# Functional genomics of brain development and developmentally related brain disease in *Drosophila*

## Inauguraldissertation

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

Thomas Loop
aus Deutschland

Basel 2004

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

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Heinrich Reichert und Prof. Dr. Karl-Friedrich Fischbach.

Basel, den 21.09.2004

Prof. Dr. Marcel Tanner

# Contents

## 1. Summary

**One of the fundamental challenges in basic neuroscience is to understand the molecular genetic networks associated with building the brain. As malfunction in these genetic pathways can lead to disorders like cancer, brain development is also a crucial research area for clinical neuroscience. In the course of this thesis, different molecular aspects of Drosophila brain development and related neoplastic disease were analyzed using high-density oligonucleotide arrays.**

**The homeotic selector gene *labial* (*lab*) plays an important role in specification of neuronal identity in the embryonic brain of *Drosophila*. In *labial* mutants presumptive neurons in the posterior tritocerebrum fail to differentiate. This leads to severe defects in tritocerebral axon pathways.**

**Using high density oligonucleotide arrays we identified downstream target genes of Labial and showed that only a limited and distinct set of genes expressed in the embryo is regulated by this homeoprotein. Furthermore, we performed genetic rescue experiments to analyze the functional equivalence of Drosophila *Hox* gene products in specification of the tritocerebral neuromere. Surprisingly, all tested homeotic proteins, with the exception of *Abd-B*, were able to rescue the *labial* mutant phenotype in the tritocerebrum. These results indicate that the specificity of homeotic gene action in embryonic brain development has to be modulated by cis-acting regulatory elements.**

**Another study circled around the homeobox transcription factor *otd* and its human homolog *Otx2*. Cross-phylum rescue experiments have shown that these genes are functionally equivalent. We used quantitative transcript imaging to analyze *otd* and *Otx* gene action in the *Drosophila* embryo at a genomic level. Our experiments suggest that about one third of the Otd-regulated transcripts in *Drosophila* can also be controlled by the human Otx2. 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*.**

***glial cells missing* (*gcm*) is a key control gene of gliogenesis. *gcm* loss-of-function leads to a transformation of glial cells into neurons and, conversely, when *gcm* is ectopically misexpressed, presumptive neurons become glia. Since *gcm* encodes a transcription factor it is supposed that a set of downstream genes are regulated by GCM that in turn execute the glial differentiation program. Again, a set of full-genome transcript profiling experiments was conducted to identify *gcm* downstream genes in a comprehensive manner. A set of several hundred candidate *gcm* target genes were identified in this screen, giving new insights into neuroglial fate specification in *Drosophila.***

**Brain tumors have been extensively studied by looking at genetic alterations and mutations that lead to malignant growth. Still, the causes of brain tumorigenesis are largely unknown. Model systems like *Drosophila* can be of great help to shed light on altered transcriptional activity in brain tumor phenotypes.**

**To investigate the *in vivo* transcriptional activity associated with a brain tumor, we conducted genome-wide microarray expression analyses of an adult brain tumor in *Drosophila* caused by homozygous mutation in the tumor suppressor gene *brain tumor* (*brat*). Two independent gene expression studies using two different oligonucleotide**

**microarray platforms were used to compare the transcriptome of adult wildtype flies with mutants displaying the adult *brat*$^{k06028}$ mutant brain tumor. Cross-validation and stringent statistical criteria identified a core transcriptional signature of *brat*$^{k06028}$ neoplastic tissue. We found highly significant expression level changes for 321 annotated genes associated with the adult neoplastic *brat*$^{k06028}$ tissue indicating elevated and aberrant metabolic and cell cycle activity, upregulation of the basal transcriptional machinery, as well as elevated and aberrant activity of ribosome synthesis and translation control. One fifth of these genes show homology to known mammalian genes involved in cancer formation. These results identify for the first time the genome-wide transcriptional alterations associated with an adult brain tumor in *Drosophila* and reveal insights into the possible mechanisms of tumor formation caused by homozygous mutation of the translational repressor *brat.***

# 2. Introduction

## 2.1. *Drosophila* as a model system

*Drosophila melanogaster* has been an important genetic model system for nearly one hundred years. The wealth of genetic, molecular and transgenic tools available today, combined with a short lifecycle of the species and low maintenance cost makes it a superior system to answer a wide variety of biological questions.

Studies conducted on the vinegar fly have elucidated many of the fundamental principles of eukaryotic genetics. The application of a wide range of genetic tools used in *Drosophila* research led to a large amount of well characterized genetic mutants that are available to the research community as stable lines from a number of stock centers. Technologies for the efficient production of genetic mosaics and transposon based methods of altering genes facilitate the identification and functional analysis of genes and their epigenetic interactions.

Comparison of data assembled with molecular cloning and analysis of genes in *Drosophila* with functional genetic data from mammalian systems has shown that a surprisingly large number of *Drosophila* genes have homologs with similar function in mammalian systems. These include transcription factors and their regulatory targets, structural proteins, chromosomal proteins, ion channels, and signaling proteins. The evolutionary conservation extends to higher-level processes, like development (Riddle and Tabin, 1999), behavior, sleep (Hendricks et al., 2000), and responses to drugs at the physiological level (Moore et al., 1998).

## 2.2. The development of the Drosophila brain

The work reported here revolves around the *Drosophila* brain. As it is still largely a mystery how such a complex organ is generated through genetic mechanisms (Fig. 1), research on brain development is a challenging task.

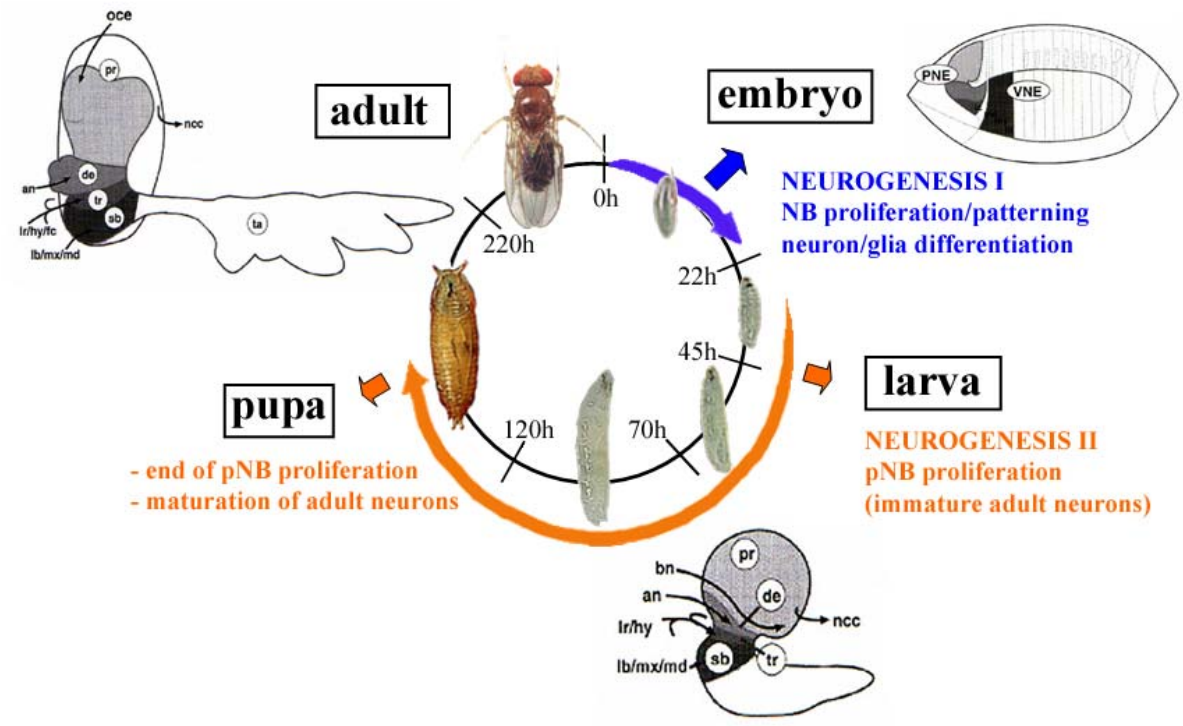



**Fig.1.** The life cycle of *Drosophila*. The middle panel shows the different stages throughout development from the embryo via the larva and pupa to the adult. The timings are shown in the circle. The panel at the upper right shows a schematic picture of an early embryo after patterning of the neuroectoderm; PNE protocerebral neuroectoderm, VNE ventral neuroectoderm . On the lower right there is a schematic representation of a larval CNS; pr protocerebrum, de deuterocerebrum, tr tritocerebrum, sb subesophageal ganglion with the labial, maxillary and mandibular neuromeres, bn Bolwig´s nerve, an antennal nerve, lr/hy labral/hypopharyngeal nerve, ncc nerve to corpora cardiaca . The schematic on the upper left shows the adult CNS; oce ocellar nerve (modified after Hirth).

