Klaes et al., 1994; Xiong and Montell, 1995; Granderath et al., 1999).

In order to identify additional *gcm* target genes, two genome-wide microarray experiments have been carried out recently using whole mount embryos in which *gcm* was misexpressed genetically throughout the entire neuroectoderm (Egger et al., 2002; Freeman et al., 2003). Both studies reported the identification a large number of differentially expressed candidate genes following *gcm* misexpression in embryos as compared to wild type-like control embryos. However, changes of gene expression could be validated by in situ hybridization or immunostaining for only a very limited number of these candidate genes (Egger et al., 2002; Freeman et al., 2003). Indeed, based on the low level of validation, an estimate of the number

of false positive results in whole embryo microarray studies of this type has been given at 88% (Freeman et al., 2003). Clearly, such a high level of false positives results would hinder the further application of microarray technology to studies of neurogliogenesis in *Drosophila*.

The low level of validation attained in conventional microarray experiments is a general problem, and one of the major reasons for this drawback appears to be the complexity of the tissue used for the microarray experiments. For the microarray experiments mentioned above, the whole embryos rather than the neuroectoderms of the embryos had been involved which certainly contributed to the high false positives. The high complexity of the tissue samples used creates a signal-to-noise problem for the specific detection of gene expression in a given microarray experiment (Barlow and Lockhart, 2002; Griffin et al., 2003; Henry et al., 2003). One way to solve problems of tissue heterogeneity is to reduce as much as possible the irrelevant tissues. This can be achieved by dissecting the part of the organisms interested or purifying specific cell types from complex tissue. There have been several successful examples of microarray experiments based on purification of specific cell types; these include the application of Laser Captured Microdissection (LCM), Fluorescent Associated Cell Separation (FACS), single cell transcript profiling or mRNA-tagging (Bryant et al., 1999; Mills et al., 2001; Roy et al., 2002; Luzzi et al., 2003; Tietjen et al., 2003). Thus, access to a homogeneous population of specific cell types facilitates the application of microarray analysis in developmental biology.

In this report we adapted the well established method of magnetic cell separation (MACS) (Safarik and Safarikova, 1999) to isolate neuroectodermal cells from *Drosophila* embryos for microarray analysis of *gcm* action in neurogliogenesis. For this purpose, neuroectodermal cells were genetically labeled with a transmembrane fusion protein consisting of murine CD8 fused with GFP (mCD8-GFP); following cell dissociation, mCD8-positive neuroectodermal cells were incubated with magnetic microbeads coupled with anti-mCD8 antibody and were subsequently enriched by magnetic sorting. Using this cell separation technique in combination with full genome microarrays, we analysed differential gene expression in wild type embryos versus embryos in which *gcm* was misexpressed throughout the neuroectoderm. For comparison, we used the same microarrays to analyse differential gene expression in the same two sets of embryos as wholemounts, i.e. without cell separation. In microarray

experiments involving sorted cells, 76 genes were identified as differentially expressed following *gcm* misexpression. This contrasted with the results of the wholemount-based experiments, in which 242 genes were judged as differentially expressed following *gcm* misexpression. Moreover validation studies of genes identified as differentially expressed by in situ hybridization revealed high rates of confirmation for the sorted cell-based microarray experiments. Taken together, our experiments imply that reduction of cell heterogeneity through cell sorting techniques leads to a marked increase in the ability of microarrays to reveal differential gene expression in the developing nervous system.

2. Material and Methods

Flies

Drosophila melanogaster stocks were raised on standard cornmeal/yeast/agar medium at 25°C. To label the embryonic neuroectoderm, virgin females from *scabrous-GAL4 (sca-GAL4)* (Klaes et al., 1994) were crossed to w^{1118};;*UAS-mCD8::GFP* males (Lee and Luo, 1999). Previous studies indicated that there is no detectable toxicity due to overexpression of the mCD8-GFP fusion protein in *Drosophila* (Lee and Luo, 1999). For ectopic expression of *gcm, sca-GAL4* virgins were crossed to w^{1118};;*UAS-mCD8::GFP,UAS-gcm* (Bernardoni et al., 1997). After a 1h pre-collection, embryos were collected in parallel for 1 h and staged to 6-7 h AEL (late stage 11). Stages are according to Campos-Ortega and Hartenstein (Campos-Ortega, 1997).

Embryo dissociation

Embryos were dechorionated in 3-4% chlorax for 4 min, collected on a mesh and rinsed thoroughly with water. After incubation with 60µg/ml Proteinase K (Invitrogen, 20mg/ml) in phosphate buffered saline (PBS) for 2 min, embryos were dried and transferred into a 15ml polypropylene screw-cap tube (Falcon) containing 1ml of 10x Trypsin-EDTA (GIBCO, Cat# 35400-027). The embryos were homogenized by passing ~20 times through a 21 G needle. After adding 4 ml PBS into the tube, the tube was fixed on a shaker and incubated for 30 min at 25º C at 800 rpm. Finally, the homogenate was filtered through a cell strainer (40µm, Falcon) to remove tissue clumps. Cells were pelleted at 1500rpm, 4°C for 5min in a tabletop centrifuge. After resuspension in 0.5ml of MACS buffer (1xPBS, 0.5% BSA and 2mM EDTA), cells were incubated with Hoechst 333342 (2µg/ml final concentration; Molecular Probes) and/or propidium iodide (10µg/ml final concentration; Molecular Probes) for 30 min at room temperature to allow dyes to equilibrate. Flow cytometer analysis was performed in a Becton Dickinson FACSCalibur flow cytometer.

Magnetic Cell Sorting (MACS)

Cell concentration was determined using haemacytometer before cells were resuspended in 90ul of MACS buffer per 10^7 total cells. 10µl of MACS CD8a(Ly-2) Microbeads (Miltenyi Biotec, Cat# 130-090-401) was added to every 10^7 cells. Free microbeads were washed away by adding 10-20 x labeling volume of MACS buffer and centrifuging at 1500rpm for 5min.

Cells were resuspended in 0.5ml MACS buffer and loaded onto a prepared MS+/RS+ type column (Miltenyi Biotec, Cat# 130-042-201) in the magnetic field of an octoMACS separator (Miltenyi Biotec, Cat# 130-042-109). Non-labeled cells flowed through the column while labeled cell were retained in the column. After washing the column four times with MACS buffer, retained cells were flushed out in 1ml MACS buffer using the plunger supplied with the column. Cells were pelleted and stored in liquid nitrogen before total RNA isolation.

Real Time Reverse Transcription Polymerase Chain Reaction

For real time RT-PCR experiments, 300 ng poly $(A)^+$ RNA was isolated (mRNA isolation kit; Roche Diagnostics) and reverse transcribed with AMV-RT and random hexamers (RT-PCR kit; Roche Diagnostics). PCR was performed with 100 pg template DNA and gene-specific primers (Seq Web, Winsconsin Package Version 10.0, GCG) on a light cycler (LightCycler, Roche Diagnostics). Continuous fluorescence observation of amplifying DNA was possible using SYBR Green I (Roche Diagnostics). After cycling, a melting curve was produced by slow denaturation of the PCR end products to validate the specificity of amplification. To compare the relative amounts of PCR products we monitored the amplification profile on a graph, displaying the log of the fluorescence intensity against the number of cycles. Relative fold changes for a given gene under both conditions (sorted *vs.* flowthrough) were calculated using the fit point method (LightCycler Manufacturer, Roche).

RNA isolation, target preparation and hybridization

Total RNA from sorted cells and embryos was extracted using a Mini RNA Isolation Kit (Zymo Research, Cat# R1005) and was eluted with RNase-free water. Quality and quantity of the RNA samples were assessed using a RNA 6000 NANO Chip (Agilent Technologies, Cat# 5065-4476). Because of the small amount of total RNA derived from the MACS sorted cells, we used a commercial Microarray Target Amplification Kit (Roche, cat# 3 310 191) to synthesize target cRNA; this involved PCR-amplification of cDNA after normal retro-transcription. Target cRNA preparation from the embryos was as previously described (Leemans et al., 2001). In both cases, 20µg of biotinylated antisense cRNA were ultimately hybridized to the arrays according to standard protocol (Montalta-He et al., 2002).

Oligonucleotide Arrays

For expression profiling, DrosGenome1, a high density oligonucleotide array (Affymetrix, cat# 900 335) was used. This array was based on the Release 1.0 of the *Drosophila* genome. (Sequences were downloaded from the Flybase database on August 25th, 2000.) Sequences on the array represented more than 13,500 predicted transcripts as well as different control genes. Each sequence is represented on the array by a set of 14 oligonucleotide probes of matching sequence and 14 oligonucleotide probes with a single nucleotide mismatch. The signal intensity is calculated by an algorithm based on the perfect match hybridization signal and on the mismatch hybridization signal, and is proportional to the abundance of a given transcript (Rajagopalan, 2003). Four replicates were performed for each experimental condition.

Data analysis

Data acquisition and processing by RACE-A was as described elsewhere(Montalta-He et al., 2002). For quantification of relative transcript abundance, the value of signal intensity was used. All arrays were normalized against the mean of the total sums of signal intensity values. For differential transcript imaging, only transcripts that showed an expression level fold change (FC) = 2.0 or = -2.0 at significance values of p = 0.01 (unpaired t-test) were considered to be differentially expressed. The complete list of the microarray expression data involved can be accessed at http://www.ncbi.nlm.nih.gov/geo/. Accession number: GSE612 (sorted experiment) and GSE613 (wholemount experiment).

Synthesis of RNA probes from PCR products

Primers for in situ hybridization probe synthesis were designed according to the coding region of the gene studied. A T3 promoter sequence, 5' ATTAACCCTC ACTAAAGGGA GA 3', was added to the 5 prime of the 3 prime end primer. Normal PCR reaction was performed using cDNA from the stage 11 embryos as the template (Kain et al., 1991). RNA probes were prepared according to the standard protocol.

In situ hybridization and immunocytochemistry

In situ hybridization was carried out according to Tautz and Pfeifle (Tautz and Pfeifle, 1989). Embryos were mounted in Canada balsam (Serva), viewed on a Zeiss Axioskop microscope

with differential interference contrast optics and photographed with a Prog/Res/3008 digital camera (Kontron). Immunocytochemical experiments were carried out as described previously(Therianos et al., 1995). The primary antibodies were rat anti-RK2/REPO diluted 1:1000 (Campbell et al., 1994). For fluorescent labeling, rat anti-mouse CD8-FITC (Miltenyi Biotec) was used 1:10 and secondary antibodies were Alexa568 and Alexa488 conjugated and diluted 1:150 (Molecular Probes). Fluorescently labeled embryos were viewed with a Leica TCS SP confocal microscope.

3. Results

Labeling of the embryonic neuroectoderm and targeted misexpression of *gcm*

For a genome-wide identification of genes that are either direct *gcm* target genes or among the initial set of downstream genes of *gcm*, we studied differential gene expression at embryonic stage 11. At this stage, the first glial marker, the direct *gcm* target gene *repo*, is expressed in the neuroectoderm. Cells in the neuroectoderm of stage 11 embryos were labeled genetically with a transmembrane protein consisting of murine CD8 fused with GFP. This was achieved by crossing a *sca-Gal4* enhancer trap line (Klaes et al., 1994) with a *UAS-mCD8-GFP* line (Lee and Luo, 1999), and resulted in uniform labeling of the cell surface of all cells in the neuroectoderm. This was referred to as the wild type-like situation. An example of the extent of this type of neuroectodermal labeling by mCD8-GFP in stage 11 embryos is shown in Figure 1A; localization of intense staining at the surface of the cells is very clear. No homologs of mCD8 or GFP exist in *Drosophila*, thus the only cells in the embryo that express mCD8 or GFP are those targeted by the GAL4-UAS system. In order to misexpress *gcm* in the labeled embryonic neuroectoderm of stage 11 embryos, the same *sca-GAL4* line was crossed with a recombinant *UAS-mCD8-GFP, UAS-gcm* line of which the neuroectoderm is also labeled by mCD8-GFP (Fig. 1B). This was referred to as the *gcm* misexpression situation.

With the exception of altered gene expression in cells of the neuroectoderm, no obvious morphological changes were seen in the stage 11 wild type-like or *gcm*-misexpression embryos.

In the wild type-like situation, endogenous *gcm* expression was seen in two small groups of neuroectodermal cells in each hemisegment during stages 10-11 (Fig. 1C). At stage 11, a single *gcm*-expressing neural precursor delaminated from each of these groups and gave rise to cells expressing the glia-specific *repo* gene, a direct target of *gcm* (Fig. 1E)(Hosoya et al., 1995; Jones et al., 1995). In contrast, in the targeted *gcm* misexpression situation, all of the cells in the neuroectoderm expressed *gcm* at stages 10-11 (Fig. 1D). In consequence, most of the neural precursor cells expressed the glia-specific *repo* gene at stage 11 (Fig. 1F) (Egger et al., 2002).

MACS allows efficient recovery of mCD8-labeled cells from dissociated embryonic neuroectoderm

Cells were dissociated from stage 11 embryos by homogenization and treatment with trypsin followed by filtration (see Materials and Methods). To assess the extent of cell dissociation and the rate of cell survival, cells were stained with Hoechst 33342 for visualization of nuclei and with propidium iodide, which is excluded from cells with intact membranes and hence stains the dead cells. Examination of dissociated cell preparations stained with Hoechst 33342 showed that the majority of the stained material corresponded to single cells (Fig. 2A). Some small clusters of non-dissociated cells were also observed. Examination of dissociated cell preparations stained with propidium iodide showed that only very few cells were stained with propidium iodide (Fig. 2B). For quantification of this, propidium iodide stained preparations were analysed by a flow cytometer; this analysis indicated that over 90% of dissociated cells were viable judged by their low levels of propidium iodide staining (Fig. 2C).

After cell dissociation, microbeads coupled with anti-mCD8 antibody were incubated with the concentrated cell solution and then applied to a separation column in a magnetic field to enrich for mCD8 expressing cells (see Figure 3 and Material and Methods). According to the MACS protocol used, mCD8-positive neuroectodermal cells were expected to be selectively retained in the column while the mCD8-negative non-neuroectodermal cells were expected to flow through. To assay the degree of cell purification obtained by the MACS procedure, both the mCD8-positive neuroectodermal cells (referred to as sorted fraction) and the mCD8-negative cells (referred to as flowthrough fraction) were characterized by studying GFP fluorescence (cell labeling was achieved with a mCD8-GFP fusion protein). As expected, only a few cells in the flowthrough fraction were GFP positive judged by fluorescence microscopy (Fig. 4 A, B). In contrast, the majority of the cells in the sorted fraction appeared as GFP positive (Fig4. C, D). For quantification and control of the degree of GFP labeling of cells in the sorted fraction, a flow cytometry analysis was carried out for each experiment performed in this study. In all cases, flow cytometry analysis indicated a >60% purity of GFP-positive cells, and in some cases values as high as 85% purity were obtained (Fig. 4E). The overall morphology of the cells in both fractions appeared to be normal. Cell separation and purification rates obtained from wild type-like embryos and from *gcm* overexpression

embryos were not significantly different. Taken together, this indicates that a marked enrichment of the mCD8-positive neuroectodermal cells was obtained through cell dissociation and magnetic cell separation.

For an independent molecular confirmation of the enrichment efficiency of the cell sorting procedure, real-time RT-PCR was performed on cDNA prepared from the sorted fraction versus the flowthrough fraction for three genes: *sca* was chosen as a positive control because sorted cells were derived from the endogenous *sca* domain; *cg9232* gene, which is expressed in the embryonic endoderm and posterior/anterior midgut primordium hence should not be enriched in the sorted neuroectodermal cells, was for negative control; finally, the ubiquitously expressed *rp49 (ribosomal protein 49)* gene was selected for base line control. 4 independent replicates, derived from two different cDNA preparations were carried out. As expected, *rp49* had similar expression levels in both fractions. The *sca* gene was enriched 5.7-fold in the sorted fraction as compared to the flowthrough fraction; this indicates a marked enrichment of endogenous *sca*-expressing neuroectodermal cells (which are also labeled transgenically with *sca-GAL4/UAS-mCD8-GFP*) in the sorted fraction. Conversely, the *cg9232* gene was enriched 4.6-fold in the flowthrough fraction as compared to the sorted fraction. These results confirm, at the molecular level, that it is possible to markedly enrich for neuroectodermal cells labeled with mCD8-GFP from the embryos using the magnetic cell separation procedure.