The embryonic brain is composed of a supraesophageal ganglion lying anteriorly and a posterior subesophageal ganglion. The supraesophageal ganglion can be divided into the protocerebrum (b1), deutocerebrum (b2) and tritocerebrum (b3), and the gnathal subesophageal ganglion can be separated into the mandibular (s1), maxillary (s2) and labial (s3) neuromeres (Reichert and Boyan, 1997; Therianos et al., 1995; Younossi-Hartenstein et al., 1996). This layout can also be found in the adult brain with the mouth parts being innervated by sensory and motor nerves from the subesophageal ganglion and the tritocerebrum, which also innervates the stomatogastric ganglion. The deutocerebrum, being the major olfactory center, receives the antennal nerves and projects onto the motorneurons of the ventral nervous system (VNC). The protocerebrum constitutes the biggest part of the brain and contains the mushroom bodies (corpora pedunculata) and their major afferent tract (antennocerebral tract), the central complex, pars intercerebralis, optic tubercles, and optic lobes (Nassif et al., 1998).

The mushroom bodies in the dorsal protocerebrum are considered to be a center for learning and memory and are composed of Kenyon cells whose axons form characteristic structures called calyx, peduncle and the α, β, and γ lobes. More medially lies the pars intercerebralis containing neurosecretory cells and interconnecting the basal brain regions and the VNC with axons that form the median bundle. The central complex, thought to be involved in flight

control, is an unpaired structure situated between the mushroom body calyxes. The central complex receives input from the ventral body at the base of the protocerebrum. These structures form the so-called forebrain which is bordered on both sides by the optic lobes. An optic lobe can be divided into lamina and medulla lying distally and the inner optic neuropil called lobula complex. The compound eyes project their axons into the lamina. Lamina and medulla project into the lobula complex, which innervated its contralateral counterpart as well as the lateral midbrain (Nassif et al., 1998).
How the complex adult brain is formed starting from the embryonic brainanlage through the larval and pupal stages of *Drosophila* is becoming a subject of detailed study.

Early in the embryo zygotically expressed genes interact in patterning the germ layers. The dorsoventral neural patterning in *Drosophila* relies on three signaling pathways. The *dorsal* (*dl*) signaling cascade is acting in formation of the ventral mesoderm and neuroectoderm, while *dpp* signaling defines the dorsal border of the neurogenic region. *Epidermal growth factor receptor* (*Egfr*) signaling is crucial for the ventral and intermediate neuroectoderm specification (Fig. 2) (Cornell and Ohlen, 2000).

The transcription factor *dorsal* is a member of the Rel/NF kappa B family and acts in the onset of dorsoventral patterning (Steward, 1987). Its gene product is located in the cytoplasm of oocytes and is transported into the nucleus after fertilization. In the early embryo a nuclear concentration gradient of DL can be found along the dorsoventral axis. The DL gradient initiates differentiation of the mesoderm, the neuroectoderm and the dorsal ectoderm by concentration-dependent regulation of several zygotically active target genes.
The mesodermal genes *twist* and *snail* are activated by high concentrations of DL on the ventral side, leading to a repression of neuroectoderm formation.
Lower levels of DL gives rise to neuroectoderm and are required to activate neural gene expression. One of the direct target genes in the neurogenic domain seems to be *short gastrulation* (*sog*). The DL gradient also functions as a context-dependent repressor that restricts the expression of genes like *decapentaplegic* (*dpp*) to dorsal regions (Stathopoulos et al., 2002).
The *dpp* gene product is a member of the transforming growth factor beta (TGF-beta) superfamily and has a role as an inhibitor of neural tissue formation. In the early embryo it is localised in the dorsal-most 40% of nuclei, defining the dorsal border of the presumptive neuroectoderm. Other essential roles are establishment of dorsal embryonic tissues including dorsal ectoderm. *dpp* loss-of-function leads to expansion of the neurogenic neuroectoderm at the expense of other dorsal tissues (Wharton et al., 1993). In the gain-of-function situation *dpp* can induce dorsal structures and inhibit neurogenic tissue formation (Ferguson and Anderson, 1992; Wharton et al., 1993).

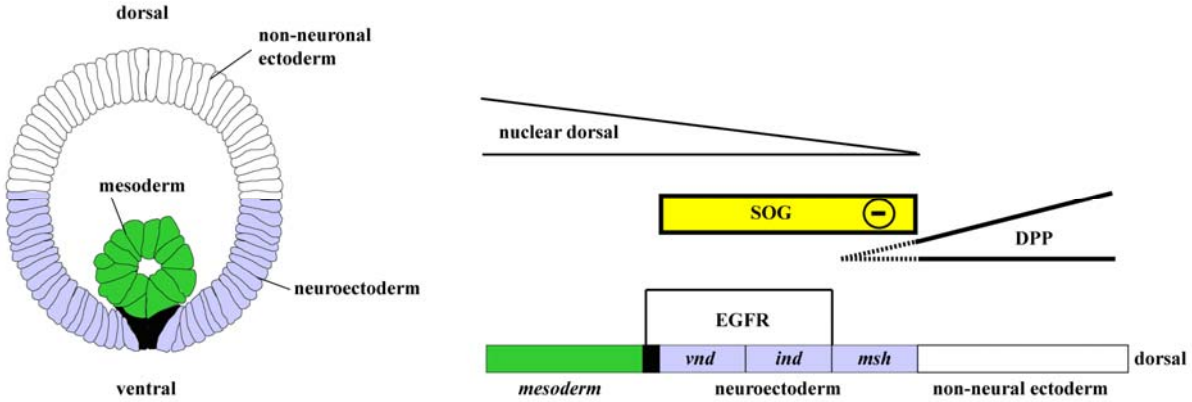



**Fig. 2.** Neuroectoderm formation in the early embryo**.** Left: Cross-section through a *Drosophila* embryo at gastrulation stages. The mesoderm invaginates into the inside of the embryo and the neuroectoderm forms at the ventrolateral side. The dorsal side gives rise to non-neural ectoderm and extra-embryonic tissue (amnioserosa). Right: In blastdoderm stages Dorsal (DL) proteins forms a nuclear gradient with highest concentration at the ventral body side. *short gastrulation* (*sog*) is expressed in ventrolateral cells comprising the neuroectoderm, and decapentaplecig (*dpp*) is expressed in dorsal cells comprising the non-neuronal ectoderm. SOG antagonizes DPP signaling in the neuroectoderm. The genes *ventral nervous system defective* (*vnd*), *intermediate neuroblast defective* (*ind*), and *muscle segment homeobox* (*msh*) pattern the neuroectoderm in three columnar domains (modified according to Cornell and Von Ohlen, 2000).

*sog* is expressed in two broad lateral stripes and is activated by a distinct concentration level within the nuclear DL gradient coinciding with the borders of the presumptive neuroectoderm on the ventral side of the embryo. The *sog* gene product is a secreted protein and forms a morphogenetic gradient, thereby acting as an antagonist for DPP. Thus SOG prevents the neuroectoderm from becoming dorsal epidermis (Ferguson, 1996). Loss of *sog* function leads to a reduction of the neuroectoderm and expansion of the dorsal epidermis.

The processes that pattern the neuroectodermal tissue along the dorsoventral axis are less well characterized than those that specify it. The early ventral neuroectoderm consists of three dorsoventral columns (Fig. 2). The neural precursors within these regions can be characterized by expression of different homeobox transcription factors. These are the products of the *ventral nervous system defective* (*vnd*), *intermediate neuroblast defective* (*ind*), and *muscle segment homeobox* (*msh*) genes. *vnd* expression can be found ventrally, *ind* is expressed intermediately, whereas *msh* expression is defining the dorsal-most column. In *Egfr* mutants dorsal genes are expressed in the intermediate neuroectoderm and intermediate neural precursors fail to form, suggesting a role of *Egfr* in controlling the limits of the intermediate column (Udolph et al., 1998; von Ohlen and Doe, 2000).
Furthermore, there exists a hierarchy of transcriptional repression where the more ventral columnar genes repress the more dorsal genes.
Loss-of-function of columnar genes leads to a strong reduction of delaminating neuroblasts in the mutant region, suggesting an important role of the columnar genes in formation and specification of neural progenitor cells (Cornell and Ohlen, 2000; von Ohlen and Doe, 2000).

### 2.2.1. The *Drosophila* neuroblasts

Neurogenesis is initiated via a set of progenitor cells called neuroblasts. These cells are specified via the proneural genes that encode for basic helix-loop-helix (bHLH) transcription factors. Known proneural genes are the genes of the *achaete-scute complex* (*asc*) which includes *acheate* (*ac*), *scute* (*sc*), *lethal of scute* (*l'sc*) and *asense* (*ase*) (Campuzano and Modolell, 1992). More recent data added the gene *atonal* (*ato*) to the proneural family (Bertrand et al., 2002). After a cell-cell interaction process called lateral inhibition, single cells are selected to adopt a neuroblast fate. Lateral inhibition works with the proneural genes inhibiting themselves in neighboring cells to thereby effectively hinder those cells in taking on the neuroblast cell fate. The signaling pathway acting in this cell communication process is the *Delta*/*Notch* system (Skeath and Thor, 2003).