Overview of gene expression profiling following gcm overexpression in the embryonic neuroectoderm

Analyses of differential gene expression in the *gcm* misexpression versus wild type-like situations were carried out using full-genome high density oligonucleotide arrays. Differential gene expression was determined by transcript profiling of sorted cells derived from *gcm* misexpression embryos as compared to transcript profiling of sorted cells derived from wild type-like embryos. Four replicates were performed for each experimental condition. These experiments identified 76 genes as differentially regulated following *gcm* overexpression. The same number of transcripts had upregulated (n=38) and downregulated (38) expression levels. All 76 genes were classifed based on molecular function according to Gene Ontology

(Ashburner et al., 2000) (Table 1). Strikingly, but not surprisingly, the majority of the differentially regulated genes (54%) were of currently unknown function. The two functional classes with the largest number of differentially regulated transcripts were enzymes (15) and nucleic acid binding (12), including 10 transcription factors.

The relatively small number of 76 genes identified as differentially regulated following *scaGAL4/UASgcm* mediated misexpression of *gcm* in the neuroectoderm obtained in this cell sorting-based microarray experiment contrasts with the larger numbers of differentially regulated genes obtained in comparable wholemount-based microarray experiments involving *gcm* misexpression (Egger et al., 2002; Freeman et al., 2003). To determine if these differences in gene numbers might be due to the different cell-sorting versus wholemount situations, we repeated out transcript profiling experiments on wholemount embryos identical to those used for cell sorting (*gcm* misexpression versus wild type-like).

In the transcript profiling experiments involving wholemount embryos, 242 transcripts were judged to be differentially regulated by *gcm* overexpression as compared to wild type. Approximately the same number of transcripts had increased (n=116) and decreased (n=126) expression levels. Classification of these differentially expressed genes (Table 1) showed that the majority of genes (51%) were of unknown function, and the two functional classes with the largest number of transcripts were again enzymes (65) and nucleic acid binding (18), including 14 transcription factors. Only 13 of the 242 transcripts judged to be differentially expressed in the wholemount-based experiment were also found as differentially expressed in the cell sorting-based experiment.

A comparison of the results obtained in the cell sorting-based experiments versus those obtained under otherwise identical conditions in wholemount based-experiments, shows clearly that much fewer genes were judged to be differentially regulated following cell sorting. It seems likely, that this is due to the use of more homogeneous target tissue in the cell sorting experiments.

Quantitative transcript profiling of differentially expressed genes in cell sorting-based experiments

Figure 5 shows the differentially regulated genes identified by transcript profiling of sorted cells derived from *gcm* misexpression embryos versus wild type-like embryos. It presents a quantitative representation of the change in expression levels for these gene transcripts. The gene with the highest increase in expression level (12.9-fold) was *cg9451*, which has been predicted to encode an adenylate kinase. Increases in expression level above 10-fold were also observed for *cg3132*, which encodes a beta-galactosidase. There were 4 transcripts which all encode proteins of unknown function and showed increases in the 5-10-fold range, namely *cg12910, cg6218, cg5822* and *cg12641*. The first three have been shown to be expressed in glial cells and be potential *gcm* regulated genes (Freeman et al., 2003). The majority of the remaining upregulated transcripts have increases in the range of 2-5 folds.

None of the genes with decreased expression levels have decrease levels below −5 fold. The gene with the largest decrease in expression level (-4.9-fold) was *cg14830*; its function is currently unknown. Interestingly, among the genes with downregulated expression levels are several genes, which have been shown to act in the neuronal cells or neuronal precursors. These are *al, ey, ewg, nrm, scrt* and *ftz* (Doe et al., 1988; DeSimone and White, 1993; Roark et al., 1995; Kammermeier et al., 2001). The downregulation of these genes is in accordance with the model of GCM action during gliogenesis that *gcm* suppresses genes functioning in the neuronal cell lineages (Giesen et al., 1997).

Validation of results of cell sorting experiments by in situ hybridization in embryonic CNS

Can the differential expression of genes that result from microarray analysis of cell sorted *gcm* misexpression embryos be confirmed by tissue-specific spatial expression studies? To address this question, we focused on the set of 17 genes among the genes of the highest increase in expression levels in the sorting-based experiment. These genes are highlighted in bold in Figure 5. Among these 17 genes, 9 have been shown by in situ hybridization to be differentially expressed following *gcm* overexpression in previous reports (Fig. 5)(Freeman et

al., 2003). The remaining genes were studied by in situ hybridization using the same transgenic lines as those employed for microarray analysis. Among them, 5 showed clear differential expression patterns in *gcm* overexpression embryos as compared to wild type-like embryos (Fig. 5 and see below for in situ data). Thus, among the 17 genes that we considered and that were studied by in situ hybridization, a total of 14 (82%) were validated by in situ hybridization as differentially expressed in embryonic tissue following *gcm* misexpression.

Interestingly, 7 of these 14 validated genes from the cell sorting-based microarray experiment were also found among the genes judged to be differentially expressed in the wholemount-based microarray experiment (Table 2). These are actually all the upregulated genes in the overlap of sorted and wholdmount experiment. This suggests that genes judged as differentially expressed in both the cell-sorted microarray experiment and in the wholemount-based microarray experiment might have high rate of validity. To investigate this, we performed in situ hybridization studies for another 4 genes judged as downregulated in both experiments (Table 2). Among these 4 genes, 3 showed clear differential expression patterns in *gcm* overexpression embryos as compared to wild type-like embryos. Thus, among the 13 genes that showed differential expression in both microarray studies, 11 have been studied by in situ hybridization, and 10 of these (91%) were validated by in situ hybridization as differentially expressed in embryonic tissue following *gcm* misexpression (Table 2, see below for in situ data).

Taken together, the in situ hybridization studies carried out here validated 8 new candidate *gcm* downstream genes. The in situ expression patterns of these 8 genes are shown for the wild type-like versus the *gcm* misexpression situation of stage 11 embryos in figure 6. In the case of *cg9541*, *gcm* misexpression resulted in an increased expression of the gene in specific, segmentally repeated domains of the ventral neuroectoderm (Fig. 6A, B). A similar situation was observed for *cg15307*, where *gcm* misexpression caused an increase in gene expression in segmentally reiterated domains of the ventral neuroectoderm as well as in specific cephalic regions (Fig. 6C, D). For *cg12641* a comparable increase in expression in the segmentally repeated domains of ventral neuroectoderm was seen following *gcm* misexpression, however, this increase was not as strong as those observed for *cg9541* and *cg15307* (Fig.6E, F). In contrast to the relatively broad expression domains of these three genes in the wild type-like

situation, the expression domains of the genes *cg3132* and *cg6560* were more localized to specific groups of neuroectodermal cells in the wild type-like situation (Fig. 6G, I). Accordingly for these two genes, *gcm* misexpression resulted in an increase of expression localized to these groups of neuroectodermal cells (Fig. 6H, J). The genes *nerfin-1* and *cg17649* were expressed in a comparable set of neuroectodermal cells in the wild type-like situation (Fig. 6K, M). For *nerfin-1*, these neuroectodermal cells have been identified as neuroblasts and ganglion mother cells (Stivers et al., 2000). Following *gcm* misexpression throughout the neuroectoderm, the expression of *nerfin-1* and *cg17649* was strongly reduced or even abolished in some of these cells (Fig. 6L, N). For the *al* gene, the wild type-like expression at stage 11 has been characterized as restricted to a segmentally repeated pattern of three thoracic and eight abdominal lateral patches of which the eighth spot conceivably labels the anterior lateral sense organs (Schneitz et al., 1993) (Fig. 6O). Misexpression of *gcm* in the neuroectoderm appeared to repress the expression of *al* in the abdominal patches (Fig. 6P).

Discussion

Genetic labeling coupled with magnetic cell separation leads to efficient isolation of viable cells from genetically specified embryonic domains.

Microarray studies involving wholemount tissues are inherently complicated by the diversity of cell populations. This problem is more prominent for studies of the nervous system where cells of interest may comprise only a fraction of the entire tissue studied, anatomical divisions between regions of the nervous system are often unclear, and precisely controlled, artifact-free dissections are difficult(Barlow and Lockhart, 2002; Griffin et al., 2003). In consequence, averaging expression levels of entire tissue regions, or of entire embryos, may minimize or conceal even large expression changes that occur in small subpopulations of cells. This problem is aggravated in studies of neuronal development due to the small size of embryonic nervous system and the difficulty in identifying the subpopulations of interest for dissection in embryos. In order to overcome this experimental obstacle, methods are needed that allow the isolation of part of the organism interested or specific cell subpopulations from wholemount tissues.

In this study we have used a combination of genetic labeling techniques and magnetic cell sorting for isolating neuroectodermal cells from *Drosophila* embryos for microarray analysis. The high spatiotemporal specificity of the *GAL4-UAS* system was used to direct expression of mCD8-GFP, a molecular label suitable for magnetic cell isolation, exclusively to the neuroectoderm of stage 11 embryos. Labeled cells were then dissociated and separated using magnetic cell sorting techniques, which permitted a high rate of purification of viable cells from the neuroectoderm as assayed by both cellular and molecular methods. Given the versatility and precision of the *GAL4-UAS* system and the increasing number of specific GAL4 lines, specific labeling of cell types or of tissue domains followed by cell sorting and microarray analysis should be possible for virtually all embryonic or postembryonic cell types and gene expression domains in *Drosophila*.

Improved identification of *gcm* downstream genes by microarray analysis of sorted neuroectodermal cells

Overexpression of *gcm* in the embryonic neuroectoderm of *Drosophila* followed by microarray analysis of differential gene expression based on wholemount embryos have been

reported by two groups. In these studies, differential regulation of 417 potential *gcm* downstream genes (Egger et al., 2002) and 153 potential *gcm* downstream genes (Freeman et al., 2003) resulted from the microarray analyses. The wholemount-based microarray analysis of *gcm* downstream genes carried out as a control in our investigation resulted in 242 genes judged as differentially expressed. These relatively large gene numbers are in marked contrast to the 76 genes that were identified as differentially expressed in our sorted cell-based microarray analysis. A direct comparison of sorted cell-based versus wholemount-based microarray analyses is possible for our data, since both experiments were performed under virtually identical conditions (identical fly lines, genetic background, RNA isolation, microarray hybridization, microarray composition, and data analysis).

Based on the results, we strongly suggest that much more homogeneous nature of the target tissue obtained by cell sorting versus whole embryos lead to the fact that much fewer genes were judged to be differentially regulated in the sorted cell-based microarray experiment. This is because the GAL4-UAS method used limits *gcm* overexpression to precisely the same tissue that is labeled for subsequent magnetic cell sorting. Thus, signal loss due to "dilution" of neuroectodermal RNA by RNA obtained from other parts of the embryo is avoided, and tissue contamination from non-neuroectodermal cells that might be non-specifically affected by *gcm* action is minimized. However, we cannot entirely rule out, that the smaller number of genes identified as differentially regulated is due to artifacts of the cell sorting or target amplification procedures used.

Independent confirmation of the improved gene identification rate obtained by microarray analysis from sorted cells is demonstrated by the results of in situ-based validation studies. Thus, among a group of 18 genes selected exclusively on their level of expression change in the microarray experiment, 9 had already been validated as *gcm* downstream genes by previous in situ hybridization studies, and a remaining 5 were validated by in situ experiments in this report. This implies a validation rate of 82%. Interestingly, an even higher validation rate (91%) was obtained for the 13 genes that were judged as differentially expressed in both the cell sorted-based and the wholemount-based experiment. Taken together, these findings suggest that identification of *gcm*-downstream genes is markedly improved by sorted cell-based microarray analysis. However, more extensive and complete validation studies must be

carried out on a gene-by-gene basis, before a full quantitative appreciation of the advantages of cell sorting for *gcm* target gene identification can be obtained.

Despite the improved gene identification that is obtained by basing microarray analysis on homogeneous cell populations, it is clear that false positive results cannot be entirely eliminated from microarray investigations. For example, hybridization errors will occur due to imperfections in the manufacture of microarrays. Moreover, the design of probe sets on the arrays that precisely represent all the genes in the genome yet eliminate cross-hybridization is difficult. Furthermore, annotation of a fully sequenced genome, such as that of *Drosophila*, remains tentative and indeed can change significantly during a period of improvement and validation (Misra et al., 2002), and this effects the ability of the microarray probe sets used on the arrays to detect the biologically correct gene. Finally, while the cell sorting method used was highly efficient in isolating cells and also gave excellent results in subsequent microarray analyses, some isolation-based artifacts cannot be entirely ruled out. Since mechanical homogenization and enzymatic digestion were used to dissociate cells, it is possible that the transcriptome of the sorted cells may not be identical to that of the same cells in situ. For example, cleavage of surface proteins during dissociation might interfere with cell signaling and result in gene expression changes. Although this concern has been controlled here by monitoring gene expression of 3 genes, change of the global transcriptome can not be excluded. Moreover, given the relatively small numbers of cells obtained by sorting (as compared to the large cell numbers obtained by using entire embryos), PCR-based RNA amplification of targets for hybridization was necessary. Although the PCR-based RNA amplification method used has been reported to be sensitive and lead to highly reproducible results (Ji et al., 2000), artifacts due to amplification also might influence in the results.

Validation of new candidate *gcm* downstream genes by in situ hybridization

In *Drosophila*, *gcm* functions as a genetic switch to control neuronal versus glial cell fate and is involved in the specification of nearly most differentiated glial cell types (Jones et al., 1995; Vincent et al., 1996; Jones, 2001). The initiation of gliogenesis is thought to occur through the direct activation of downstream target genes of *gcm* (Jones, 2001). In order to identify *gcm* downstream genes, we analyzed gene expression profiles following *gcm* misexpression

in the neuroectoderm at an embryonic stage in which the first glial-specific genes such as *repo* (a direct *gcm* target gene) are expressed. This procedure should identify genes that act in neural precursors and are involved in the determination of glial versus neuronal cell lineage.

In total, 76 genes were judged as differentially expressed following *gcm* misexpression in this study; these genes are posited to be either direct *gcm* target genes or among the initial set of genes regulated by *gcm*. Among these 76 genes, 8 have for the first time been validated as potential *gcm* targets by in situ hybridization. Two of these genes have been studied previously. *nerfin-1* which encodes a transcription factor with a zinc finger. *nerfin-1* message is found in most, if not all, early neuroblasts although the protein is detected only in the nucleus of GMCs that will divide once to produce neurons (Stivers et al., 2000). As judged by *nerfin-1* and *repo* co-immunostaining which showed no overlapping in expression pattern of these two genes and our observation, it seems that *gcm* silences *nerfin-1*(Odenwald Ward, personal communication). Analysis of *nerfin-1* mutants shows normal glial and neuronal development but disrupted axon guidance (Odenwald Ward, personal communication). It requires more detail analysis to investigate the interaction between *gcm* and *nerfin-1* and its effect of gliogenesis. *al* encodes a transcription factor containing a paired type homeodomain. It is expressed in specific sensory organ anlagen during embryogenesis (Schneitz et al., 1993; Tomancak et al., 2002). *al* is involved in both embryonic development and pattern formation in appendages(Campbell, 2002). Since the isolation of complete null alleles of *al* has not yet been achieved, further study is needed for *gcm* action on *al*.

A possible involvement of the remaining 6 genes in the embryonic nervous system of *Drosophila* has not been reported previously. However, several of these genes have gene homologs which are implicated in human nervous system development and disease. For example, the human homolog of *cg9541* encodes an adenylate kinase, which shows increased activity in the cerebrospinal fluid of patients with acute onset ischaemic brain damage and in patients with specific brain tumors (Van Rompay et al., 1999). The human homolog of *cg3132* is associated with severe cerebral degeneration and accumulation of ganglioside in neurons(Giugliani et al., 1985; Caciotti et al., 2003). The human homolog of *cg6560* encodes a protein predicted to be an ARF small monomeric GTPase belonging to the Ras GTPase superfamily which is involved in cell survival (Ehrhardt et al., 2002). While these findings

suggest that cases of molecular homology might also characterize neuroglial development in insects and mammals, a more detailed analysis involving mutant analyses in fly and murine models will be necessary before a more accurate assessment of the function and conservation of these genes can be attained.

In summary, we have succeeded in applying MACS to isolate neuroectodermal cells from stage 11 embryos of *Drosophila* for microarray experiment which give much less false positive results than those based on whole mount tissue. Given that MACS only requires very simple and economic experimental settings and the specificity and versatility of the *GAL4/UAS* system, it will definitely facilitate the application of microarray techniques in *Drosophila*. In the meanwhile, the viable and relatively pure cells obtained via MACS provide important materials for experiments depending on access to specific cell types.

Acknowledgements:
We would like to thank specially Antonius Rolink and people in his lab for allowing us and helping us (specially, Ulf Grawunder, Eva Harfst and Gina Balciunaite) to use the FACSCalibur which is essential for this work, Markus Affolter and Carlos Ribeiro and Alain Jung in his lab for initiation of the idea of MACS in *Drosophila*, Sandra Klur for her generous help regarding the PCR-based microarray target preparation; Urs Kloter, Leandro Hermida, Philippe Demougin and Micheal Primig for technical support and helpful discussions; Bruno Bello and Martin Müller for critical comments on the manuscript. This work was supported by the Swiss NSF.