In *Drosophila* there are about 100 individual neuroblasts in the procephalic neurogenic region that divide in a stereotype spatio-temporal fashion. The process in which the neuroblast leaves the epithelium and begins to divide along its apical-basal axis is called delamination (Fig. 3a). Each neuroblast is characterized by a unique cascade of genes expressed during delamination and production of daughter cells (Urbach and Technau, 2003).
The neuroblasts divide asymmetrically to produce a ganglion mother cell (GMC) and a neuroblast which can go on to divide. The GMCs in turn divide symmetrically to produce two neuronal precursor cells that ultimately differentiate into neurons (Fig. 3) (Jan and Jan, 1998). The gene expression profiles of neuroblasts and GMCs are different. Examples for neural precursor genes expressed in neuroblasts and repressed in GMCs are *ase* and *deadpan* (*dpn*) (Bier et al., 1992; Brand et al., 1993). *even-skipped* (*eve*) and *fushi tarazu* (*ftz*), on the other hand, are expressed in GMCs, where they may help confer GMC identity (Doe et al., 1988a; Doe et al., 1988b).
On the basal side of a mitotic neuroblast Miranda (*mira*) tethers Prospero (*pros*) to the cortex (Matsuzaki et al., 1998; Shen et al., 1998). On the apical side another group of proteins is responsible for orienting the mitotic spindle in a way that ensures segregation of Prospero and Miranda exclusively into the daughter cell – the GMC. These proteins include Bazooka (*baz*), DmPar-6, DaPKC, inscuteable (*insc*) and Rapsynoid (*raps*). Inscuteable is the only component of these complexes that cannot be found in the neuroepitheleal cells before neuroblast specification (Chia and Yang, 2002; Doe et al., 1998; Jan and Jan, 2000).
Prospero, which is translocated into the nucleus of the new-born GMC, represses transcription of cell-cycle genes, effectively limiting the mitotic potential of the cell to one division.
Numb protein is also preferentially segregated into the basal GMC and has been shown to be crucial for cell fate determination of sibling neurons in certain GMC divisions in the CNS (Buescher et al., 1998; Skeath and Doe, 1998).

Although the neuroblasts of the larval brain are of embryonic origin, their properties in terms of division modes and transcriptional characteristics have been shown to be different. In early larvae, neuroblasts divide symmetrically to add to the pool of stem cells. In the later larval stages the mode is asymmetric again producing the usual GMC that again divides quickly to

produce neuronal offspring. The switch in mode seems to be associated with the localization of Inscuteable, which is only apically localized in the case of asymmetric divisions. Another difference to the embryonic stages is the missing colocalisation of Prospero with Miranda in the cortical crescent of mitotic neuroblasts (Ceron et al., 2001).

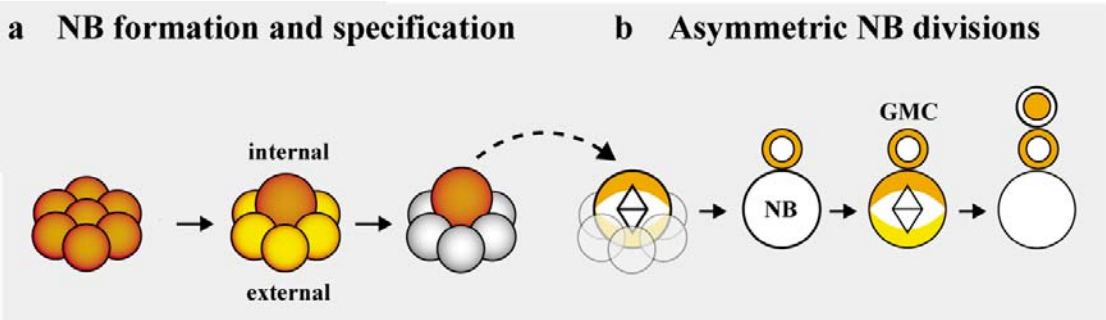



**Fig. 3.** Specification and divison of a neuroblast. a: neuroblast delaminates from the neuroectoderm. b: It has produced two ganglion mother cells, which divide once giving rise to neurons or/and glia cells (modified according to Skeath 2003).

Temporal control over cell fate of GMCs is based on the sequential expression of five genes in neuroblasts, GMCs and neurons. These genes are *hunchback* (*hb*), *Kruppel* (*Kr*), *nubbin (nub, formerly pdm)*, *castor* (*cas*) and *grainy head* (*grh*). The gene expression of one gene always represses the expression of that after the next in the cascade. This mechanism adds another layer of complexity to the abovementioned spatial coordinate system and opens up the possibility for more precision in genetically defining cell fate (Fig. 4, (Isshiki et al., 2001; Skeath and Thor, 2003) reviewed in (Skeath and Thor, 2003)).
The postmitotic neurons created by the GMC are characterised by differentiation marker gene expression, such as *embryonic lethal, abnormal vision* (*elav*).

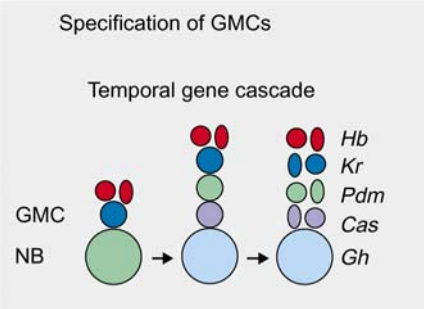



**Fig. 4.** The temporal gene cascade. Different genes are expressed at different times during specification of GMCs (modified according to Skeath 2003).

### 2.2.2. Gliogenesis in *Drosophila*

Another important cell type of neural origin in the brain is the glia cell. Glia have been found to carry out various crucial functions throughout development and in the adult. These actions include secretion of neurotrophic factors (Xiong and Montell, 1995) and control of

proliferative rates of neighboring neurons (Ebens et al., 1993). Other functions comprise guidance of axonal growth via interaction with the neurons (Klaes et al., 1994) and, in vertebrates, the electrical insulation of axons.
In the *Drosophila* central nervous system (CNS) glia are derived from the mesoderm and the procephalic and ventral neuroectoderm. The mesodermal glia are enveloping the commissural axon bundles, the neuroectodermal glia are generated from neuroglial precursor cells and comprise a variety of functions.

Control of neuronal versus glial fate in the *Drosophila* neuroectoderm is dependent of a transcription factor encoded by the *glial cells missing (gcm)* gene. The expression of *gcm* in the CNS appears first transiently at embryonic stage 10. It can then be found in two small groups of neuroectodermal cells per hemisegment. The fist glial markers can be found in a single gcm-expressing precursor delaminating from each of these two groups. This process goes on through embryogenesis leading to increasing numbers of gcm-positive progenitors. Transient expression of *gcm* seems to be a trait of all gliogenic cells emanating from the neuroectoderm throughout the CNS and the peripheral nervous system (PNS). At the end of embryogenesis, no GCM protein is detectable anymore (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996). The mesodermally derived midline glia follow a different transcriptional program that is independent of *gcm* action (Granderath and Klambt, 1999).

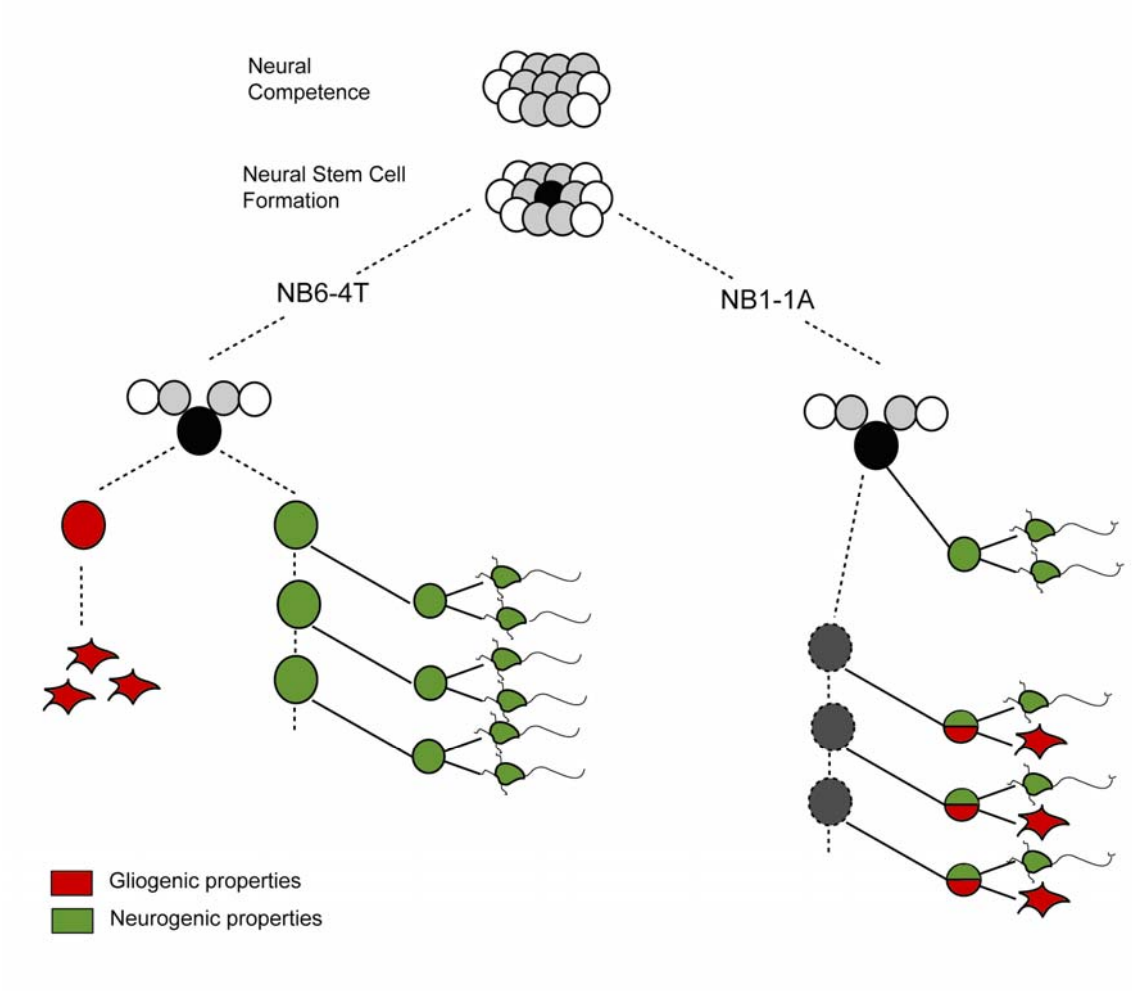



**Fig. 5.** Two division patterns in neuroglial development. Neuroglioblasts (NGB) have singled out through *Notch*/*Delta* mediated lateral inhibition. NGBs delaminate from proneural neuroectoderm clusters (grey) and generate neuronal and glial cells in a mixed lineage. Two different types of NGBs exist. Left: A first type of NGBs bifurcates into a glioblast (GB) and a neuroblast (NB) during their first division and hence creates precursors with restricted developmental potential that give rise either to neurons or to glial cells (e.g. NB6-4T). Right: In contrast, a second type of NGBs creates intermediate precursors that have the potential to generate neurons as well as glia via asymmetric cell division (e.g. NB1-1A). Notch is used for specifying the glial part of the lineage (modified according to Udolph et al., 2001 and Egger 2003).

In summary, this makes *gcm* a key regulator of glial specification acting as a binary switch between neuronal and glial fate in the neuroectoderm.

In *Drosophila* and in mammals there exist neural stem cell lineages generating both neurons and glial cells. In the embryonic neuroectoderm of *Drosophila* about 30 neural progenitor cells per hemisegment delaminate in five discrete waves (Doe, 1992), among which are seven neuroglioblasts (NGB). These give rise to neurons and glial cells, whereas neuroblasts (NB) and glioblasts (GB) exclusively generate neurons or glial cells, respectively (Bossing et al., 1996; Schmid et al., 1999; Schmidt et al., 1997). Glial cells can be produced by NGBs in two distinct ways. Either the NGB divides asymmetrically, producing a neuroblast and a glioblast, or a GMC divides into a neuron and a glia. Both ways are *gcm*-dependent.
The decision to produce neurogenic or gliogenic cells can happen at various time points (Fig. 5). It is possible that an NGB first produces a row of cells with neuronal fate and then switches to a glial mode (Bernardoni et al., 1999).

### 2.2.3. Patterning the brain via the *otd* and the *homeotic* genes

On a larger scale the embryonic brain of *Drosophila* is patterned via an evolutionary conserved set of genes including the cephalic gap genes *orthodenticle* (*otd*) and *empty spiracles* (*ems*) and the homeotic genes *labial* (*lab*), *Deformed* (*Dfd*) and *proboscipedia* (*pb*) (Fig. 6).
Loss-of-function experiments have shown that *otd* plays a critical role in patterning the most anterior part of the brain (Hirth et al., 1995). In homozygous *otd* mutants the protocerebral anlage is deleted because the specification of cells in the neuroectoderm is disrupted, leading to a loss of neuroblasts in this region. The role of *otd* brain patterning is evolutionary conserved. In the mouse there are two homologous genes called *Otx1* and *Otx2*. Both genes are expressed in the embryonic head and brain. Early in embryonic mouse development *Otx2* is expressed in the forebrain and midbrain regions right down to the midbrain-hindbrain boundary (reviewed in (Reichert, 2002)). Loss-of-function experiments with *Otx2* have shown that the resulting mice are early embryonic lethal. Among the reasons for this lethality is the fact that the anterior neurectoderm is not specified. This structure is in wildtype mice responsible for building the forebrain, midbrain and the rostral hindbrain (reviewed in (Reichert, 2002). Cross-phylum rescue experiments in *Drosophila* and *Mus* have shown the strong evolutionary conservation of the Otx proteins. The human *Otx1* and *Otx2* genes where overexpressed in the *Drosophila otd* mutant (Leuzinger et al., 1998) and the murine *Otx1* gene was replaced with the *Drosophila otd* gene (Acampora et al., 1998). In both cases a reasonably good rescue of mutant defects was observed.

The products of the homeotic genes are setting up the posterior regions of the embryonic brain. In *Drosophila*, the homeotic genes are arranged in one cluster, but map to the separated *Antennapedia* (*ANT-C*) and *Bithorax* (*BX-C*) complexes, which are collectively referred to as the *Homeotic complex* (*HOM-C*). The *ANT-C* confines the genes *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), and *Antennapedia* (*Antp*). The *BX-C* contains the genes *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*) (Akam, 1989). An interesting feature of the homeotic genes is their spatial co-linearity in terms of chromosomal arrangement and expression pattern. More 3' located genes are expressed more anteriorly along the body axis of the embryo, whereas more 5' located genes are expressed more posterior. This colinearity is also present on the functional level, as more posteriorly expressed *Hox* genes are functionally dominant over more anteriorly expressed *Hox* genes - a phenomenon termed "posterior prevalence" (Duboule and Morata, 1994). The homeotic gene with the most defined anterior expression domain in the embryonic brain is *lab*, which is expressed in the posterior tritocerebrum. *lab* expression is followed by non-overlapping domains of *Dfd*, *Scr* and *Antp* expression in the mandibular, maxillary and labial neuromeres, respectively (Fig. 6). The *BX-C* genes are expressed in the more posterior thoracic and abdominal neuromeres (Hirth et al., 1998).
Loss-of-function experiments clearly show the importance of the homeotic genes in controlling development of the central nervous system. Knock-out of *labial* or *Deformed* lead

to axonal patterning defects in the tritocerebrum or in the mandibular neuromere, respectively. These defects are not due to loss of cells in the brain, but happen because of loss of neuronal identity in the defective brain neuromeres. This, in turn, leads to a failure of longitudinal and commissural axonal pathway formation because the postmitotic cells adjacent to the mutant neuromeres do not extend their axons into the area containing the unspecified cells (Hirth et al., 1998).

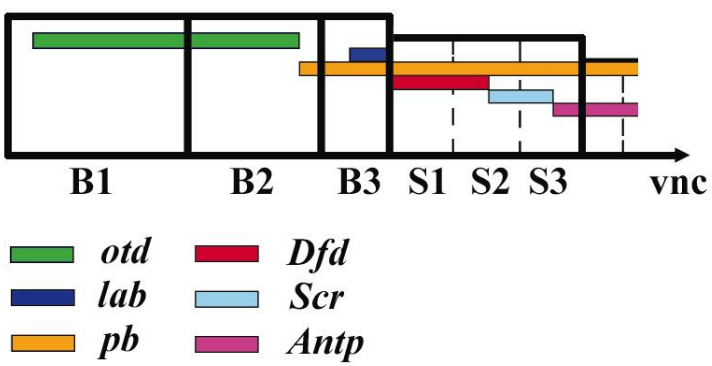



**Fig. 6.** Expression domains of the homeotic genes and of *otd* in the brain and in adjacent neuromeres.
B1: protocerebrum, B2 deutocerebrum, B3: tritocerebrum, S1 mandibular, S2: maxillary, S3: labial neuromere.

It has been shown that a lot of the genetic mechanisms involved in building a brain are evolutionary conserved between *Drosophila* and vertebrates. The striking similarity in structure and function of genes related to neurogenesis could be demonstrated for many basic pathways controlling brain development, such as for the proneural genes (Lee, 1997), the neurogenic genes (Chan and Jan, 1999), the *Hox* genes (Lumsden and Krumlauf, 1996), and the *otd*/*Otx* genes (Acampora and Simeone, 1999).

### 2.2.4. The three phases of neurogenesis

Summarizing brain development in *Drosophila*, there are three phases of neurogenesis. The first phase takes place in the embryo including neuroectodermal proliferation of neuroblasts, patterning of the neuraxis and differentiation into neurons and glia that constitute the brain of the 1st instar larvae. The optic lobe and central brain neuroblasts become quiescent towards the end of embryogenesis. Exceptional in this regard are the neuroblasts building the mushroom bodies, as they continue to proliferate through all developmental stages. During the larval stages a second wave of neurogenesis begins to build the structures for the adult brain. This second phase is based on re-activation of quiescent embryonic neuroblasts. The third phase of neurogenesis happens during metamorphosis in the pupal stages where the proliferation of neuroblasts stops and the neurons terminally differentiate into mature adult cells (Figure 1, (Truman and Bate, 1988).

## 2.3. The *Drosophila* genome and high density oligonucleotide arrays

The high degree of functional conservation between flies and mammals has made *Drosophila* an invaluable tool to facilitate entry into functional genetic studies of mammalian models.

However, in the study of genetic pathways and epigenetic interrelations the lack of information on the full genomic scale has always been a major obstacle in trying to understand how complex organs like the brain are generated during development.
In the year 2000 the genome sequence of *Drosophila melanogaster* has been published (Adams et al., 2000). The current annotated release (as per September 2004 – Release 3) of the genome contains about 14000 genes. The fact that the genome is relatively compact and contains few duplicate genes adds to the attractiveness of the model.
With the sequence available, full-genome experiments have become possible, such as massively parallel analysis of gene expression. In the past decade microarrays (Schena et al., 1995) and oligonucleotide arrays (Lipshutz et al., 1999) have become major tools for comprehensive analysis of gene expression. This new quality in functional study on the nucleic acid level opens new possibilities for looking into the genetic pathways that underlie development in complex organisms.
Microarrays are manufactured by spotting thousands of cDNA clones taken from the 3´ region of the transcript onto a coated glass surface using a high-speed robot. During the main course of a microarray experiment, total RNA from both the test and reference sample is fluorescently labeled with two different fluorescent markers. The fluorescent targets are pooled and hybridized under stringent conditions to the clones on the microarray. The intensity of the two fluorescent signals is measured using a confocal scanning laser microscope and the resulting images are pseudo-colored, merged and the measured spots are annotated with information including gene name, intensity values, intensity ratio, normalization constant and confidence interval. Increase or decrease of transcript levels between experimental conditions can be can be viewed as a normalized ratio of the two fluorescent intensities (Duggan et al., 1999).