Figure 1. Labeling of the embryonic neuroectodermal cells and targeted misexpression of gcm leads to gain of glial cells at the expense of neuronal cells

(A,B) Immunostaing with anti-mCD8 in late stage-11 wild type-like embryos (A) and *gcm* misexpression embryos (B), which were labeled genetically with a transmembrane protein consisting of murine CD8 fused with GFP. Immunostaining (in green) shows the extent of labeling of the neuroectoderm; ventral view, anterior towards the left. (C, D) In situ hybridization of *gcm* in stage10 embryos shows expression in wild type-like embryos (C) and in *gcm* misexpression embryos (D), lateral views, anterior towards the left. In the wild type-like embryos, small clusters of cells in the neuroectoderm of the VNC express *gcm*; in *gcm* misexpression embryos, all cells of the neuroectoderm express *gcm.* (E, F) Immunostaining with anti-REPO antibody in wild type-like embryos (E) and in *gcm* misexpression embryos (F); laser confocal microscopy of stage 11 embryos, ventral views of the VNC, anterior is towards the left. In the wild type-like embryos, single glial precursors (which also express *gcm*; not shown) in each hemisegment express the *repo* gene. In *gcm* misexpression embryos, virtually all of the neuronal and glial precursor cells are REPO positive.

Figure 2. Assessment of extent of cell dissociation and determination of cell survival rate before sorting

(A, B) Staining of dissociated cells with Hoechst 33342 (A) and with Propidium iodide (B); Hoechst 33342 staining reveals a majority of single cells as well as a few small clusters of non-dissociated cells. Examination of dissociated cell preparations with propidium iodide staining shows that only very few cells were stained with propidium iodide. (C) Quantification of propidium iodide staining with flow cytometry. X-axis is the forward scatter. Y-axis is the fluorescent signal of propidium iodide staining. Each spot represents one cell. Cells within the rectangle R1 are judged as viable based on the low levels of propidium iodide staining. The number of cells in the rectangle R1 divided by the total cells counted gives a survival rate of over 90%.

Figure 1

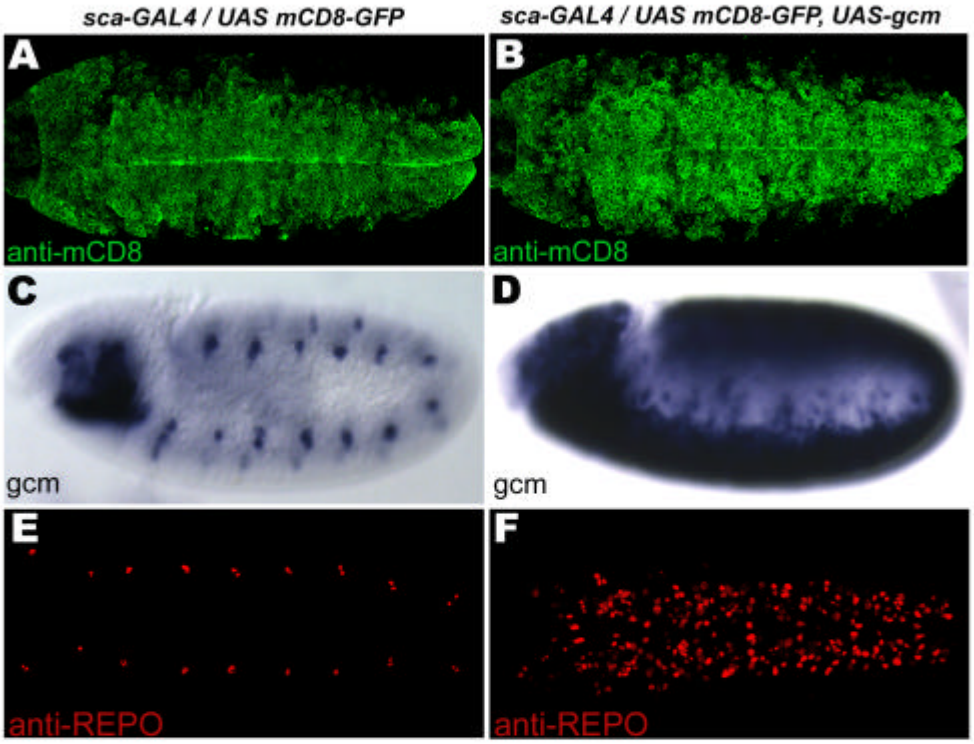

sca-GAL4 / UAS mCD8-GFP
sca-GAL4 / UAS mCD8-GFP, UAS-gcm
A
anti-mCD8
B
anti-mCD8
C
gcm
D
gcm
E
anti-REPO
F
anti-REPO

Figure 2

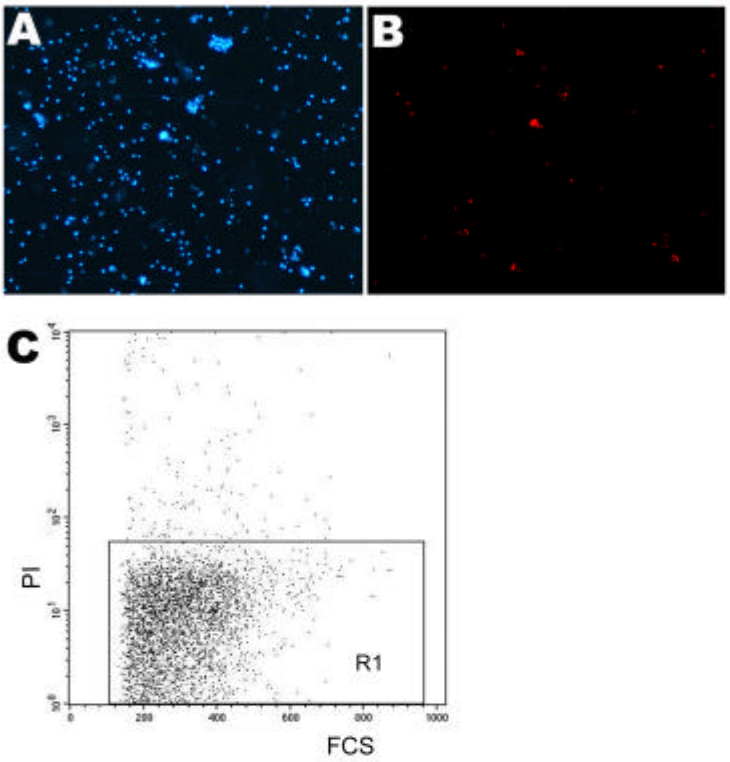


Figure 3. Outline of MACS procedure

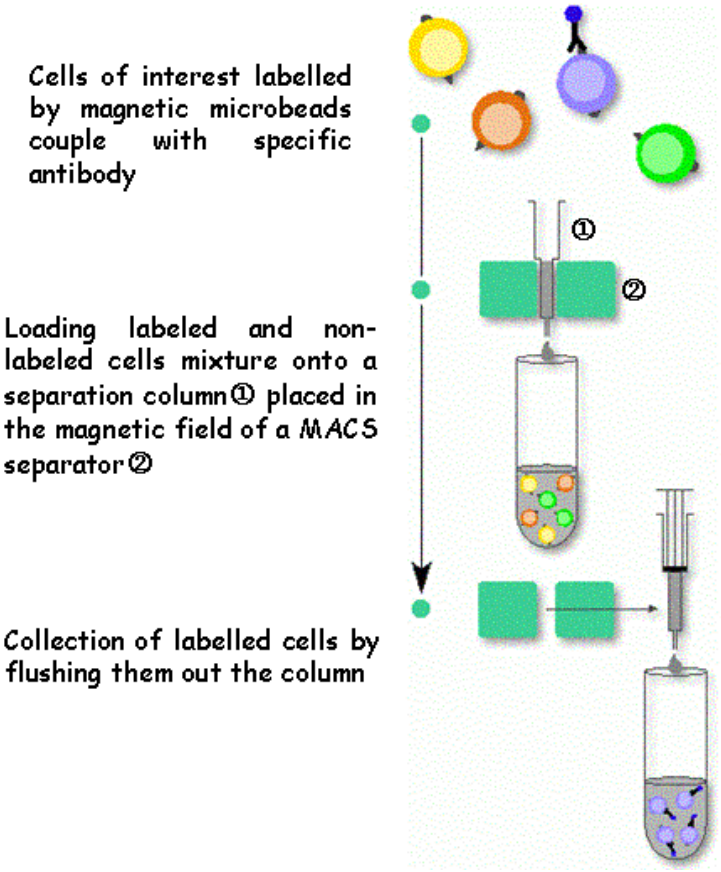


Figure 4. Analysis of the degree of cell purification obtained by the MACS procedure according to the GFP fluorescent signal

(A, B) Fluorescence microscopy of cells in the flowthrough fraction viewed under normal filter (A) and GFP filter (B); as expected, only a few cells in the flowthrough fraction were GFP positive as judged by fluorescence microscopy. (C, D) Fluorescence microscopy of cells in the sorted fraction viewed under normal filter (C) and GFP filter (B); in contrast to the flow though fraction, the majority of the cells are GFP positive. (E) Quantification of the degree of GFP labeling of cells in the sorted fraction by flow cytometry. X-axis represents the GFP signal; Y-axis is the number of cells counted. M1 indicates the region in which cells are judged GFP positive; this region is delimited towards low intensities by the values of autofluorescence observed for GFP negative cells. The number of cells in M1 divided by the number of all cells counted gives the percentage of purification.

Figure3

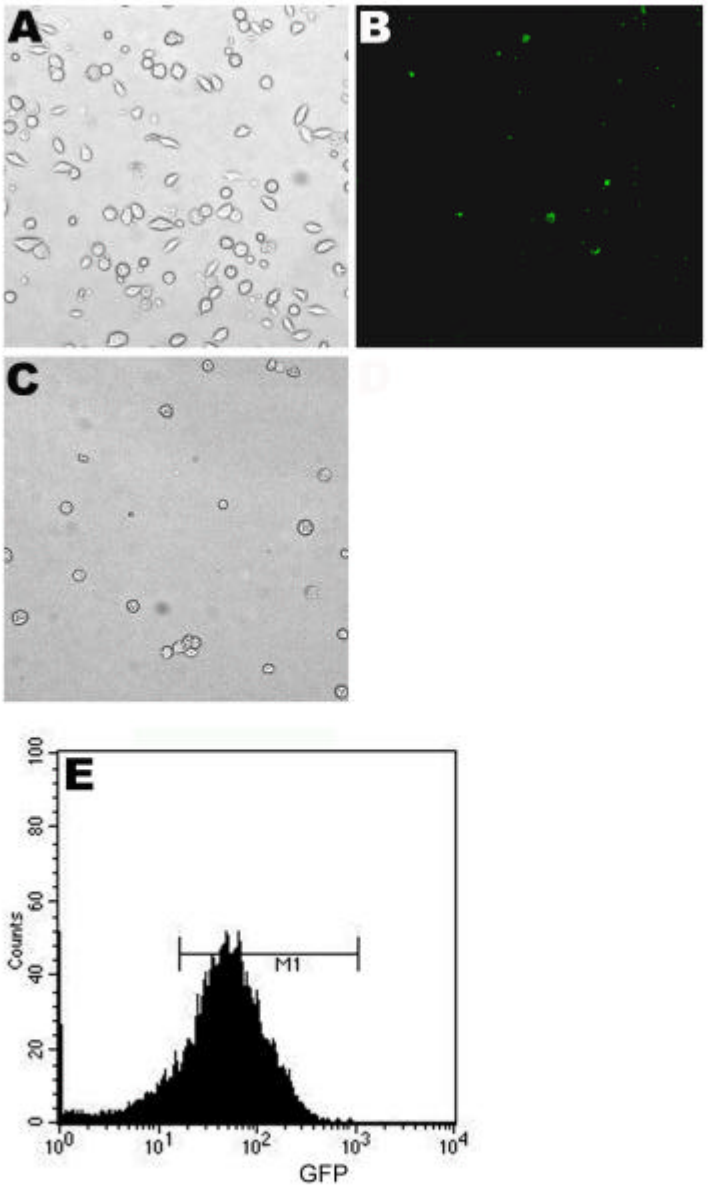


Table 1. Classification of differentially expressed genes according to molecular function.

Molecular function(s)	Number of transcripts	
	sorted	wholemount
function unknown	41	124
enzyme	15	65
kinase /phosphatase	3	6
Protein/ carbohydrate/ nucleotide binding	2	7
nucleic acid binding	12	18
transcription factor	11	14
signal transducer	1	11
transcriptional regulator	0	2
transporter	1	6
cell adhesion molecule	1	2
chaperone	0	2
structural molecule	3	4
antioxidant	0	1
Total	76	242

Figure 5. Quantitative transcript profiling of differentially expressed genes following gcm misexpression in cell sorting-based experiments

Genes differentially expressed in response to misexpression of *gcm* in the sorted experiment were grouped according to functional classes. Bars represent the fold changes in expression level of cells sorted from the wild type-like embryos versus those from the *gcm* misexpression embryos. Positive values indicate that the relative expression level of a gene is increased (upregulation) following *gcm* misexpression and negative values indicate a decrease (downregulation). Ranges of signal intensity are given for the *gcm* misexpression condition as follows: white bars, signal intensity < 100; gray bars, signal intensity from 100-1,000; black bars, signal intensity > 1,000. 17 genes selected for in situ validation are indicated in bold: *genes validated previously and **genes validated in this study.