The other main technology used to conduct gene expression studies on the genomic scale are high density oligonucleotide arrays (HDOAs) also called GeneChips. This is also the method used in the studies described throughout this thesis. These arrays are manufactured by using light directed synthesis of nucleotides on a glass surface (Fig. 7). By this means it is possible to fit hundreds of thousands of different oligonucleotides, named probes, onto a small surface.

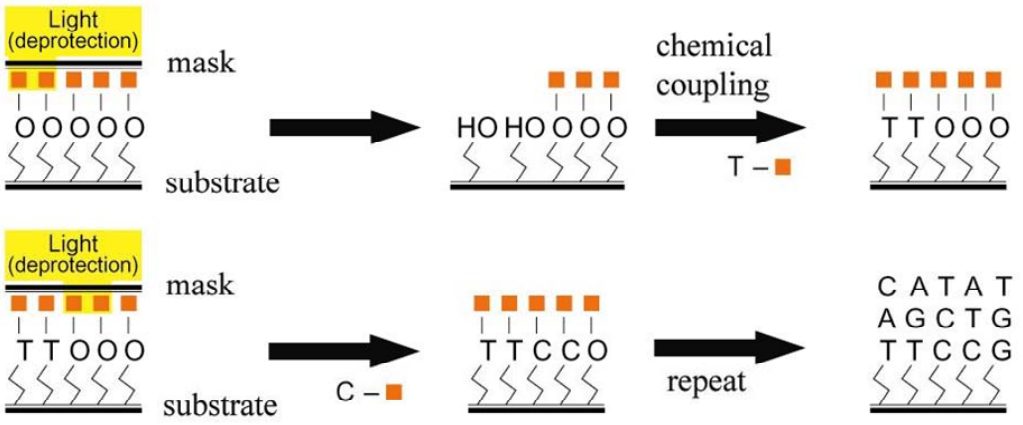



**Fig. 7.** Light directed oligonucleotide synthesis. A solid support is derivatized with a covalent linker molecule terminated with a photolabile protecting group. Light is directed through a mask to deprotect and activate selected sites, and protected nucleotides couple to the activated sites. The process is repeated, activating different sets of sites and coupling different bases allowing arbitrary DNA probes to be constructed at each site. (modified from Lipshutz, 1999 #27).

A gene is not represented by a single spot of a cDNA species but by up to 20 spots of 25-mers that are taken from different regions of the gene of interest. These spots of oligonucleotides are synthesized on the surface in pairs, with one of the two spots containing the perfect match sequence and the other spot having a single-base mismatch in the middle of the sequence (Lipshutz et al., 1999). The full complement of probes representing a single gene is called a probeset (Fig. 8). The HDOAs used in some of the studies reported here contain probesets representing all genes in the genome of *Drosophila melanogaster*. In case of the larger vertebrate genomes the genes are split into several arrays.

In an experiment using HDOAs, biotinylated cRNA prepared from total RNA and extracted from the tissue of interest is hybridized to the array which is subsequently scanned with a confocal laser scanner. The resulting image is annotated and analyzed using a variety of algorithms. The relatively complex setup of molecules on the array containing surface adds the possibility of controlling for unspecific cross-hybridizations. The actual relative expression level of the genes represented on the array is derived statistically from the measured intensities of the probes. The result takes into account the signals for the mismatch probes leading to a high confidence measure of the abundance of hybridized labeled target cRNA species (Lipshutz et al., 1999).
Microarrays and GeneChips have become major tools in functional genomics as they make it feasible to take genome-wide snapshots of gene expression in one single experiment. The range of array-based studies conducted is quickly getting wider.
In *Drosophila* research microarray technology has been used to study a wide array of topics. One approach in designing an array experiment is to sample the transcriptional profiles of wildtype animals or tissues, looking for spatially or temporally differential expression of genes during development. This has been done in studies that established a transcriptional time course of *Drosophila* embryonic development (Tomancak et al., 2002) or even profiled

through all stages including adulthood (Arbeitman et al., 2002). Other studies within this paradigm fished for spatially restricted gene expression during dorso-ventral patterning (Stathopoulos et al., 2002), in wing development (Butler et al., 2003), mesoderm development (Furlong et al., 2001), and metamorphosis (White et al., 1999). Further surveys looked at the establishment of olfactory learning (Dubnau et al., 2003) and at transcriptional differences between the sexes (Parisi et al., 2004).
Another conceptual way to conduct full-genome expression profiling is to set up a comparison of different conditions in terms of external influences. This has been done with studies

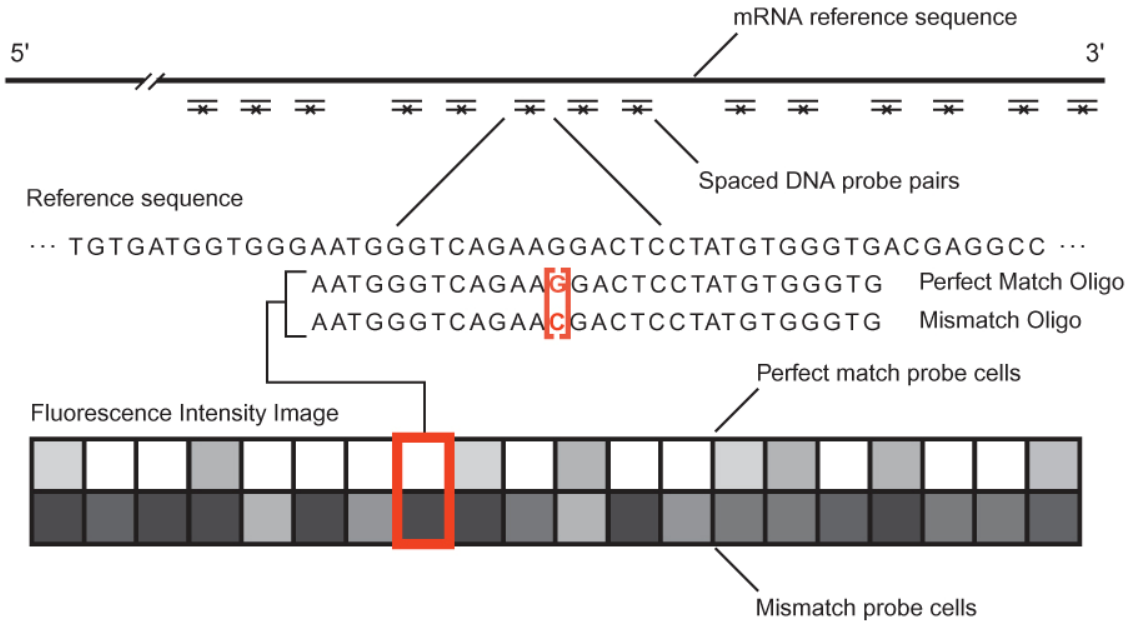



**Fig 8.** Expression probe and array design. Oligonucleotide probes are chosen based on uniqueness criteria and composition design rules. For eukaryotic organisms, probes are chosen typically from the 3´ end of the gene or transcript (nearer to the poly(A) tail) to reduce problems that may arise from the use of partially degraded mRNA. The use of the PM minus MM differences averaged across a set of probes greatly reduces the contribution of background and cross-hybridization and increases the quantitative accuracy and reproducibility of the measurements (modified after {Lipshutz, 1999 #27}).

examining the immune response of flies (Irving et al., 2001) or the resistance to drugs or toxins (Le Goff et al., 2003; Pedra et al., 2004).
Functional genomic studies can also be carried out by using loss- and gain-of function genotypes, often in search for target genes of the protein of interest. This approach has been exploited by labs looking into trithorax genes (Beltran et al., 2003), Ras overexpression in the hematopoietic system (Asha et al., 2003) and into genes regulating neuronal morphogenesis (Brenman et al., 2001).

## 2.4. Drosophila as a model in cancer research

Cancer research has produced a large body of data during the past 25 years. The discovery of oncogenes, producing dominant gain-of-function phenotypes and tumor suppressors displaying recessive loss-of-function phenotypes was only the beginning leading the field into the extremely complex biology of this disease.

However, yet no small set of principles governing tumorigenesis have emerged. The transformation of a normal cell into a cancer cell seems to involve a complex series of genomic alterations. Combined with the more than 100 known types of cancer this body of evidence makes the identification of the critical genetic pathways a difficult task.
From a functional point of view, most vertebrate cancer cells can be characterized by a set of features including self-sufficiency in growth signals, insensitivity to growth inhibition, ignorance of apoptotic signals, sustained angiogenesis, unrestrained replicative potential, and the ability to invade other tissues and produce metastases. The events that lead to these cellular capabilities can be very disparate, with single events sometimes enabling multiple traits (reviewed in (Hanahan and Weinberg, 2000)).