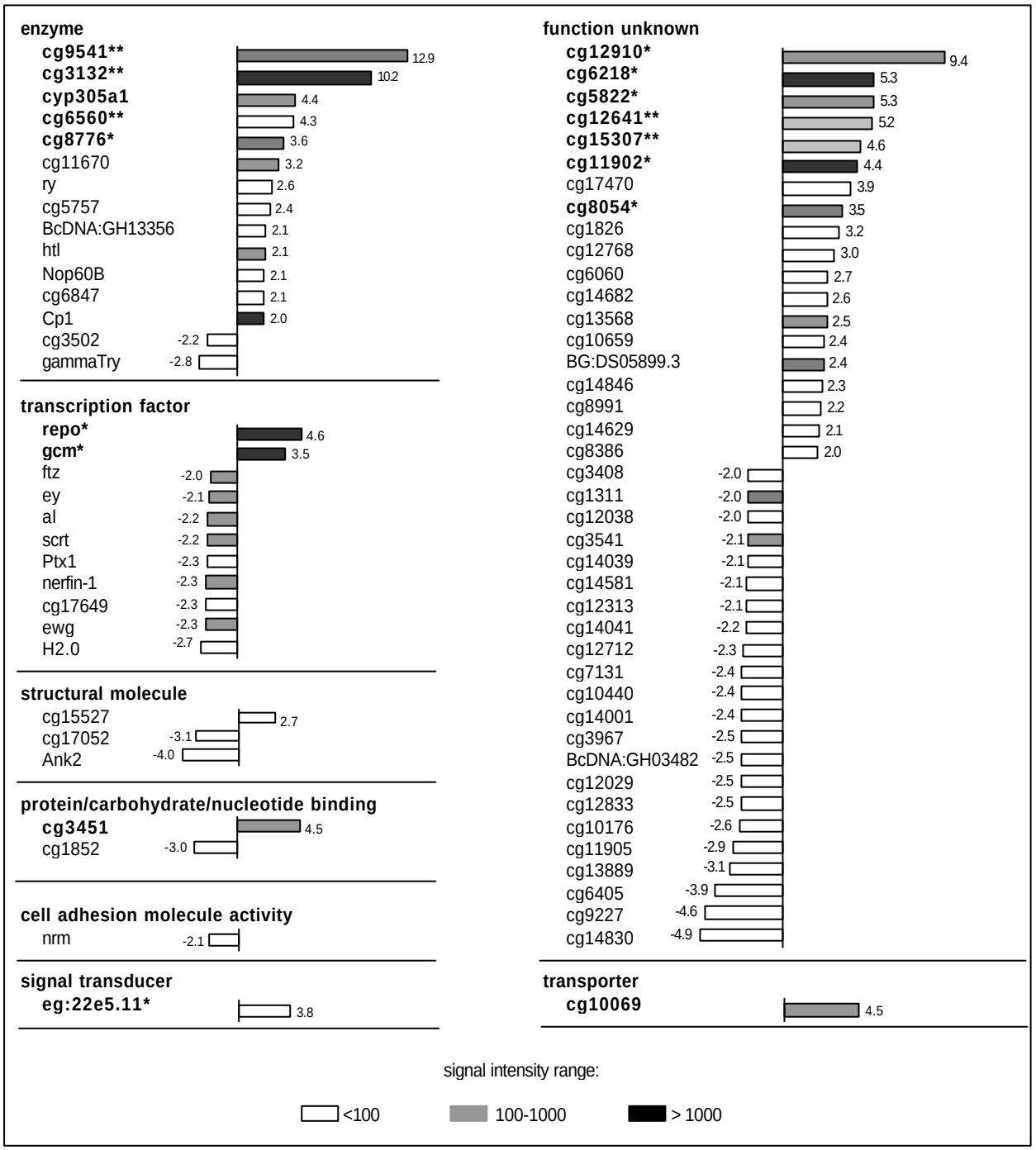

enzyme
cg9541**
cg3132**
cyp305a1
cg6560**
cg8776*
cg11670
ry
cg5757
BcDNA:GH13356
htl
Nop60B
cg6847
Cp1
cg3502
gammaTry
12.9
10.2
4.4
4.3
3.6
3.2
2.6
2.4
2.1
2.1
2.1
2.1
2.0
-2.2
-2.8
transcription factor
repo*
gcm*
ftz
ey
al
scrt
Ptx1
nerfin-1
cg17649
ewg
H2.0
4.6
3.5
-2.0
-2.1
-2.2
-2.2
-2.3
-2.3
-2.3
-2.3
-2.7
structural molecule
cg15527
cg17052
Ank2
2.7
-3.1
-4.0
protein/carbohydrate/nucleotide binding
cg3451
cg1852
4.5
-3.0
cell adhesion molecule activity
nrm
-2.1
signal transducer
eg:22e5.11*
3.8
function unknown
cg12910*
cg6218*
cg5822*
cg12641**
cg15307**
cg11902*
cg17470
cg8054*
cg1826
cg12768
cg6060
cg14682
cg13568
cg10659
BG:DS05899.3
cg14846
cg8991
cg14629
cg8386
cg3408
cg1311
cg12038
cg3541
cg14039
cg14581
cg12313
cg14041
cg12712
cg7131
cg10440
cg14001
cg3967
BcDNA:GH03482
cg12029
cg12833
cg10176
cg11905
cg13889
cg6405
cg9227
cg14830
9.4
5.3
5.3
5.2
4.6
4.4
3.9
3.5
3.2
3.0
2.7
2.6
2.5
2.4
2.4
2.3
2.2
2.1
2.0
-2.0
-2.0
-2.0
-2.1
-2.1
-2.1
-2.1
-2.1
-2.2
-2.3
-2.4
-2.4
-2.4
-2.5
-2.5
-2.5
-2.5
-2.6
-2.9
-3.1
-3.9
-4.6
-4.9
transporter
cg10069
4.5
signal intensity range:
<100
100-1000
> 1000

Table 2. Summary of gene validation of genes in the overlap of sorted and wholemount experiment

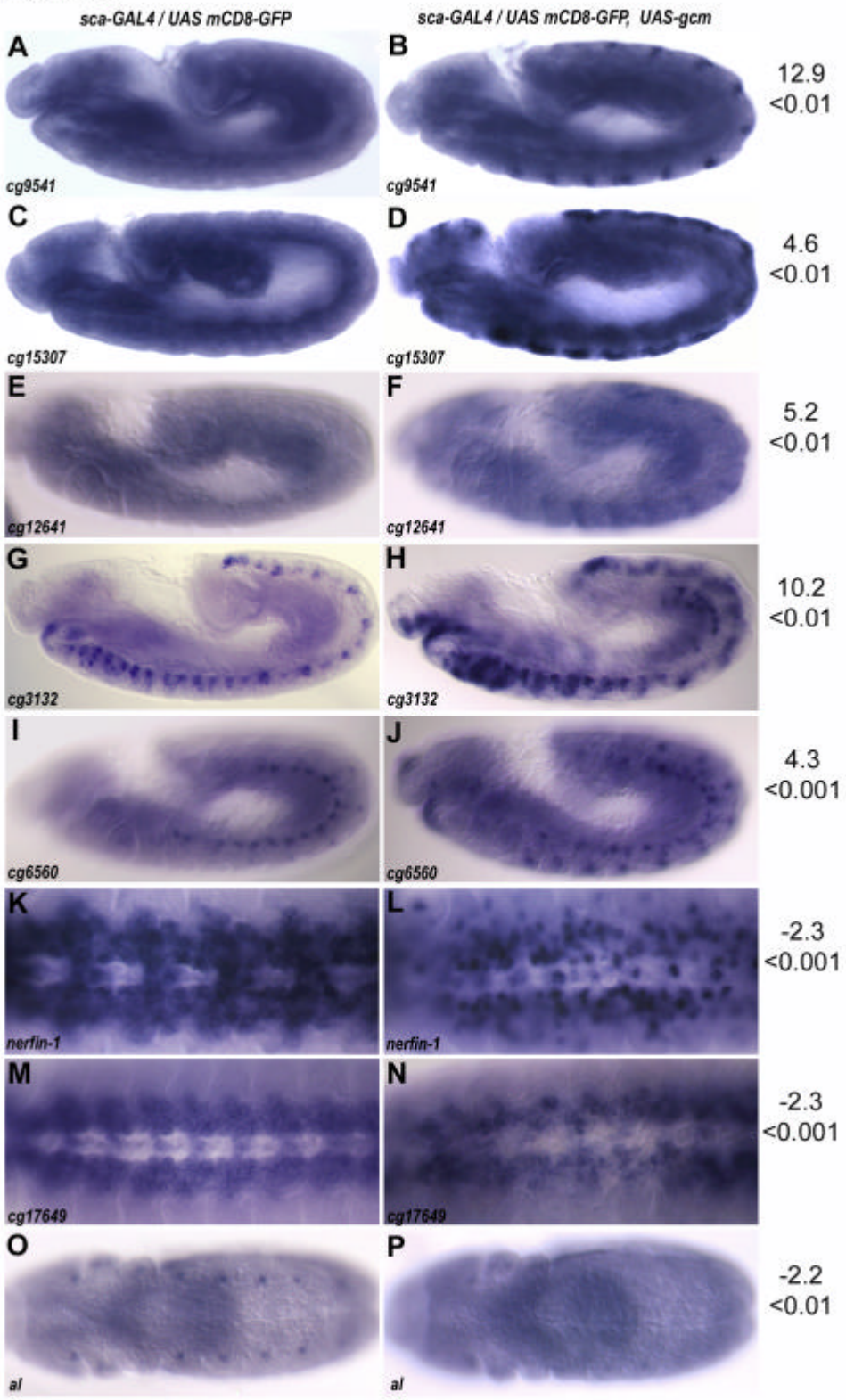
Previously: this gene has been validated as a potential *gcm* target by in situ hybridization in the previous published paper. In this study: this gene has been validated as a potential *gcm* target by in situ hybridization in this study.

Symbol	Fold changes	validation
cg3132	10.2	in this study
cg12910	9.4	previously
cg5822	5.3	previously
cg6218	5.3	previously
repo	4.6	in this study
cg6560	4.3	in this study
gcm	3.5	previously
cg3408	-2.0	--
al	-2.2	in this study
cg14041	-2.2	--
cg17649	-2.3	in this study
BcDNA:GH03482	-2.5	not validated
Nerfin-1	-2.6	in this study

Figure 6. Spatial expression of selected candidate *gcm* downstream genes by in situ hybridization

Whole-mount in situ hybridization shows expression of differentially regulated genes in wild type-like and *gcm* misexpression embryos. Lateral views (A-J) and ventral views (K-P) of stage 11 embryos, anterior is towards the left. Differential expression fold changes and p values from corresponding microarray experiments are indicated on the right. (A, C, E) The *cg9541*, *cg15307* and *cg12641* genes are expressed broadly in the neuroectoderm of wild type-like embryos; (B, D, F) *gcm* misexpression in the neuroectoderm results in ectopic expression of *cg9541*, *cg15307* and *cg12641* in specific, segmentally repeated domains of the ventral neuroectoderm. (G, I) The expression domains of the genes *cg3132* and *cg6560* are restricted to specific groups of neuroectodermal cells in the wild type-like situation. (H, J) Following *gcm* misexpression embryos, increased expression of *cg3132* and *cg6560* results around these groups of neuroectodermal cells. (K, M) The genes *nerfin-1* and *cg17649* were expressed in a comparable set of cells in the neuroectoderm in the wild type-like situation. (L, N) Following *gcm* misexpression throughout the neuroectoderm, the expression of *nerfin-1* and *cg17649* was strongly reduced or even abolished in some of these cells. (O, P) In the wild type-like embryos, a segmentally repeated pattern of *al* expression is seen in three thoracic and eight abdominal lateral patches. Following *gcm* misexpression in the neuroectoderm, the expression pattern of *al* in the abdominal patches is repressed.

Figure 5


sca-GAL4 / UAS mCD8-GFP
sca-GAL4 / UAS mCD8-GFP, UAS-gcm
A
B
cg9541
cg9541
12.9
<0.01
C
D
cg15307
cg15307
4.6
<0.01
E
F
cg12641
cg12641
5.2
<0.01
G
H
cg3132
cg3132
10.2
<0.01
I
J
cg6560
cg6560
4.3
<0.001
K
L
nerfin-1
nerfin-1
-2.3
<0.001
M
N
cg17649
cg17649
-2.3
<0.001
O
P
al
al
-2.2
<0.01

7. Impressive expressions: a systematic database of gene expression patterns in *Drosophila* embryogenesis

Haiqiong Montalta-He, Heinrich Reichert

Institute of Zoology, Biocenter/Pharmacenter, Klingelbergstrasse 50, University of Basel, CH-4056 Basel, Switzerland.

Genome Biology (2003) **4**(2): 205.

Summary

Systematic high throughput in situ **hybridization studies on whole-mount *Drosophila* embryos provide a powerful basis for determining the expression patterns of all genes in the genome and creating an integrated public resource of image-oriented gene expression data. The ongoing establishment of an** in situ **gene expression pattern database, together with new information on the reannotated *Drosophila* genome and several recent microarray-based genomic analyses of *Drosophila* development, present us with a novel, comprehensive level of resolution in developmental molecular genetics in this model system.**

These are exciting days in developmental genetics; the rapid advance of functional genomic analyses of key model systems is creating possibilities that were only scientific fantasies a few years ago. One of these scientific fantasies was to know not only all of the genes in your favorite model organism, but also the expression patterns, in time and space, of all of these genes during development. Imagine, for example, what you could do as a developmental biologist interested in the formation of midline structures if you had access to a database that revealed the identity of all of the fly genes expressed at the midline during embryogenesis. Instead of spending time and money establishing a subtractive library or doing differential display to look for midline-specific genes, you could simply go to the database and query for all the genes that are expressed at the midline. Then, knowing all potential genes of interest, you could proceed directly to the functional analysis of these genes and their genetic network. A recent report in Genome Biology on the systematic determination of patterns of gene expression during *Drosophila* embryogenesis by Tomancak et al. shows that this fantasy is rapidly becoming a reality (Tomancak et al., 2002). The systematic establishment of a gene expression database, together with a flurry of new information on the annotated *Drosophila* genome and several other recent microarray-based, functional genomic analyses of *Drosophila* development, present us with a new, comprehensive level of resolution in developmental molecular genetics (Bergman et al., 2002; Celniker et al., 2002; Hoskins et al., 2002; Kaminker et al., 2002; Lewis et al., 2002; Misra et al., 2002b; Mungall et al., 2002; Ohler et al., 2002).

Driven by advances in DNA-sequencing technology, early genomic projects were principally large-scale sequencing of whole genomes. Currently, approximately 100 genomes have been completely sequenced including 90 microbial genomes and 8 eukaryotic genomes, and the complete genome sequences of an increasing number of model organisms are now becoming available on Pubmed: http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/org.htm. A first annotated version of the *Drosophila melanogaster* genome was released in March 2000 (Adams et al., 2000; Myers et al., 2000; Reese et al., 2000). It was the first metazoan genome successfully sequenced by the whole-genome shotgun (WGS) method (Hartl, 2000; Loder, 2000). In the subsequent two years, the *Drosophila* genome has been reannotated twice, and the most recent of these reannotations, Release 3 of the *Drosophila* genome, has now been finished and is available online in Flybase (Celniker et al., 2002). Established by human curators with the

help of sophisticated new software and significantly increased amounts of experiment-derived EST and cDNA data, Release 3 provides a euchromatic sequence that is virtually gap free and of high accuracy(Hoskins et al., 2002; Kaminker et al., 2002; Misra et al., 2002a; Mungall et al., 2002; Stapleton et al., 2002). Compared to the preceding annotated Release 2, the number of genes does not change much, however, more exons and more transcripts are reported, and importantly, changes in over 40% of the predicted protein sequences have resulted(Misra et al., 2002b). It is believed that this new release of the *Drosophila* genome sequence is now a reliable resource for molecular and genetic experimentations as well as for biocomputational analysis.

With the rapid progress of genome sequencing projects, functional genomic experiments using microarrays have become powerful and popular tools to investigate biological questions at a genome-wide level. In terms of developmental biology of *Drosophila*, the adoption of microarray technology was rather slow initially, but the use of microarrays has accelerated markedly especially in the past year(Livesey, 2002). One of the first microarray-based analysis of *Drosophila* development focused on the process of metamorphosis using microarrays containing cDNAs corresponding to several thousand gene sequences and was carried out before sequence information on the entire genome became available (Ashburner et al., 1999). Similar microarrays were combined with automated embryo sorting by Furlong and colleagues to identify the targets of the transcription factor *twist* which plays a key role in mesoderm development(Furlong et al., 2001). More recently, a systematic study of gene expression throughout *Drosophila* development with microarrays has been carried out, and approximately one third of all genes were surveyed at different stages of embryos, larva, pupa and adults (Arbeitman et al., 2002). In two further recent investigations, whole-genome oligonucleotide arrays containing the entire protein coding capacity of the *Drosophila* genome (>13500 genes) have been used to study specific aspects of embryogenesis in the fly. Stathopoulos and colleagues have focused on dorsal-ventral patterning in the *Drosophila* embryo and have used whole-genome microarrays to identify targets of the transcription factor *dorsal*(Stathopoulos and Levine, 2002). Their work identified over 40 novel *dorsal* target genes as well as several new tissue-specific enchancers of *dorsal* targets. Egger and colleagues (2002) have studied gliogenesis in *Drosophila* embryos by using whole-genome

microarrays to identify downstream targets of the *glial cells missing* gene, which controls the determination of glial versus neuronal cell fate(Egger et al., 2002).

While these microarray experiments have provided a quantitative overview of changes gene expression levels across developmental time or between different experimental conditions(Spellman et al., 1998; Furlong et al., 2001; Kim, 2001), they still suffer from several limitations. Transcripts of low abundance, which are often involved in regulatory processes and thus may be of high interest for understanding development, are typically under-represented in RNA probe pools and are therefore hard to detect in microarray experiments(Chudin et al., 2002). Moreover, in multicellular organisms, cell division and differentiation leads to an increase in tissue complexity throughout development, but whole-animal microarray analysis cannot document this spatial information. One can try to isolate mRNA from every tissue at each developmental stage and then define gene expression information in different tissues at different times. However, this is a very formidable task and requires reliable methods for tissue-specific mRNA isolation and probe preparation to be established. Furthermore, false-positive results can be due to technical problems such as cross-hybridization properties of target-probe pairs or incorrect annotation of genome sequences leading to false gene model predictions(Schena et al., 1995; Chudin et al., 2002). For all of these reasons, validation of the microarray data with histological methods such as RNA in situ hybridization becomes both important and necessary. Indeed, all of the recent whole-genome microarray studies of *Drosophila* development incorporate selected in situ hybridization experiments to confirm and localize expression for a subset of the studied genes(Furlong et al., 2001; Arbeitman et al., 2002; Egger et al., 2002; Klebes et al., 2002; Stathopoulos and Levine, 2002). Given the massive quantitative expression dataset that is deriving from whole-genome microarray experiments, it now becomes important to have access to equally massive whole-genome in situ hybridization data. Ideally, one would like to have access to the expression patterns of all genes in the genome in all major embryonic tissues at all embryonic stages. This is the goal of the in situ gene expression atlas that Tomancak and colleagues are assembling online (Tomancak et al., 2002).