A lot of the abovementioned prerequisites can be directly linked to basic mechanisms of brain development. The growing group of genes involved in basic developmental and cell biological pathways that have links to cancer is growing quickly. Prime candidates for this group are genes controlling cell growth, differentiation and death (Wechsler-Reya and Scott, 2001). Also, there is growing evidence that the functional conservation between *Drosophila* and mammals also extends to tumor suppressors. The strong evolutionary conservation of gene pathways and the striking similarity in cellular processes between *Drosophila* and mammals show that research on fly tumorigenesis can add substantially to the understanding of human brain tumors (reviewed in (Potter et al., 2000)). Also, there is growing evidence that functional conservation between *Drosophila* and mammals also extends to tumor suppressors. Examples for this are genes like *brain tumor* (*brat*), *malignant brain tumor* (*mbt*), *lethal giant larvae* (*lgl*) and *disc large* (*dlg*). Mutations in those genes lead to tumor phenotypes in the larval brains consisting of malignant and invasive neoplasms (reviewed in (Arama et al., 2000)).
It can be very productive to use *Drosophila* as a sidestep in research on mammalian cancer to take advantage of the easier genetic accessibility, the smaller genome and the wealth of tools that exist for this model system. Knowledge gained in looking at fly genetic pathways can often be directly used in vertebrate systems. This has been done with the *patched*/*hedegehog* pathway. The human *patched* acts as a tumor suppressor and is mutated in the nevoid basal cell carcinoma syndrome. Investigation of the *Drosophila patched*/*hedgehog* pathway led to the discovery of several other genes acting in vertebrate tumor formation (Hahn et al., 1996; Xie et al., 1998).

### 2.4.1. The *brain tumor* gene acts as a tumor suppressor in *Drosophila*

The *brat* gene is a prime example for a *Drosophila* tumor suppressor gene. The *brat* locus spans about 38 000 base pairs on chromosome 2L at position 37C6. There are at least five exons and four introns with a coding region that is confined to exon 5 only. cDNA analysis suggested two transcripts. The derived amino acid sequence of Brat indicates a protein of 1037 amino acids. Analysis of the protein sequence revealed two B-box zinc-finger motifs, a coiled-coil domain, and a C-terminal beta-propeller domain with the amino acid sequence of

the Brat C-terminal repeats being highly similar to the NHL consensus sequence. All these motifs are reported to be involved in protein - protein interactions (Arama et al., 2000).

Brat belongs to the B-box family of proteins which contains a large number of members involved in functions like axial patterning, growth control, differentiation, transcriptional regulation and cancer formation. The B-box family genes have been implied to have roles in very fundamental biological processes including human tumor formation (Jensen et al., 2001; Torok and Etkin, 2001). Homologous proteins can be found in Caenorhabditis elegans, Mus Musculus, Rattus norvegicus and Homo sapiens (Arama et al., 2000). The highest similarity can be found in the C. elegans protein NCL-1, which has been shown to be a functional homolog of *brat*. This is supported by the fact that Drosophila brat can functionally replace ncl-1 in a cross-phylum rescue experiments. NCL-1 is an inhibitor of cell growth and negatively regulates RNA polymerase I and III transcription in nematodes (Frank and Roth, 1998). A mammalian homolog of brat with some known properties is the Rattus protein BERP, which is interacting with alpha-actinin-4 (El-Husseini et al., 2000). The other mammalian homologs of Brat are of mostly unknown function and include KIAA0517 in Homo and the Mus protein HAC1 (Arama et al., 2000).

In the embryo, brat expression can be found in the developing brain, in the ventral nerve cord and in the peripheral nervous system from embryonic stage 11 onwards. Apart from expression in ganglion mother cells (GMCs) and neuroblasts in the central nervous system, transcript can also be detected in the differentiating sensilla and in the chordotonal organs (Arama et al., 2000). Larval expression can be found uniformly in the entire brain hemispheres including the optic lobes of third instar larvae. Weaker but also quite uniformly distributed levels of transcript are detected in the imaginal discs of third instar larvae. In eye discs brat is expressed in small cell clusters along the morphogenetic furrow (Frank et al., 2002). Among findings about Brat function is the translational repression of hunchback (hb) mRNA in the abdomen of the early Drosophila embryo. This repression requires the recruitment of Brat by Nanos and Pumilio to form a quaternary complex with the 3' untranslated region of hb mRNA (Sonoda and Wharton, 2001). It has been suggested that the regulation of hb translation can take place by removal of the poly(A) tail of hb mRNA and by a poly(A)-independent process that directly affects translation (Chagnovich and Lehmann, 2001). Furthermore, misexpression experiments have shown that Brat can block cell proliferation in a variety of tissues and organs, and that it can inhibit cell division in culture cells (Sarnatzki et al., 2003). Also, there is evidence that brat mutant cells have larger nucleoli and excess rRNA whereas an overexpression of brat negatively regulates the level of cellular rRNA. brat overexpression in eye- and wingdiscs leads to a decrease in organ size, inhibiting cell growth and slowing down cell division (Frank et al., 2002). Taken together, the functional data support a role for brat in cell proliferation, translational repression and RNA metabolism. There are several brat alleles available in Drosophila and some of them have been shown to lead to a severe overproliferation of the larval brain. Most of the alleles that display these larval neoplasms have mutations in the beta-propeller domain (Arama et al., 2000). Neither the critical time for tumorigenesis in development nor the cellular nature of the tumor is

known to date. Some evidence suggests that the tumor arises from a population of cells in the third instar larval brain that is only as big as a few hundred, representing only a small percentage of the complete cellular pool. When transplanted into wildtype hosts, primary tumors from mutant larval brains form secondary tumors and invade tissues that are remote from the place of implantation (Woodhouse et al., 1998).

## 2.6. Scope of this thesis

This thesis draws a bow from functional genetics of *Drosophila* brain development to cancer research in a full-genome transcriptional analysis of a *Drosophila* brain tumor.

As a pilot study in functional genomics of *Drosophila* custom-made high density oligonucleotide arrays were used to capture the transcriptional profile of 1500 genes under standard conditions and in response to heatshock. The analysis of differential gene expression following heat shock application 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* (Leemans et al., 2000).

Analyzing the genetics of brain patterning in the embryo of *Drosophila* we studied the functional equivalence of Hox gene products in specification of the tritocerebrum. During embryonic development of the Drosophila brain, the Hox gene labial is required for the regionalized specification of the tritocerebral neuromere. In the absence of labial, the cells in this brain region do not acquire a neuronal identity leading to major axonal pathfinding deficits. Using genetic methods, the *labial* gene was replaced in the tritocerebrum with all other known *Drosophila* Hox genes. We could show, that with the exception of *Abdominal-B* all Hox genes are able to functionally replace *labial* in its role of specifying tritocerebral identity. The rescue efficiency correlated with the arrangement of the Hox genes on the chromosome (Hirth et al., 2001).

Homeotic protein action is based on the regulation of specific downstream genes in a variety of developmental processes. However, only a small number of target genes of homeodomain transcription factors have been found. By using transgenic flies carrying the *lab* gene under the control of a heat-inducible promoter, *lab* was ubiquitously overexpressed following heat-shock treatment in *Drosophila* embryos. Expression profiles suing HDOAs identified differences in transcript levels in response to *lab* overexpression. The results demonstrate that overexpression of *lab* leads to activation and repression of genes encoding proteins that cover a wide variety of molecular functions. The analysis identified a novel set of candidate LAB target genes and thus, provides a set of data for further functional analysis of hometic gene action (Leemans et al., 2001).

Furthermore, full-genome transcriptional profiling experiment was conducted to analyse the functional equivalence of *Drosophila otd* and vertebrate *Otx2* gene action. This was accomplished by comparing *Drosophila* embryos overexpressing the fly *otd* gene with embryos overexpressing the human *Otx2* gene. The experiments showed that approximately one third of the otd-regulated transcripts also respond to overexpression of 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).

Also, oligonucleotide arrays were used to compare the complete transcriptome of wildtypoe *Drosophila* embryos with embryos overexpressing the *glial cells missing* (*gcm*) gene throughout the complete neuroectoderm. The experiment was conducted for two different timepoints in embryonic development. Results included hundreds of genes that were differentially expressed following *gcm* misexpression. These genes are potentially involved in aspects of glial development. *In vivo* validation of the array data was performed for a subset of the data.
Being one of the first full-genome analyses of gene expression events downstream of a key developmental transcription factor this study puts forth reveals new insights into the genetic pathways involved in initiation and maintenance of cell fate determination in the CNS (Egger et al., 2002).

The main focus of this thesis was the transcriptional analysis of an adult *Drosophila* brain tumor using full-genome high density oligonucleotide arrays. The tumor was caused by a mutation in the *brain tumor* (*brat*) gene. Two independent genome-wide gene expression studies using two different oligonucleotide microarray platforms were used to compare the transcriptome of adult wildtype flies with mutants displaying the adult bratk06028 mutant brain tumor. Cross-validation and stringent statistical criteria identified a core transcriptional signature of bratk06028 neoplastic tissue. We find significant expression level changes for 321 annotated genes associated with the adult neoplastic bratk06028 tissue indicating elevated and aberrant metabolic and cell cycle activity, upregulation of the basal transcriptional machinery, as well as elevated and aberrant activity of ribosome synthesis and translation control. One fifth of these genes show homology to known mammalian genes involved in cancer formation. Our results identify for the first time the genome-wide transcriptional alterations associated with an adult brain tumor in Drosophila and reveal insights into the possible mechanisms of tumor formation caused by homozygous mutation of the translational repressor brat (Loop et al., 2004).