To achieve this formidable task, these authors have devised a high-throughput whole-mount in situ hybridization protocol in which RNA probes are generated from the set of cDNA

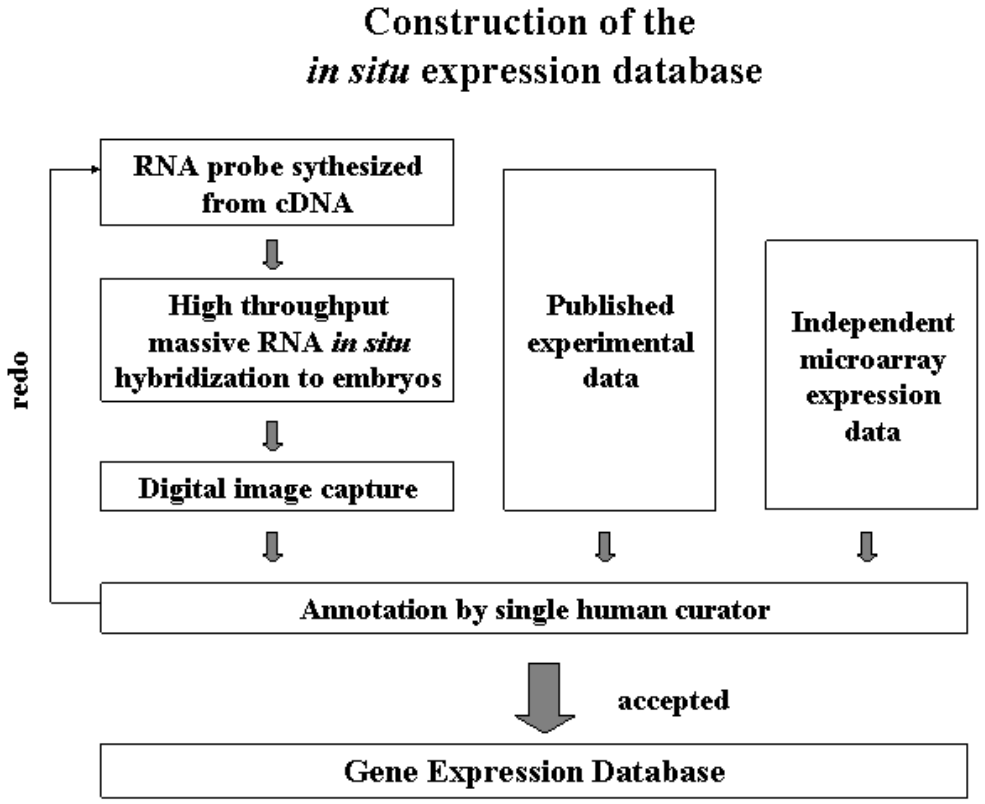
clones that comprise the *Drosophila* Gene Collections (Rubin, 2000; Stapleton et al., 2002) and then are hybridized to *Drosophila* embryos in 96-well plates. Gene expression patterns are documented by assembling digital photographs of individual embryos ordered according to developmental stage in order to visualize time-dependent expression changes. To facilitate subsequent analysis, the expression patterns of all genes are annotated by a single human curator using a controlled vocabulary that describes developmental and spatial relationships between embryonic tissues. Hierarchical clustering is further used to group together genes with similar expression patterns as well as embryonic tissues with similar sets of expressed genes. All of this data, digital images as well as annotations, are stored in a relational database and presented in a searchable form on the web: http://toy.lbl.gov:8888/cgi-bin/ex/insitu.pl). This enables any interested researcher to query the database rapidly and compare results in a rigorous manner. In addition, quantitative expression levels determined by whole-genome microarrays is obtained for each gene and each developmental stage studied, and this data is also presented along with the images and annotations of in situ expression patterns in the database, thus making a direct comparison of the two complementing data sets possible. Figure 1 shows the pipeline of the construction of the database.

Currently, over 2000 genes, representing about one sixth of all *Drosophila* genes have been examined by in situ hybridization in embryos and over 25000 digital photographs of gene expression patterns have been taken, annotated and stored in the database. Considering current production rates, the authors estimate that a first pass through the existing cDNA collections which represent about 70% of the *Drosophila* genes should be finished within a year; probes for genes that lack a suitable cDNA clone, but that show significant expression by microarray analysis will be generated by genomic PCR so that expression patterns for these genes will be determined as well (Tomancak et al., 2002). This will represent a major step towards the overall goal of determining the expression patterns of all genes in the fly genome and creating an integrated public resource of image-oriented gene expression data analogous to the repositories of DNA sequences. However, the project will not stop there, rather it will continue to be refined as more accurate information on gene sequences, coding regions, and cDNAs becomes available. Release 3 of the fly genome has already presented marked improvements in all of these areas, and regular updates of the fly genome and the in

situ expression database are planned (Bergman et al., 2002; Celniker et al., 2002; Hoskins et al., 2002; Kaminker et al., 2002; Lewis et al., 2002; Misra et al., 2002b; Mungall et al., 2002; Ohler et al., 2002; Stapleton et al., 2002; Tomancak et al., 2002).

Systematic high throughput in situ hybridization of whole-mount embryos of the type described provides a powerful method for the global survey of gene expression in embryos(Simin et al., 2002; Tomancak et al., 2002). Combined with data obtained by microarray analysis, this method makes it possible to investigate gene expression profiles in both quantitative and qualitative manner. Analysis of this type of gene expression dataset will provide a rich source of developmental genetic information and should also make it possible to identify genes involved in developmental processes that have been missed by traditional, mutagenesis-based genetic analysis. According to published estimates for flies and other animals, less than one-third of genes lead to obvious phenotypes when mutated (*Caenorhabditis elegans* Sequencing Consortium, 1998; Thatcher et al., 1998; Ashburner et al., 1999), so there is a lot that remains to be discovered. The exciting days of developmental genetics have only just begun.

Figure 1 Overview of the pipeline for the database construction.



8. Discussion

8.1 Identification of *lab* downstream genes

Homeotic genes are involved in different important developmental processes as key regulators that are highly conserved throughout evolution. It has been proposed that the they do not act directly to specify morphological differences but rather control a battery of subordinate genes encoding cellular functions directly required in differentiation (Garcia-Bellido, 1975; Pradel and White, 1998). Various strategies such as enhancer trapping, immunoprecipitation of chromatin fragments, subtractive hybridization, and heat-shock induced overexpression, have been used to search for these target genes of homeotic genes (Gehring et al., 1994b; Nasiadka and Krause, 1999; Nasiadka et al., 2000). Nevertheless, only a few genes acting downstream of homeotic proteins have been identified.

We focused on the homeotic gene *lab* in this thesis mainly because the *lab* gene and its vertebrate *Hox1* orthologs are among the best-characterized examples of evolutionary conservation of structure, expression and function of *Hox* genes during embryonic brain patterning in animal development (Hirth and Reichert, 1999).To address which and how many downstream genes are regulated by *lab*, we used a combination of *in vivo* overexpression techniques and quantitative trancription imaging with oligonucleotide arrays. *lab* was ubiquitously overexpressed in the *Drosophila* embryos by heat-shock treatment using transgenic flies carrying the *lab* gene under the control of a heat-inducible promoter. Subsequently, changes of expression profiles in response to *lab* overexpression were identified by microarray analysis.

Using inducible overexpression and quantitative transcript imaging through oligonucleotide arrays, we have identified 96 genes, whose expression levels change significantly following *lab* overexpression, as novel candidate downstream genes of *lab*. This represents 6.3% of the 1,513 identified genes represented on the oligonucleotide array. These genes belong to all functional classes represented on the array and, in consequence, cover a wide variety of molecular functions. Thus, our results identify a number of novel candidate downstream genes for lab and suggest that this homeoprotein differentially regulates a limited and distinct

set of embryonically expressed *Drosophila* genes. These findings will provide a solid basis for further genetic analysis. (Leemans et al., 2001).

8.2 Molecular basis of functional equivalence: common potential downstream genes of *otd/Otx* gene

Intensive efforts over the past two decades have revealed that homeodomain-containing transcription factors are involved in the regulation of a vast number of developmental events in different animals. Furthermore, it is now generally accepted that the majority of these factors evolved from common ancestral genes, and hence retain strong sequence and functional conservation across animal phyla. One of the clearest examples of this conservation is the *Drosophila* gene otd and it vertebrate homologs Otx1 and Otx2. These genes share a highly conserved DNA-binding motif, the homeodomain, and are almost invariably expressed in the most-anterior part of embryos which gives rise to the forebrain and midbrain (Boncinelli and Morgan, 2001). Cross-phylum gene replacement experiments have shown that the fly *otd* gene and the homologous human *Otx* genes are functionally equivalent *in vivo*, in that overexpression of either gene in *Drosophila otd* null mutants can lead to the restoration of defects in cephalic and brain development (Acampora et al., 1998; Leuzinger et al., 1998; Nagao et al., 1998; Boyl et al., 2001a).

Given the fact that otd and Otx genes share a highly conserved homeodomain, one might predict that the *in vivo* functional equivalence of *otd*/*Otx* genes, as demonstrated in the cross-phylum rescue experiments, is due to the fact that both *otd* and *Otx* genes can control a comparable set of downstream target genes, irrespective of whether the *otd*/*Otx* genes are expressed in flies or in mammals (Acampora et al., 2001b). However, currently, little is known about the downstream targets of either *otd* or *Otx* genes in flies or in mammals, and no information on common targets of *otd* and *Otx* genes is available in any species context (Acampora et al., 2001b; Boncinelli and Morgan, 2001).

In order to address this issue in a comprehensive manner we have combined cross-phylum overexpression experiments with genome-wide expression analysis based on oligonucleotide arrays in order to identify the number and identity of downstream target genes that are affected by overexpression of *otd* and of human *Otx2* in *Drosophila*. For this we used

transgenic flies, which carried either the fly *otd* gene or the human *Otx2* gene under the control of a heat-inducible promoter, to ubiquitously overexpress these transgenes in *Drosophila* embryos and then employed high-density oligonucleotide arrays representing the entire annotated fly genome for large-scale detection and quantification of induced gene expression(Lockhart et al., 1996; Lipshutz et al., 1999; Adams et al., 2000; Leemans et al., 2000; Leemans et al., 2001).

These experiments identified 287 annotated genes that showed highly significant (p=0.001) changes in expression levels in response to *otd* overexpression in *Drosophila*. Among these genes, 93 also showed highly significant differential expression changes in response to *Otx2* overexpression. Moreover the expression levels of 90 of these 93 genes were influenced in the same direction, either upregulated or downregulated, by *otd* and by *Otx2* overexpression. Thus, approximately one third of the candidate *otd* downstream target genes in *Drosophila* can be controlled in a comparable manner by the human *Otx2* gene homolog. From a genome-wide perspective, it is likely that the conserved genetic control of these common *otd*/*Otx2* downstream genes forms the molecular genetic basis for the striking *in vivo* functional equivalence of *otd* and *Otx* gene action in *Drosophila*(Leemans et al., 2001).

8.3 Methodological adaptation of oligonucleotide array experiments in developmental neurobiology

Although we have succeeded in identifying a number of novel potential downstream genes of *otd* and *lab* in the microarray experiments mentioned above, it should be noted that differential gene expression for several potential *lab* or *otd* regulated genes could be only confirmed by real time RT-PCR analysis (Lightcycler); for these genes, in situ hybridization was not able to visualize differential gene expression and validate oligonucleotide array data. In our experimental settings, overexpression of *lab* and *otd* is under control of heat-shock inducible promoter although the spatial and temporal control of endogenous expression of *lab* and *otd* is essential for the function of the genes. On the one hand, this might lead to unspecific induction of gene expression ectopically in the embryo, which would be difficult to detect by in situ hybridization if the change level is not high enough. In this case, in situ hybridization is not suitable for validation. On the other hand, differential expression might be averaged or masked throughout the embryo since *lab* and *otd* indeed function differently in

specific tissues. By taking the whole embryos for quantitative expression profiling, it may not be possible to avoid numerous false positives.

The problem of false positives is not specific to these two microarray experiments. In fact, microarray studies of the nervous system are inherently complicated by the diversity of cell populations and the lack of homogeneitiy in nervous tissue (Barlow and Lockhart, 2002; Griffin et al., 2003). Thus, cells of interest may comprise only a fraction of the entire tissue studied, anatomical divisions between regions of the nervous system are often unclear, and precisely controlled, artifact-free dissections are difficult. In consequence, averaging expression levels of entire tissue regions, or of entire embryos, may minimize or conceal even large expression changes that occur in small subpopulations of cells. This problem is aggravated in studies of neuronal development due to the small size of the embryonic nervous system and the difficulty in identifying the subpopulations of interest for dissection in embryos. In order to overcome this experimental obstacle, methods are needed that allow the isolation of specific neural cell subpopulations from embryos.

In the course of this thesis, we have developed and applied a microarray analysis based on genetic labeling techniques and magnetic cell sorting for isolating neuroectodermal cells from *Drosophila* embryos. The high spatiotemporal specificity of the *GAL4-UAS* system was used to direct expression of mCD8-GFP, a molecular label suitable for magnetic cell isolation, exclusively to the neuroectoderm. Labeled cells were then dissociated and separated using magnetic cell sorting techniques, which permitted a high rate of purification of viable cells from the neuroectoderm. Given the versatility and precision of the *GAL4-UAS* system, specific labeling of cell types or of tissue domains followed by cell sorting and microarray analysis should be possible for virtually all embryonic or postembryonic cell types and gene expression domains.

8.4 Identification of new GCM candidate downstream genes

This above-mentioned method for isolating neuroectodermal cells from *Drosophila* embryos was applied to study the potential downstream genes of *gcm*. In *Drosophila*, *gcm* functions as a genetic switch to control neuronal versus glial cell fate and is involved in the specification of nearly most differentiated glial cell types (Jones et al., 1995; Vincent et al., 1996; Jones,

2001). The initiation of gliogenesis is thought to occur through the direct activation of downstream target genes of *gcm* (Jones, 2001). In order to identify *gcm* downstream genes, we analyzed gene expression profiles following *gcm* misexpression in the neuroectoderm at an embryonic stage in which the first glial-specific genes such as *repo* (a direct *gcm* target gene) are expressed. This procedure should identify genes that act in neural precursors and are involved in the determination of glial versus neuronal cell lineage.

A total of 76 genes were detected as potential gcm downstream targets in these cell-sorted experiments. Of these, 8 genes have for the first time been validated as *gcm* targets by in situ hybridization. Two of these genes have been studied previously; *nerfin-1* which is also *called neral-finger-1* is expressed during embryonic development in neuroblasts and subsequently in ganglion mother cells (Stivers et al., 2000), and *al* is expressed in specific sensory organ anlagen during embryogenesis (Schneitz et al., 1993; Tomancak et al., 2002). An involvement of the remaining 6 genes in the embryonic nervous system of *Drosophila* has not been reported previously. However, several of these genes have gene homologs in humans, which are implicated in human nervous system development and disease. For example, the human homolog of *cg9541* encodes an adenylate kinase, which shows increased activity in the cerebrospinal fluid of patients with acute onset ischaemic brain damage and in patients with specific brain tumors (Van Rompay et al., 1999). The human homolog of *cg3132* is associated with severe cerebral degeneration and accumulation of ganglioside in neurons(Giugliani et al., 1985; Caciotti et al., 2003). The human homolog of *cg6560* encodes a protein predicted to be an ARF small monomeric GTPase belonging to the Ras GTPase superfamily which is involved in cell survival(Ehrhardt et al., 2002). While these findings are interesting and suggest that cases of molecular homology might also characterize neuroglial development in insects and mammals, a more detailed analysis involving mutant analyses in fly and murine models will be necessary before a more accurate assessment of the function and conservation of these genes can be attained.

REFERENCES

Acampora, D., Avantaggiato, V., Tuorto, F., Barone, P., Perera, M., Choo, D., Wu, D., Corte, G. and Simeone, A. (1999). Differential transcriptional control as the major molecular event in generating Otx1-/- and Otx2-/- divergent phenotypes. *Development* **126**, 1417-26.

Acampora, D., Avantaggiato, V., Tuorto, F., Barone, P., Reichert, H., Finkelstein, R. and Simeone, A. (1998). Murine *Otx1* and *Drosophila otd* genes share conserved genetic functions required in invertebrate and vertebrate brain development. *Development* **125**, 1691-702.

Acampora, D., Gulisano, M., Broccoli, V. and Simeone, A. (2001a). *Otx* genes in brain morphogenesis. *Progress in Neurobiology* **64**, 69-95.

Acampora, D., Mazan, S., Avantaggiato, V., Barone, P., Tuorto, F., Lallemand, Y., Brulet, P. and Simeone, A. (1996). Epilepsy and brain abnormalities in mice lacking the *Otx1* gene. *Nature Genetics* **14**, 218-22.

Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A. and Brulet, P. (1995). Forebrain and midbrain regions are deleted in *Otx2-/-* mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development* **121**, 3279-90.

Acampora, D., Pilo Boyl, P., Signore, M., Martinez-Barbera, J. P., Ilengo, C., Puelles, E., Annino, A., Reichert, H., Corte, G. and Simeone, A. (2001b). OTD/OTX2 functional equivalence depends on 5' and 3' UTR-mediated control of *Otx2* mRNA for nucleo-cytoplasmic export and epiblast-restricted translation. *Development* **128**, 4801-13.

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

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

Akiyama-Oda, Y., Hosoya, T. and Hotta, Y. (1998). Alteration of cell fate by ectopic expression of *Drosophila glial cells missing* in non-neural cells. *Development Genes and Evolution* **208**, 578-85.

Anderson, D. J. (2001). Stem cells and pattern formation in the nervous system: the possible versus the actual. *Neuron* **30**, 19-35.

Andrew, D. J. and Scott, M. P. (1992). Downstream of the homeotic genes. *New Biol* **4**, 5-15.

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

Ang, S. L., Jin, O., Rhinn, M., Daigle, N., Stevenson, L. and Rossant, J. (1996). A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. *Development* **122**, 243-52.

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

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9.

Ashburner, M., Misra, S., Roote, J., Lewis, S. E., Blazej, R., Davis, T., Doyle, C., Galle, R., George, R., Harris, N. et al. (1999). An exploration of the sequence of a 2.9-Mb region of the genome of Drosophila melanogaster: the Adh region. *Genetics* **153**, 179-219.

Barbosa, V., Yamamoto, R. R., Henderson, D. S. and Glover, D. M. (2000). Mutation of a Drosophila gamma tubulin ring complex subunit encoded by discs degenerate-4 differentially disrupts centrosomal protein localization. *Genes Dev* **14**, 3126-39.

Barlow, C. and Lockhart, D. J. (2002). DNA arrays and neurobiology--what's new and what's next? *Curr Opin Neurobiol* **12**, 554-61.

Benos, P. V., Gatt, M. K., Murphy, L., Harris, D., Barrell, B., Ferraz, C., Vidal, S., Brun, C., Demaille, J., Cadieu, E. et al. (2001). From first base: the sequence of the tip of the X chromosome of Drosophila melanogaster, a comparison of two sequencing strategies. *Genome Res* **11**, 710-30.

Bergman, C., Pfeiffer, B., Rincón-Limas, D., Hoskins, R., Gnirke, A., Mungall, C., Wang, A., Kronmiller, B., Pacleb, J., Park, S. et al. (2002). Assessing the impact of comparative genomic sequence data on the functional annotation of the Drosophila genome. *Genome Biol* **3**, research0086.1-0086.20.

Bernardoni, R., Miller, A. A. and Giangrande, A. (1998). Glial differentiation does not require a neural ground state. *Development* **125**, 3189-200.

Bernardoni, R., Vivancos, V. and Giangrande, A. (1997). glide/gcm is expressed and required in the scavenger cell lineage. *Dev Biol* **191**, 118-30.

Bertram, M. J. and Pereira-Smith, O. M. (2001). Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif. *Gene* **266**, 111-21.

Biggin, M. D. and McGinnis, W. (1997). Regulation of segmentation and segmental identity by *Drosophila* homeoproteins: the role of DNA binding in functional activity and specificity. *Development* **124**, 4425-33.

Boncinelli, E. and Morgan, R. (2001). Downstream of Otx2, or how to get a head. *Trends Genet* **17**, 633-6.

Bossing, T., Udolph, G., Doe, C. Q. and Technau, G. M. (1996). The embryonic central nervous system lineages of *Drosophila melanogaster*. I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Developmental Biology* **179**, 41-64.

Botas, J. and Auwers, L. (1996). Chromosomal binding sites of Ultrabithorax homeotic proteins. *Mech Dev* **56**, 129-38.

Boyl, P. P., Signore, M., Acampora, D., Martinez-Barbera, J. P., Ilengo, C., Annino, A., Corte, G. and Simeone, A. (2001a). Forebrain and midbrain development requires epiblast-restricted Otx2 translational control mediated by its 3' UTR. *Development* **128**, 2989-3000.

Boyl, P. P., Signore, M., Annino, A., Barbera, J. P., Acampora, D. and Simeone, A. (2001b). Otx genes in the development and evolution of the vertebrate brain. *Int J Dev Neurosci* **19**, 353-63.

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

Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G. G. et al. (2003). ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* **31**, 68-71.

Brown, P. O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat Genet* **21**, 33-7.

Bryant, Z., Subrahmanyan, L., Tworoger, M., LaTray, L., Liu, C. R., Li, M. J., van den Engh, G. and Ruohola-Baker, H. (1999a). Characterization of differentially expressed genes in purified *Drosophila* follicle cells: toward a general strategy for cell type- specific developmental analysis. *Proc Natl Acad Sci U S A* **96**, 5559-64.

Bryant, Z., Subrahmanyan, L., Tworoger, M., LaTray, L., Liu, C. R., Li, M. J., van den Engh, G. and Ruohola-Baker, H. (1999b). Characterization of differentially expressed genes in purified Drosophila follicle cells: toward a general strategy for cell type-specific developmental analysis. *Proc Natl Acad Sci U S A* **96**, 5559-64.

Caciotti, A., Bardelli, T., Cunningham, J., D'Azzo, A., Zammarchi, E. and Morrone, A. (2003). Modulating action of the new polymorphism L436F detected in the GLB1 gene of a type-II GM1 gangliosidosis patient. *Hum Genet* **113**, 44-50.

***Caenorhabditis elegans* Sequencing Consortium.** (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. *Science* **282**, 2012-8.

Campbell, G. (2002). Distalization of the Drosophila leg by graded EGF-receptor activity. *Nature* **418**, 781-5.

Campbell, G., Goring, H., Lin, T., Spana, E., Andersson, S., Doe, C. Q. and Tomlinson, A. (1994). RK2, a glial-specific homeodomain protein required for embryonic nerve cord condensation and viability in *Drosophila*. *Development* **120**, 2957-66.

Campos-Ortega, J. and Hartenstein, V. (1997). The embryonic development of *Drosophila melanogaster*. Heidelberg: Springer.

Campos-Ortega, J. a. H., V. (1997). The Embryonic Development of *Drosophila melanogaster*. *Heidelberg: Springer*.

Capecchi, M. (1997). The role of Hox genes in hindbrain development. In *Molecular and cellular approaches to neural development*, (ed. W. Cowan T. Jessel and S. Zipursky), pp. 334-355. Oxford: Oxford University Presse.

Carr, A. and Biggin, M. D. (1999). A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila* embryos. *Embo J* **18**, 1598-608.

Carroll, S. B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479-85.

Cecchi, C. (2002). *Emx2*: a gene responsible for cortical development, regionalization and area specification. *Gene* **291**, 1-9.

Celniker, S. E., Wheeler, D. A., Kronmiller, B., Carlson, J. W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S. P., Frise, E. et al. (2002). Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol* **3**, RESEARCH0079.

Certa, U., de Saizieu, A. and Mous, J. (2001). Hybridization analysis of labeled RNA by oligonucleotide arrays. *Methods Mol Biol* **170**, 141-56.

Chan, S. K. and Mann, R. S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. *Proc Natl Acad Sci U S A* **93**, 5223-8.

Chudin, E., Walker, R., Kosaka, A., Wu, S. X., Rabert, D., Chang, T. K. and Kreder, D. E. (2002). Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. *Genome Biol* **3**, RESEARCH0005.

Cohen, S. M. and Jurgens, G. (1990). Mediation of Drosophila head development by gap-like segmentation genes. *Nature* **346**, 482-5.

Cohen, S. X., Moulin, M., Hashemolhosseini, S., Kilian, K., Wegner, M. and Müller, C. W. (2003). Structure of the GCM domain-DNA complex: a DNA-binding domain with a novel fold and mode of target site recognition. *Embo J* **22**, 1835-45.

Consortium, T. F. (1999). The FlyBase database of the Drosophila Genome Projects and community literature. *Nucleic Acids Res* **27**, 85-8.

Craig, E. A., Ingolia, T. D. and Manseau, L. J. (1983). Expression of *Drosophila* heat-shock cognate genes during heat shock and development. *Dev Biol* **99**, 418-26.

Cribbs, D. L., Benassayag, C., Randazzo, F. M. and Kaufman, T. C. (1995). Levels of homeotic protein function can determine developmental identity: evidence from low-level expression of the Drosophila homeotic gene proboscipedia under Hsp70 control. *Embo J* **14**, 767-78.

Crow, J. F. (1999). Unmasking a cheating gene. *Science* **283**, 1651-2.

Cutforth, T. and Rubin, G. M. (1994). Mutations in *Hsp83* and *cdc37* impair signaling by the *sevenless* receptor tyrosine kinase in *Drosophila*. *Cell* **77**, 1027-36.

DeSimone, S. M. and White, K. (1993). The Drosophila erect wing gene, which is important for both neuronal and muscle development, encodes a protein which is similar to the sea urchin P3A2 DNA binding protein. *Mol Cell Biol* **13**, 3641-9.

Deyholos, M. K. and Galbraith, D. W. (2001). High-density microarrays for gene expression analysis. *Cytometry* **43**, 229-38.

Diederich, R. J., Merrill, V. K., Pultz, M. A. and Kaufman, T. C. (1989). Isolation, structure, and expression of *labial*, a homeotic gene of the Antennapedia Complex involved in *Drosophila* head development. *Genes Dev* **3**, 399-414.

Dobie, K. W., Hari, K. L., Maggert, K. A. and Karpen, G. H. (1999). Centromere proteins and chromosome inheritance: a complex affair. *Curr Opin Genet Dev* **9**, 206-17.

Doe, C. Q., Smouse, D. and Goodman, C. S. (1988). Control of neuronal fate by the *Drosophila* segmentation gene *even-skipped*. *Nature* **333**, 376-8.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev Suppl*, 135-42.

Duboule, D. and Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet* **10**, 358-64.

Egger, B., Leemans, R., Loop, T., Kammermeier, L., Fan, Y., Radimerski, T., Strahm, M. C., Certa, U. and Reichert, H. (2002). Gliogenesis in *Drosophila*: genome-wide analysis of downstream genes of *glial cells missing* in the embryonic nervous system. *Development* **129**, 3295-309.

Ehrhardt, A., Ehrhardt, G. R., Guo, X. and Schrader, J. W. (2002). Ras and relatives--job sharing and networking keep an old family together. *Exp Hematol* **30**, 1089-106.

Feger, G. (1999). Identification and complete cDNA sequence of the missing Drosophila MCMs: DmMCM3, DmMCM6 and DmMCM7. *Gene* **227**, 149-55.

Ferrante, A. W., Jr., Reinke, R. and Stanley, E. R. (1995). Shark, a Src homology 2, ankyrin repeat, tyrosine kinase, is expressed on the apical surfaces of ectodermal epithelia. *Proc Natl Acad Sci U S A* **92**, 1911-5.

Field, C. M. and Alberts, B. M. (1995). Anillin, a contractile ring protein that cycles from the nucleus to the cell cortex. *J Cell Biol* **131**, 165-78.

Finkelstein, R. and Boncinelli, E. (1994). From fly head to mammalian forebrain: the story of otd and Otx. *Trends Genet* **10**, 310-5.

Finkelstein, R. and Perrimon, N. (1990). The *orthodenticle* gene is regulated by *bicoid* and *torso* and specifies *Drosophila* head development [see comments]. *Nature* **346**, 485-8.

Finkelstein, R., Smouse, D., Capaci, T. M., Spradling, A. C. and Perrimon, N. (1990). The *orthodenticle* gene encodes a novel homeo domain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures. *Genes Dev* **4**, 1516-27.

Freeman, M. R., Delrow, J., Kim, J., Johnson, E. and Doe, C. Q. (2003). Unwrapping glial biology: Gcm target genes regulating glial development, diversification, and function. *Neuron* **38**, 567-80.

Furlong, E. E., Andersen, E. C., Null, B., White, K. P. and Scott, M. P. (2001). Patterns of gene expression during Drosophila mesoderm development. *Science* **293**, 1629-33.

Galliot, B. and Miller, D. (2000). Origin of anterior patterning. How old is our head? *Trends Genet* **16**, 1-5.

Gallitano-Mendel, A. and Finkelstein, R. (1997). Novel segment polarity gene interactions during embryonic head development in Drosophila. *Dev Biol* **192**, 599-613.

Gao, Q., Wang, Y. and Finkelstein, R. (1996). *Orthodenticle* regulation during embryonic head development in *Drosophila*. *Mech Dev* **56**, 3-15.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. *Ciba Found Symp* **29**, 161-82.

Gavalas, A., Studer, M., Lumsden, A., Rijli, F. M., Krumlauf, R. and Chambon, P. (1998). *Hoxa1* and *Hoxb1* synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. *Development* **125**, 1123-36.

Gehring, W. J., Affolter, M. and Burglin, T. (1994a). Homeodomain proteins. *Annu Rev Biochem* **63**, 487-526.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G. and Wuthrich, K. (1994b). Homeodomain-DNA recognition. *Cell* **78**, 211-23.

Gellon, G. and McGinnis, W. (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *Bioessays* **20**, 116-25.

Giesen, K., Hummel, T., Stollewerk, A., Harrison, S., Travers, A. and Klämbt, C. (1997). Glial development in the *Drosophila* CNS requires concomitant activation of glial and repression of neuronal differentiation genes. *Development* **124**, 2307-16.

Giugliani, R., Dutra, J. C., Pereira, M. L., Rotta, N., Drachler Mde, L., Ohlweiller, L., Pina Neto, J. M., Pinheiro, C. E. and Breda, D. J. (1985). GM1 gangliosidosis: clinical and laboratory findings in eight families. *Hum Genet* **70**, 347-54.

Graba, Y., Aragnol, D. and Pradel, J. (1997). *Drosophila* Hox complex downstream targets and the function of homeotic genes. *Bioessays* **19**, 379-88.

Granderath, S., Stollewerk, A., Greig, S., Goodman, C. S., O'Kane, C. J. and Klämbt, C. (1999). *loco* encodes an RGS protein required for *Drosophila* glial differentiation. *Development* **126**, 1781-91.

Granjeaud, S., Bertucci, F. and Jordan, B. R. (1999). Expression profiling: DNA arrays in many guises. *Bioessays* **21**, 781-90.

Grieder, N. C., Marty, T., Ryoo, H. D., Mann, R. S. and Affolter, M. (1997). Synergistic activation of a *Drosophila* enhancer by HOM/EXD and DPP signaling. *Embo J* **16**, 7402-10.

Griffin, R. S., Mills, C. D., Costigan, M. and Woolf, C. J. (2003). Exploiting microarrays to reveal differential gene expression in the nervous system. *Genome Biol* **4**, 105.

Grossniklaus, U., Cadigan, K. M. and Gehring, W. J. (1994). Three maternal coordinate systems cooperate in the patterning of the *Drosophila* head. *Development* **120**, 3155-71.

Gunawardena, S. and Goldstein, L. S. (2001). Disruption of axonal transport and neuronal viability by amyloid precursor protein mutations in Drosophila. *Neuron* **32**, 389-401.

Haass, C., Klein, U. and Kloetzel, P. M. (1990). Developmental expression of *Drosophila melanogaster* small heat-shock proteins. *J Cell Sci* **96**, 413-8.

Hartl, D. L. (2000). Fly meets shotgun: shotgun wins. *Nat Genet* **24**, 327-8.

Hartmann, B., Hirth, F., Walldorf, U. and Reichert, H. (2000). Expression, regulation and function of the homeobox gene *empty spiracles* in brain and ventral nerve cord development of *Drosophila*. *Mech Dev* **90**, 143-53.

Hartmann, B. and Reichert, H. (1998). The genetics of embryonic brain development in *Drosophila*. *Mol Cell Neurosci* **12**, 194-205.

Hartmann, B., Reichert, H. and Walldorf, U. (2001). Interaction of gap genes in the Drosophila head: tailless regulates expression of empty spiracles in early embryonic patterning and brain development. *Mech Dev* **109**, 161-72.

Heck, M. M., Pereira, A., Pesavento, P., Yannoni, Y., Spradling, A. C. and Goldstein, L. S. (1993). The kinesin-like protein KLP61F is essential for mitosis in Drosophila. *J Cell Biol* **123**, 665-79.

Henning, K. A., Peterson, C., Legerski, R. and Friedberg, E. C. (1994). Cloning the *Drosophila* homolog of the *xeroderma pigmentosum complementation group C* gene reveals homology between the predicted human and *Drosophila* polypeptides and that encoded by the yeast *RAD4* gene. *Nucleic Acids Res* **22**, 257-61.

Henry, G., Zito, K. and Dubnau, J. (2003). Chipping away at brain funciton: mining for insights with microarrays. *Curr Opin Neurobiol.* **13**, 1-7.