## 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

Correspondence should be addressed to Dr. Ronny.Leemans, Institute of Zoology, University of Basel, Rheinsprung 9, CH-4051 Basel, Switzerland. Tel. (41-61) 2673491; Fax (41-61) 2673457; e-mail: Ronny.Leemans@unibas.ch

## Abstract

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.

Recently the genome of the first multicellular eukaryote *C. elegans* has been completely elucidated (1). Sequencing of the *Drosophila melanogaster* genome has also been carried out and currently the corresponding putative open reading frames are being defined (2). 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 arrays or oligonucleotide arrays (3, 4), both of which allow the simultaneous monitoring of hundreds of mRNA expression profiles (5, 6). 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 (7-9). We analyzed the relative abundance levels of hundreds of embryonically expressed genes under normal physiological conditions and in response to heat shock (10). 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 (9). 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 (11). 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 (12, 13).

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'-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 (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 analyzed 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 (4). *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 (13).

**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 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 wildtype (condition 1) as well as heat shock treated wildtype (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 (14). 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 1519 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 wildtype embryos raised under standard conditions (25˚C). Transcript imaging revealed a total of 563 (37%) of the 1519 *Drosophila* genes as expressed in embryonic stages 10-17. *(Expression levels and statistical data on all expressed transcripts will be made available according to*

*editorial policy)*. 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>5000, 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 (1315); 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) (10, 15, 16). 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* (17, 18). 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 (19, 20). 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 (20). The heat shock cognate genes (Hsc) have been reported to be expressed at normal temperatures but are not further induced by heat shock (21, 22). 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 (23) 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* (24), 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 (25). 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 (26).

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 (27, 28). 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* (29). With the imminence of whole genome sequence data for *Drosophila* (2) 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.

## 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 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.

## References


1. The *C. elegans* Consortium. (1998) *Science* **282,** 2012-2018.
2. Burtis, K. C. & Hawley, R. S. (1999) *Nature* **401,** 125, 127.
3. Granjeaud, S., Bertucci, F. & Jordan, B. R. (1999) *Bioessays* **21,** 781-790.
4. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. (1999) *Nat. Genet.* **21,** 20-24.
5. Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., et al. (1996) *Nat. Biotechnol.* **14,** 1675-1680.
6. Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O. & Davis, R. W. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 13057-13062.
7. Akam, M. (1987) *Development* **101,** 1-22.
8. Pankratz, M. J. & Jäckle, H. (1993) in *The Development of Drosophila melanogaster*, eds. Bate, M. & Martinez-Arias, A. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), Vol.1, pp. 467-516.
9. Campos-Ortega, J. & Hartenstein, V. (1997) *The embryonic development of Drosophila melanogaster* (Springer, Heidelberg).
10. Nover, L. & Scharf, K. D. (1997) *Cell. Mol. Life Sci.* **53,** 80-103.
11. Walter, M. F., Petersen, N. S. & Biessmann, H. (1990) *Dev. Genet.* **11,** 270-279.
12. Mahadevappa, M. & Warrington, J. A. (1999) *Nat. Biotechnol.* **17,** 1134-1136.
13. Certa, U., de Saizieu, A. & Mous, J. (2000) *Methods Mol. Biol.*, in press.
14. Tautz, D. & Pfeifle, C. (1989) *Chromosoma* **98,** 81-85.
15. Lindquist, S. & Craig, E. A. (1988) *Annu. Rev. Genet.* **22,** 631-677.
16. Schlesinger, M. J. (1990) *J. Biol. Chem.* **265,** 12111-12114.
17. Pauli, D. & Tissières, A. (1990) in *Stress Proteins in Biology and Medicine*, eds. Morimoto, R., Tissières, A. & Georgopoulos, C. (Cold Spring Harbour Laboratory Press, New York), pp. 361-378.
18. Michaud, S., Marin, R. & Tanguay, R. M. (1997) *Cell. Mol. Life Sci.* **53,** 104-113.
19. Haass, C., Klein, U. & Kloetzel, P. M. (1990) *J. Cell. Sci.* **96,** 413-418.
20. Vazquez, J., Pauli, D. & Tissières, A. (1993) *Chromosoma* **102,** 233-248.
21. Craig, E. A., Ingolia, T. D. & Manseau, L. J. (1983) *Dev. Biol.* **99,** 418-426.
22. Rubin, D. M., Mehta, A. D., Zhu, J., Shoham, S., Chen, X., Wells, Q. R. & Palter, K. B. (1993) *Gene* **128,** 155-163.
23. Ferrante, A. W., Jr., Reinke, R. & Stanley, E. R. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1911-1915.
24. Cutforth, T. & Rubin, G. M. (1994) *Cell* **77,** 1027-1036.
25. Ishikawa, T., Igarashi, T., Hata, K. & Fujita, T. (1999) *Biochem. Biophys. Res. Commun.* **254,** 566-571.
26. Wang, X. Z., Kuroda, M., Sok, J., Batchvarova, N., Kimmel, R., Chung, P., Zinszner, H. & Ron, D. (1998) *EMBO J.* **17,** 3619-3630.
27. Henning, K. A., Peterson, C., Legerski, R. & Friedberg, E. C. (1994) *Nucleic Acids Res.* **22,** 257-261.

28. Kraemer, C., Weil, B., Christmann, M. & Schmidt, E. R. (1998) *Gene* **216,** 267-276.
29. White, K. P., Rifkin, S. A., Hurban, P. & Hogness, D. S. (1999) *Science* **286,** 2179-2184.

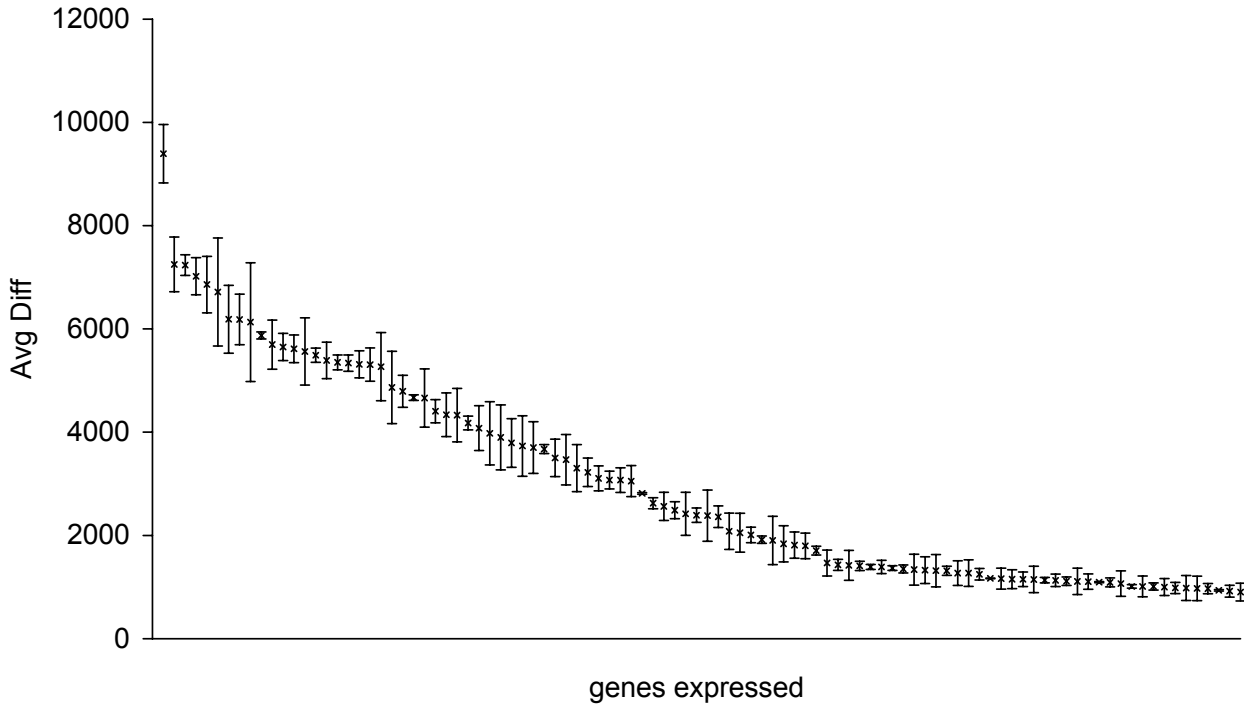



**Fig. 1.** Gene expression monitoring of stage 10-17 wildtype embryos raised under standard conditions (25˚C). Compilation of the 100 genes expressed with the highest Average Difference values (Avg Diff; for details see *Materials and Methods*) and the corresponding standard deviations (SD; indicated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