Heuer, J. G. and Kaufman, T. C. (1992). Homeotic genes have specific functional roles in the establishment of the *Drosophila* embryonic peripheral nervous system. *Development* **115**, 35-47.

Hirth, F., Hartmann, B. and Reichert, H. (1998). Homeotic gene action in embryonic brain development of *Drosophila*. *Development* **125**, 1579-89.

Hirth, F., Loop, T., Egger, B., Miller, D. F. B., T.C., K. and Reichert, H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of *Drosophila*. *Development* **128**, 4781-788.

Hirth, F. and Reichert, H. (1999). Conserved genetic programs in insect and mammalian brain development. *Bioessays* **21**, 677-84.

Hirth, F., Therianos, S., Loop, T., Gehring, W. J., Reichert, H. and Furukubo-Tokunaga, K. (1995). Developmental defects in brain segmentation caused by mutations of the homeobox genes *orthodenticle* and *empty spiracles* in *Drosophila*. *Neuron* **15**, 769-78.

Holland, L. Z. and Holland, N. D. (1999). Chordate origins of the vertebrate central nervous system. *Curr Opin Neurobiol* **9**, 596-602.

Hoppler, S. and Bienz, M. (1994). Specification of a single cell type by a Drosophila homeotic gene. *Cell* **76**, 689-702.

Hoskins, R., Smith, C., Carlson, J., Carvalho, A., Halpern, A., Kaminker, J., Kennedy, C., Mungall, C., Sullivan, B., Sutton, G. et al. (2002). Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome Biol* **3**, research0085.1-0085.16.

Hosoya, T., Takizawa, K., Nitta, K. and Hotta, Y. (1995). *glial cells missing*: a binary switch between neuronal and glial determination in *Drosophila*. *Cell* **82**, 1025-36.

Ishikawa, T., Igarashi, T., Hata, K. and Fujita, T. (1999). *c-fos* induction by heat, arsenite, and cadmium is mediated by a heat shock element in its promoter. *Biochem Biophys Res Commun* **254**, 566-71.

Ji, W., Cai, L., Wright, M. B., Walker, G., Salgam, P., Vater, A. and Lindpaintner, K. (2000). Preservation of gene expression ratios among multiple complex cDNAs after PCR amplification: application to differential gene expression studies. *J Struct Funct Genomics* **1**, 1-7.

Jones, B. W. (2001). Glial cell development in the *Drosophila* embryo. *Bioessays* **23**, 877-87.

Jones, B. W., Fetter, R. D., Tear, G. and Goodman, C. S. (1995). *glial cells missing*: a genetic switch that controls glial versus neuronal fate. *Cell* **82**, 1013-23.

Kablar, B., Vignali, R., Menotti, L., Pannese, M., Andreazzoli, M., Polo, C., Giribaldi, M. G., Boncinelli, E. and Barsacchi, G. (1996). Xotx genes in the developing brain of Xenopus laevis. *Mech Dev* **55**, 145-58.

Kain, K. C., Orlandi, P. A. and Lanar, D. E. (1991). Universal promoter for gene expression without cloning: expression-PCR. *Biotechniques* **10**, 366-74.

Kaminker, J., Bergman, C., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D., Lewis, S., Rubin, G. et al. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol* **3**, research0084.1-0084.20.

Kammermeier, L., Leemans, R., Hirth, F., Flister, S., Wenger, U., Walldorf, U., Gehring, W. J. and Reichert, H. (2001). Differential expression and function of the *Drosophila* Pax6 genes *eyeless* and *twin of eyeless* in embryonic central nervous system development. *Mech. Dev.* **103**, 71-8.

Kaufman, T. C., Seeger, M. A. and Olsen, G. (1990). Molecular and genetic organization of the *antennapedia* gene complex of *Drosophila* melanogaster. *Adv Genet* **27**, 309-62.

Kim, S. K. (2001). Http://C. elegans: mining the functional genomic landscape. *Nat Rev Genet* **2**, 681-9.

Klaes, A., Menne, T., Stollewerk, A., Scholz, H. and Klambt, C. (1994). The Ets transcription factors encoded by the Drosophila gene pointed direct glial cell differentiation in the embryonic CNS. *Cell* **78**, 149-60.

Klämbt, C. (1993). The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells. *Development* **117**, 163-76.

Klämbt, C. and Goodman, C. S. (1991). Role of the midline glia and neurons in the formation of the axon commissures in the central nervous system of the *Drosophila* embryo. *Ann. N. Y. Acad. Sci.* **633**, 142-59.

Klambt, C., Jacobs, J. R. and Goodman, C. S. (1991). The midline of the Drosophila central nervous system: a model for the genetic analysis of cell fate, cell migration, and growth cone guidance. *Cell* **64**, 801-15.

Klebes, A., Biehs, B., Cifuentes, F. and Kornberg, T. B. (2002). Expression profiling of *Drosophila* imaginal discs. *Genome Biol* **3**, RESEARCH0038.

Kraemer, C., Weil, B., Christmann, M. and Schmidt, E. R. (1998). The new gene *DmX* from *Drosophila melanogaster* encodes a novel WD-repeat protein. *Gene* **216**, 267-76.

Krumlauf, R. (1994). Hox genes in vertebrate development. *Cell* **78**, 191-201.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O. and Davis, R. W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A* **94**, 13057-62.

Lawrence, P. A. and Morata, G. (1994). Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* **78**, 181-9.

Ledent, V. and Vervoort, M. (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Res* **11**, 754-70.

Lee, J. K., Coyne, R. S., Dubreuil, R. R., Goldstein, L. S. and Branton, D. (1993). Cell shape and interaction defects in alpha-spectrin mutants of Drosophila melanogaster. *J Cell Biol* **123**, 1797-809.

Lee, T. and Luo, L. (1999). Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451-61.

Leemans, R., Egger, B., Loop, T., Kammermeier, L., He, H., Hartmann, B., Certa, U., Hirth, F. and Reichert, H. (2000). Quantitative transcript imaging in normal and heat-shocked Drosophila embryos by using high-density oligonucleotide arrays. *Proc Natl Acad Sci U S A* **97**, 12138-43.

Leemans, R., Loop, T., Egger, B., He, H., Kammermeier, L., Hartmann, B., Certa, U., Reichert, H. and Hirth, F. (2001). Identification of candidate downstream genes for the homeodomain transcription factor Labial in Drosophila through oligonucleotide-array transcript imaging. *Genome Biol* **2**.

Leuzinger, S., Hirth, F., Gerlich, D., Acampora, D., Simeone, A., Gehring, W. J., Finkelstein, R., Furukubo-Tokunaga, K. and Reichert, H. (1998). Equivalence of the fly *orthodenticle* gene and the human *OTX* genes in embryonic brain development of *Drosophila*. *Development* **125**, 1703-10.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565-70.

Lewis, S., Searle, S., Harris, N., Gibson, M., lyer, V., Richter, J., Wiel, C., Bayraktaroglir, L., Birney, E., Crosby, M. et al. (2002). Apollo: a sequence annotation editor. *Genome Biol* **3**, research0082.1-0082.14.

Liang, Z. and Biggin, M. D. (1998). *Eve* and *ftz* regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* **125**, 4471-82.

Lindquist, S. and Craig, E. A. (1988). The heat-shock proteins. *Annu Rev Genet* **22**, 631-77.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat Genet* **21**, 20-4.

Livesey, R. (2002). Have microarrays failed to deliver for developmental biology? *Genome Biol* **3**, comment2009.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**, 1675-80.

Loder, N. (2000). Celera's shotgun approach puts Drosophila in the bag. *Nature* **403**, 817.

Lutz, B., Lu, H. C., Eichele, G., Miller, D. and Kaufman, T. C. (1996). Rescue of *Drosophila labial* null mutant by the chicken ortholog *Hoxb-1* demonstrates that the function of Hox genes is phylogenetically conserved. *Genes Dev* **10**, 176-84.

Luzzi, V., Mahadevappa, M., Raja, R., Warrington, J. A. and Watson, M. A. (2003). Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. *J Mol Diagn* **5**, 9-14.

Maconochie, M., Nonchev, S., Morrison, A. and Krumlauf, R. (1996). Paralogous Hox genes: function and regulation. *Annu Rev Genet* **30**, 529-56.

Mahadevappa, M. and Warrington, J. A. (1999). A high-density probe array sample preparation method using 10- to 100- fold fewer cells. *Nat Biotechnol* **17**, 1134-6.

Manak, J. R. and Scott, M. P. (1994). A class act: conservation of homeodomain protein functions. *Dev Suppl*, 61-77.

Mann, R. S. (1995). The specificity of homeotic gene function. *Bioessays* **17**, 855-63.

Mann, R. S. and Affolter, M. (1998). Hox proteins meet more partners. *Curr Opin Genet Dev* **8**, 423-9.

Mann, R. S. and Chan, S. K. (1996). Extra specificity from *extradenticle*: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet* **12**, 258-62.

Mann, R. S. and Morata, G. (2000). The developmental and molecular biology of genes that subdivide the body of *Drosophila*. *Annu Rev Cell Dev Biol* **16**, 243-71.

Mannervik, M. (1999). Target genes of homeodomain proteins. *Bioessays* **21**, 267-70.

Mastick, G. S., McKay, R., Oligino, T., Donovan, K. and Lopez, A. J. (1995). Identification of target genes regulated by homeotic proteins in *Drosophila melanogaster* through genetic selection of Ultrabithorax protein-binding sites in yeast. *Genetics* **139**, 349-63.

Matsuo, I., Kuratani, S., Kimura, C., Takeda, N. and Aizawa, S. (1995). Mouse Otx2 functions in the formation and patterning of rostral head. *Genes Dev* **9**, 2646-58.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302.

Merrill, V. K., Diederich, R. J., Turner, F. R. and Kaufman, T. C. (1989). A genetic and developmental analysis of mutations in *labial*, a gene necessary for proper head formation in *Drosophila melanogaster*. *Dev Biol* **135**, 376-91.

Michaud, S., Marin, R. and Tanguay, R. M. (1997). Regulation of heat shock gene induction and expression during *Drosophila* development. *Cell Mol Life Sci* **53**, 104-13.

Micklem, D. R. (1995). mRNA localisation during development. *Dev Biol* **172**, 377-95.

Mills, J. C., Roth, K. A., Cagan, R. L. and Gordon, J. I. (2001). DNA microarrays and beyond: completing the journey from tissue to cell. *Nat Cell Biol* **3**, E175-8.

Misra, S., Crosby, M., Mungall, C., Matthews, B., Campbell, K., Hradecky, P., Huang, Y., Kaminker, J., Millburn, G., Prochnik, S. et al. (2002a). Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. *Genome Biol* **3**, research0083.1-0083.22.

Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E. et al. (2002b). Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol* **3**, RESEARCH0083.

Mlodzik, M., Fjose, A. and Gehring, W. J. (1988). Molecular structure and spatial expression of a homeobox gene from the *labial* region of the Antennapedia-complex. *Embo J* **7**, 2569-78.

Mohler, J. (1995). Spatial regulation of segment polarity gene expression in the anterior terminal region of the *Drosophila* blastoderm embryo. *Mech Dev* **50**, 151-61.

Montalta-He, H., Fan, Y., Egger, B., Certa, U. and Reichert, H. (2003). Isolation of genetically labeled cells from the neuroectoderm of *Drosophila* embryos for genome-wide microarray analysis of gcm downstream genes in gliogenesis. *Submitted.*

Montalta-He, H., Leemans, R., Loop, T., Strahm, M., Certa, U., Primig, M., Acampora, D., Simeone, A. and Reichert, H. (2002). Evolutionary conservation of *otd*/*Otx2* transcription factor action: a genome-wide microarray analysis in *Drosophila*. *Genome Biol* **3**, RESEARCH0015.

Montalta-He, H. and Reichert, H. (2003). Impressive expressions: developing a systematic database of gene-expression patterns in *Drosophila* embryogenesis. *Genome Biol* **4**, 205.

Morata, G. (1993). Homeotic genes of *Drosophila*. *Curr Opin Genet Dev* **3**, 606-14.

Morata, G. and Sanchez-Herrero, E. (1999). Patterning mechanisms in the body trunk and the appendages of *Drosophila*. *Development* **126**, 2823-8.

Mungall, C., Misra, S., Berman, B., Carlson, J., Frise, E., Harris, N., Marshall, B., Shu, S., Kaminker, J., Prochnik, S. et al. (2002). An integrated computational pipeline and database to support whole-genome sequence annotation. *Genome Biol* **3**, research0081.1-0081.11.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A. et al. (2000). A whole-genome assembly of Drosophila. *Science* **287**, 2196-204.

Nagao, T., Leuzinger, S., Acampora, D., Simeone, A., Finkelstein, R., Reichert, H. and Furukubo-Tokunaga, K. (1998). Developmental rescue of Drosophila cephalic defects by the human Otx genes. *Proc Natl Acad Sci U S A* **95**, 3737-42.

Nasiadka, A., Grill, A. and Krause, H. M. (2000). Mechanisms regulating target gene selection by the homeodomain- containing protein Fushi tarazu. *Development* **127**, 2965-76.

Nasiadka, A. and Krause, H. M. (1999). Kinetic analysis of segmentation gene interactions in *Drosophila* embryos. *Development* **126**, 1515-26.

Nover, L. and Scharf, K. D. (1997). Heat stress proteins and transcription factors. *Cell. Mol. Life Sci.* **53**, 80-103.

Ohler, U., Liao, G., Niemann, H. and Rubin, G. (2002). Computational analysis of core promoters in the Drosophila genome. *Genome Biol* **3**, research0087.1-0087.12.

Pankratz, M. J., Jäckle, H. (1993). In *The Development of Drosophila melanogaster*, vol. 1. New York: Cold Spring Harbor Lab. Press, Plainview,.

Parkes, T. L., Elia, A. J., Dickinson, D., Hilliker, A. J., Phillips, J. P. and Boulianne, G. L. (1998). Extension of Drosophila lifespan by overexpression of human SOD1 in motorneurons. *Nat Genet* **19**, 171-4.

Pauli, D. T. (1990). In *Stress Proteins in Biology and Medicine*. New York: Cold Spring Harvor Lab. Press.

Philip, A. V. (1998). Mitotic sister-chromatid separation: what Drosophila mutants can tell us. *Trends Cell Biol* **8**, 150.

Pöpperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd*/*pbx*. *Cell* **81**, 1031-42.

Pradel, J. and White, R. A. (1998). From selectors to realizators. *Int J Dev Biol* **42**, 417-21.

Prokopenko, S. N., He, Y., Lu, Y. and Bellen, H. J. (2000). Mutations affecting the development of the peripheral nervous system in Drosophila: a molecular screen for novel proteins. *Genetics* **156**, 1691-715.

Rajagopalan, D. (2003). A comparison of statistical methods for analysis of high density oligonucleotide array data. *Bioinformatics* **19**, 1469-76.

Reese, M. G., Hartzell, G., Harris, N. L., Ohler, U., Abril, J. F. and Lewis, S. E. (2000). Genome annotation assessment in Drosophila melanogaster. *Genome Res* **10**, 483-501.

Regel, R., Matioli, S. R. and Terra, W. R. (1998). Molecular adaptation of Drosophila melanogaster lysozymes to a digestive function. *Insect Biochem Mol Biol* **28**, 309-19.

Reichert, H. (2002). Conserved genetic mechanisms for embryonic brain patterning. *Int J Dev Biol* **46**, 81-7.

Reichert, H. and Simeone, A. (1999). Conserved usage of gap and homeotic genes in patterning the CNS. *Curr Opin Neurobiol* **9**, 589-95.

Reichert, H. and Simeone, A. (2001). Developmental genetic evidence for a monophyletic origin of the bilaterian brain. *Philos Trans R Soc Lond B Biol Sci* **356**, 1533-44.

Roark, M., Sturtevant, M. A., Emery, J., Vaessin, H., Grell, E. and Bier, E. (1995). *scratch*, a pan-neural gene encoding a zinc finger protein related to snail, promotes neuronal development. *Genes Dev* **9**, 2384-98.

Roy, P. J., Stuart, J. M., Lund, J. and Kim, S. K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. *Nature* **418**, 975-9.

Royet, J. and Finkelstein, R. (1995). Pattern formation in Drosophila head development: the role of the orthodenticle homeobox gene. *Development* **121**, 3561-72.

Rubin, D. M., Mehta, A. D., Zhu, J., Shoham, S., Chen, X., Wells, Q. R. and Palter, K. B. (1993). Genomic structure and sequence analysis of *Drosophila melanogaster* HSC70 genes. *Gene* **128**, 155-63.

Rubin, G. M. (2000). Biological annotation of the Drosophila genome sequence. *Novartis Found Symp* **229**, 79-82; discussion 82-3.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. and Pendleton, J. W. (1994). Evolution of Hox genes. *Annu Rev Genet* **28**, 423-42.

Safarik, I. and Safarikova, M. (1999). Use of magnetic techniques for the isolation of cells. *J Chromatogr B Biomed Sci Appl* **722**, 33-53.

Saitoe, M., Schwarz, T. L., Umbach, J. A., Gundersen, C. B. and Kidokoro, Y. (2001). Absence of junctional glutamate receptor clusters in Drosophila mutants lacking spontaneous transmitter release. *Science* **293**, 514-7.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-70.

Schlesinger, M. J. (1990). Heat shock proteins. *J Biol Chem* **265**, 12111-4.

Schmid, A., Chiba, A. and Doe, C. Q. (1999). Clonal analysis of *Drosophila* embryonic neuroblasts: neural cell types, axon projections and muscle targets. *Development* **126**, 4653-89.

Schmidt, H., Rickert, C., Bossing, T., Vef, O., Urban, J. and Technau, G. M. (1997). The embryonic central nervous system lineages of *Drosophila melanogaster*. II. Neuroblast lineages derived from the dorsal part of the neuroectoderm. *Dev Biol* **189**, 186-204.

Schmidt-Ott, U., Gonzalez-Gaitan, M., Jackle, H. and Technau, G. M. (1994). Number, identity, and sequence of the Drosophila head segments as revealed by neural elements and their deletion patterns in mutants. *Proc Natl Acad Sci U S A* **91**, 8363-7.

Schneitz, K., Spielmann, P. and Noll, M. (1993). Molecular genetics of Aristaless, a prd-type homeo box gene involved in the morphogenesis of proximal and distal pattern elements in a subset of appendages in Drosophila. *Genes Dev* **7**, 911.

Sharman, A. C. and Brand, M. (1998). Evolution and homology of the nervous system: cross-phylum rescues of otd/Otx genes. *Trends Genet* **14**, 211-4.

Shen, J., Wu, H. and Gudas, L. J. (2000). Molecular cloning and analysis of a group of genes differentially expressed in cells which overexpress the *Hoxa-1* homeobox gene. *Exp Cell Res* **259**, 274-83.

Shinozaki, K., Miyagi, T., Yoshida, M., Miyata, T., Ogawa, M., Aizawa, S. and Suda, Y. (2002). Absence of Cajal-Retzius cells and subplate neurons associated with defects of tangential cell

migration from ganglionic eminence in Emx1/2 double mutant cerebral cortex. *Development* **129**, 3479-92.

Simeone, A. (1998). *Otx1* and *Otx2* in the development and evolution of the mammalian brain. *Embo J* **17**, 6790-8.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. and Boncinelli, E. (1992a). Nested expression domains of four homeobox genes in developing rostral brain. *Nature* **358**, 687-90.

Simeone, A., Acampora, D., Mallamaci, A., Stornaiuolo, A., D'Apice, M. R., Nigro, V. and Boncinelli, E. (1993). A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo. *Embo J* **12**, 2735-47.

Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M. and Boncinelli, E. (1992b). Two vertebrate homeobox genes related to the Drosophila empty spiracles gene are expressed in the embryonic cerebral cortex. *Embo J* **11**, 2541-50.

Simeone, A., Puelles, E. and Acampora, D. (2002). The Otx family. *Curr Opin Genet Dev* **12**, 409-15.

Simin, K., Scuderi, A., Reamey, J., Dunn, D., Weiss, R., Metherall, J. E. and Letsou, A. (2002). Profiling patterned transcripts in Drosophila embryos. *Genome Res* **12**, 1040-7.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* **9**, 3273-97.

Sprecher, S. G. (2003). The urbilaterian brain: developmental insights into the evolutionary origin of the brain in insects and vertebrates. *Arthropod Structure & Development* **32**, 141-156.

St Johnston, D. and Nusslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201-19.

Stapleton, M., Carlson, J., Brokstein, P., Yu, C., Champe, M., George, R., Guarin, H., Kronmiller, B., Pacleb, J., Park, S. et al. (2002). A Drosophila full-length cDNA resource. *Genome Biol* **3**, research0080.1-0080.8.

Stathopoulos, A. and Levine, M. (2002). Dorsal gradient networks in the *Drosophila* embryo. *Dev Biol* **246**, 57-67.

Stivers, C., Brody, T., Kuzin, A. and Odenwald, W. F. (2000). Nerfin-1 and -2, novel Drosophila Zn-finger transcription factor genes expressed in the developing nervous system. *Mech Dev* **97**, 205-10.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F. M., Chambon, P. and Krumlauf, R. (1998). Genetic interactions between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. *Development* **125**, 1025-36.

Su, T. T. and O'Farrell, P. H. (1998). Size control: cell proliferation does not equal growth. *Curr Biol* **8**, R687-9.

Suda, Y., Matsuo, I., Kuratani, S. and Aizawa, S. (1996). Otx1 function overlaps with Otx2 in development of mouse forebrain and midbrain. *Genes Cells* **1**, 1031-44.

Tautz, D. and Pfeifle, C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* **98**, 81-5.

Thatcher, J. W., Shaw, J. M. and Dickinson, W. J. (1998). Marginal fitness contributions of nonessential genes in yeast. *Proc Natl Acad Sci U S A* **95**, 253-7.

Therianos, S., Leuzinger, S., Hirth, F., Goodman, C. S. and Reichert, H. (1995). Embryonic development of the *Drosophila* brain: formation of commissural and descending pathways. *Development* **121**, 3849-60.

Thor, S. (1995). The genetics of brain development: conserved programs in flies and mice. *Neuron* **15**, 975-7.

Tietjen, I., Rihel, J. M., Cao, Y., Koentges, G., Zakhary, L. and Dulac, C. (2003). Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161-75.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E. et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biol* **3**, RESEARCH0088.

Treisman, J. E. (1999). A conserved blueprint for the eye? *Bioessays* **21**, 843-50.

Van Rompay, A. R., Johansson, M. and Karlsson, A. (1999). Identification of a novel human adenylate kinase. cDNA cloning, expression analysis, chromosome localization and characterization of the recombinant protein. *Eur J Biochem* **261**, 509-17.

Vazquez, J., Pauli, D. and Tissieres, A. (1993). Transcriptional regulation in *Drosophila* during heat shock: a nuclear run-on analysis. *Chromosoma* **102**, 233-48.

Vincent, S., Vonesch, J. L. and Giangrande, A. (1996). Glide directs glial fate commitment and cell fate switch between neurones and glia. *Development* **122**, 131-9.

Wakimoto, B. T. (1998). Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. *Cell* **93**, 321-4.

Walldorf, U. and Gehring, W. J. (1992). *empty spiracles*, a gap gene containing a homeobox involved in *Drosophila* head development. *Embo J* **11**, 2247-59.

Wallrath, L. L. (1998). Unfolding the mysteries of heterochromatin. *Curr Opin Genet Dev* **8**, 147-53.

Walter, M. F., Petersen, N. S. and Biessmann, H. (1990). Heat shock causes the collapse of the intermediate filament cytoskeleton in Drosophila embryos. *Dev Genet* **11**, 270-9.

Wang, X. Z., Kuroda, M., Sok, J., Batchvarova, N., Kimmel, R., Chung, P., Zinszner, H. and Ron, D. (1998). Identification of novel stress-induced genes downstream of *chop*. *Embo J* **17**, 3619-30.

White, K. P., Rifkin, S. A., Hurban, P. and Hogness, D. S. (1999). Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**, 2179-84.

Wieschaus, E., Perrimon, N. and Finkelstein, R. (1992). orthodenticle activity is required for the development of medial structures in the larval and adult epidermis of Drosophila. *Development* **115**, 801-11.

Wimmer, E. A., Cohen, S. M., Jackle, H. and Desplan, C. (1997). *buttonhead* does not contribute to a combinatorial code proposed for *Drosophila* head development. *Development* **124**, 1509-17.

Wimmer, E. A., Simpson-Brose, M., Cohen, S. M., Desplan, C. and Jackle, H. (1995). Trans- and cis-acting requirements for blastodermal expression of the head gap gene *buttonhead*. *Mech Dev* **53**, 235-45.

Xiong, W. C. and Montell, C. (1995). Defective glia induce neuronal apoptosis in the *repo* visual system of *Drosophila*. *Neuron* **14**, 581-90.

Yang, X., Li, D. M., Chen, W. and Xu, T. (2001). Human homologue of Drosophila lats, LATS1, negatively regulate growth by inducing G(2)/M arrest or apoptosis. *Oncogene* **20**, 6516-23.

Yang, Y. H. and Speed, T. (2002). Design issues for cDNA microarray experiments. *Nat Rev Genet* **3**, 579-88.

Younossi-Hartenstein, A., Green, P., Liaw, G. J., Rudolph, K., Lengyel, J. and Hartenstein, V. (1997). Control of early neurogenesis of the *Drosophila* brain by the head gap genes *tll*, *otd*, *ems*, and *btd*. *Dev Biol* **182**, 270-83.

Younossi-Hartenstein, A., Nassif, C., Green, P. and Hartenstein, V. (1996). Early neurogenesis of the *Drosophila* brain. *J Comp Neurol* **370**, 313-29.

Zakin, L., Reversade, B., Virlon, B., Rusniok, C., Glaser, P., Elalouf, J. M. and Brulet, P. (2000). Gene expression profiles in normal and Otx2-/- early gastrulating mouse embryos. *Proc Natl Acad Sci U S A* **97**, 14388-93.

Zimowska, G., Aris, J. P. and Paddy, M. R. (1997). A Drosophila Tpr protein homolog is localized both in the extrachromosomal channel network and to nuclear pore complexes. *J Cell Sci* **110**, 927-44.

ACKNOWLEDGEMENTS

I am grateful that my supervisor, Heinrich Reichert, offered me this opportunity to work in his group for my PhD so that I can realize my dream to experience *Drosophila* as a wonderful animal model. It is great that Heinrich has been supporting and open for discussion during these years which has been of great help for me to grow scientifically. Many thanks go to Esther Stoeckli and Volker Schmid for serving in my thesis committee.

I would like to say thanks to:

- Former and present members in the Reichert lab: Bruno Bello for his very inspiring discussion and who let me know more about British than French; The real Swiss colleagues, Boris Egger and Martin Müller from whom I learned to do things slowly but with excellent quality; Yun Fan, my Chinese colleague in the lab; Susanne Flister for interesting discussions and her very experienced suggestions; Filomena Forte for whose work our flies survive; Beate Hartmann for her kindness and patient in explaining me questions; Frank Hirth for his critical views and stimulating attitude; Lars Kammermeier, who makes sure that we have lunch on time and is always helping as much as he can; Ronny Leemans who took care of me on the first day in the lab and from whom I got to know the western culture much better: anybody touch that, just kill him!! Thomas Loop who, hopefully, is not bored by my questions about computer and software; Pascale Piguet who has been helping me a lot in English writing and encouraging me always; Urs Stiefel our lab manager; last but not least, the youngsters in the lab: Robert Lichtneckert, Stefan Meier, Simon Sprecher, Christopher Schweizer, Niggi Holbro, Regina Miller, Eva Bönzli and Bettina Roth.
- People in the Schimd group, especially Nathalie Yanze for her technical support and discussions, and people in the Stoeckli group.
- People on the second floor, especially the swimming gang --Urs Kloter, Makiko Seimiya and Grieder Nicole, for technical supports as well as interesting and cheering discussions.
- Hans Vogt, Leo Faletti, Eleniana Petitjean, Christine Parry, Keith Potter, Jsabelle Altherr, Ingrid Singh for their work making the lab functioning.
- Friends who share invaluable friendship and make my life here so colorful and enjoyable in these years.

My thanks specially go to my parents --it is they who grew me up and made it possible for me to come here for my study -- as well as my relatives for their supports in these years.

Very uniquely, it is really a highlight in my life that I met Stefan and his love has been supporting me ever since.

Curriculum Vitae

Haiqiong MONTALTA-HE

Personal Data

Date of birth	Sept. 16th, 1973
Place of birth	Guangzhou, Guangdong, People's Republic of China
Marital Status	Married
Nationality	Chinese

Education

Oct. 1999 – present **Ph.D.Candidate** (Supervisor: Prof. Heinrich Reichert)
Developmental Neurobiology, Basel University, Switzerland.

Sep. 1996 –Jul. 1999 **M.S.** (Supervisor: Prof. Jinfa Wang)
Molecular Biology, Zhongshan University, China

Sep. 1992 – Jul. 1996 **B.S.** (Supervisor: Prof. Jinfa Wang)
Biochemistry, Zhongshan University, China

Teaching Experience

2001-2003	Supervisor for the Block Course of Neruobiologie in the biocenter
2002	Supervisor in the ELTEM Summer School of Neurogenomics
2000-2001	Tutor for the course of Biology in University of Basel (Code: 7320 Einführung in Die Biologie)
1997-1998	Organizing and teaching the course of Basic Experiments in Genetic Engineering (Zhongshan University, China)

Invited Talk

9th Biennial meeting of Neurobiology of *Drosophila*, 2001, New York, USA
1st Neurex ELTEM meeting on Drosophila Neurobiology, 2002, Engelberg

Poster Presentation at Conferences

43rd annual *Drosophila* Research Conference, 2002, San Diego, U.S.A.
4th meeting of the German Neuroscience Society- 28th Göttingen Neurobiology Conference, 2001, Göttingen, Germany
18th International Congress of Genetics, 1998, Beijing, China

Publicaitons

Research paper

Montalta-He H., Fan Y., Egger B., Certa U. and Reichert H. (2003) Isolation of genetically labeled cells from the neuroectoderm of *Drosophila* embryos for genome-wide microarray analysis of *gcm* downstream genes in gliogenesis. *Submitted*

Montalta-He H, Leemans R, Loop T, Strahm M., Certa U, Primig M., Acampora D., Simeone A. and Reichert H. (2002) Evolutionary conservation of *otd/Otx2* transcription factor action: a genome-wide microarray analysis in *Drosophila*. *Genome Biol.* 3(4): RESEARCH0015.1-0015.15.

Leemans R, Loop T, Egger B, **He H**, Kammermeier L, Hartmann B, Certa U, Reichert H, Hirth F. (2001) Identification of candidate downstream genes for the homeodomain transcription factor Labial in Drosophila through oligonucleotide-array transcript imaging. *Genome Biol.* 2(5): RESEARCH0015

Leemans R, Egger B, Loop T, Kammermeier L, **He H**, Hartmann B, Certa U, Hirth F, Reichert H. (2000). Quantitative transcript imaging in normal and heat-shocked Drosophila embryos by using high-density oligonucleotide arrays. *Proc Natl Acad Sci USA*;97(22): 12138-43

Wang JF, Li YK, **He H**, Chen ZJ, Liu LS. (2000). Promoter Function of a Rice Repetitive Sequence RRD3 in Transgenic Plants. *The Journal of Plant.* 10.

Wang JF, **He H**, Li YK, Wang HB, Liu LS. (2000). The promoter activity of a rice-derived repetitive sequence RRD3 in CHO and 3T3 cell lines. *The Journal of Biochemistry and Molecular Biology in China*. 6.

He H, Wang JF, Chen Y, Li HJ, Tu GH. (1998). The bioactivity assay of recombinant human Epidermal Growth Factor (rhEGF). *Guangdong Medical Journal* 19(5): 324-325.

Review Paper

Montalta-He H, Reichert H. (2003). Impressive expressions: a systematic database of gene expression patterns in *Drosophila* embryogenesis. *Genome Biol.*, **4**(2): 205.

He H, Wang JF. (1998). The study of the structure and gene regulation of human aromatase gene, CYP19. *Progress of Biochemistry and Biophysics*. 25(6): 35-37.

Books

Wang JF, **He H**, Chen ZJ, Yang L, He XL, Yi HF, Wang HB. (2000). Collection of exercise of Molecular and Cell Biology. *The Science Press.*

Wang JF, **He H**, Chen ZJ, Yang L, He XL, Yi HF, Wang HB. (1999). Nomenclature Guide of Genetics. *The Science Press.*

Wang JF, **He H**, Chen Y, Shen HX, Chen ZJ. (1996). Basic experiments for the principle of gene engineering. *Zhongshan University Press.*

ERKLÄRUNG

Ich erkläre, dass ich die Dissertation

FUNCTIONAL GENOMIC ANALYSIS OF DEVELOPMENTAL CONTROL GENE ACTION IN THE EMBRYONIC NERVOUS SYSTEM OF DROSOPHILA

nur mit der darin angegebenen Hilfe verfasst und bei keiner anderen Fakultät eingereicht habe.

Basel, den 06. October 2003

Haiqiong Montalta-He