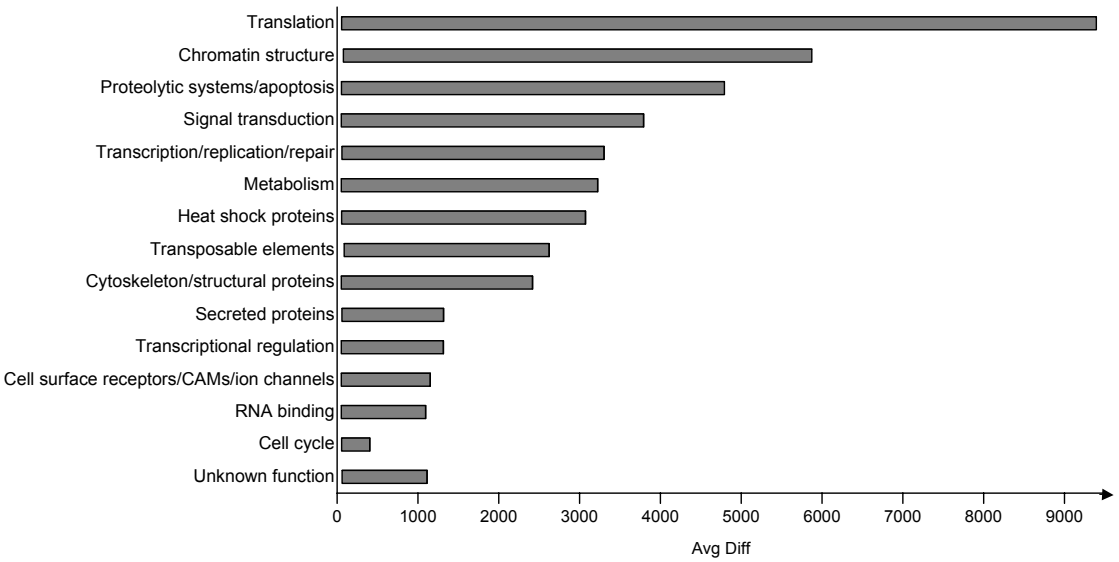



**Fig. 2**. Range of Avg Diff values for expressed genes, grouped according to their functional classes. Translation (min Avg Diff 56 - max Avg Diff 9394), Chromatin structure (78-5873), Proteolytic systems/ apoptosis (53-4792), Signal transduction (52-3791), Tanscription/ replication/ repair (59-3303), 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).

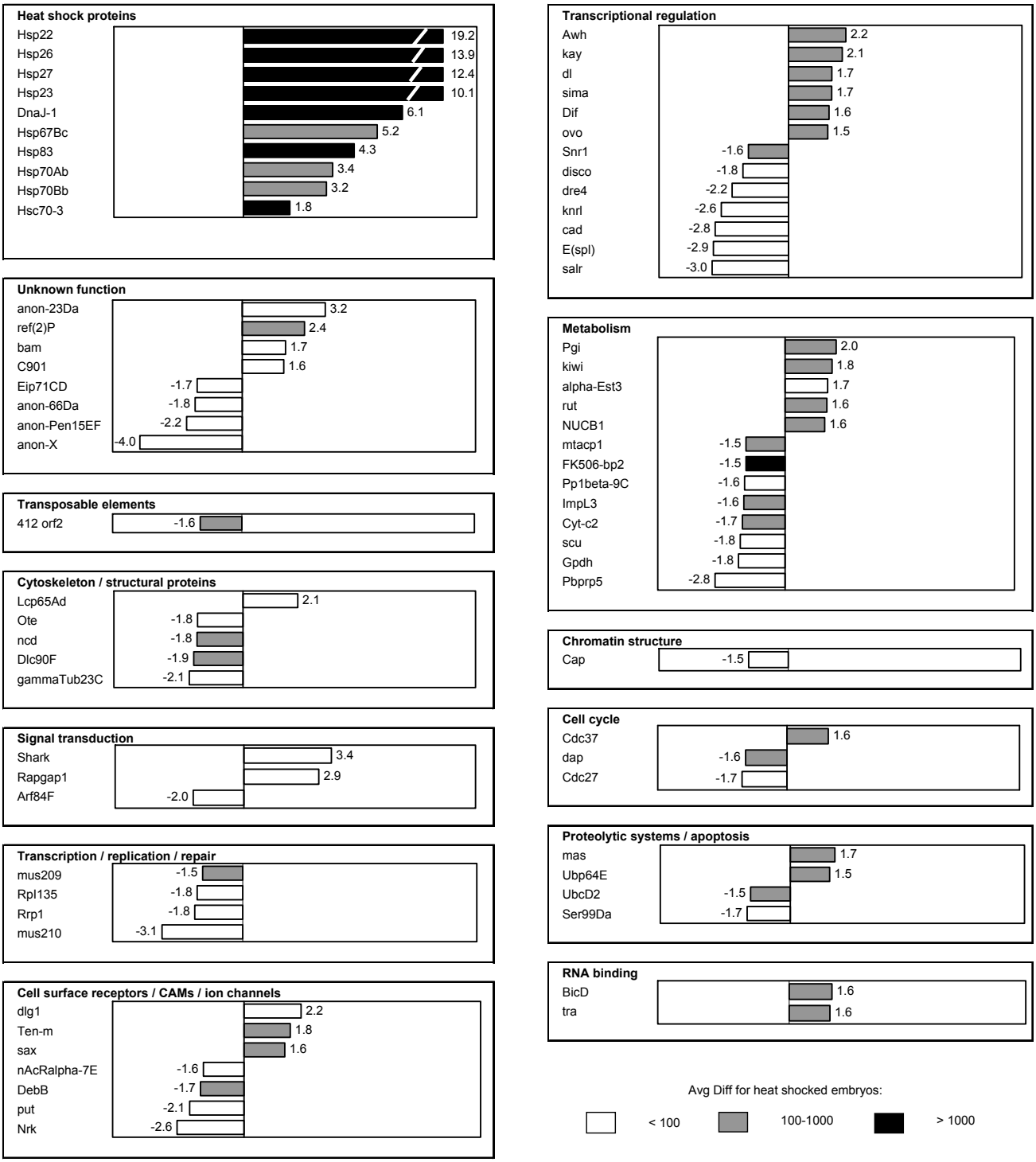



**Fig. 3.** Differentially expressed genes observed in heat shocked versus non-heat shocked stage 10-17 wildtype embryos, grouped according to functional classes. Bars represent the fold-change of differentially expressed genes in the heat shock versus standard condition. Positive values indicate that 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.

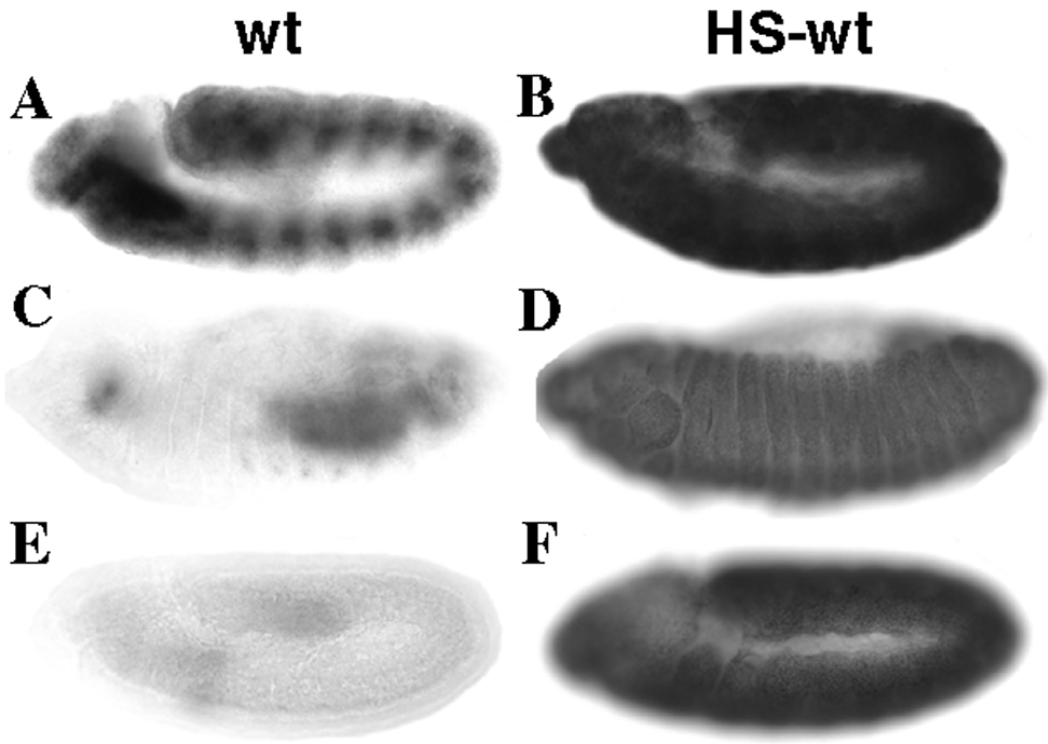



**Fig. 4.** Comparison of whole mount *in situ* hybridizations between non-heat shocked and heat shocked wildtype embryos. (*A-F*) lateral views, anterior to the left. (*A, C, E*) non-heat shocked wildtype, (*B, D, F*) heat shocked wildtype embryos. (*A, B*) at stage 11 *Hsp22* expression is confined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

**Table 1.** *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 |
| | $\Sigma N = 1519$ | $\Sigma n = 563$ | | |

Genes expressed in stage 10-17 wildtype 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 $\Sigma N=1519$. Number of genes expressed within a functional group (n); total number of transcripts detected $\Sigma n=563$. (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.

**Table 2.** 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 |

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

## 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], Ullrich 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

*Author for correspondence and material request: Dr. Frank Hirth, Institute of Zoology, Biocenter/Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland. Tel. (41-61) 2671617; Fax (41-61) 2671613; e-mail: Frank.Hirth@unibas.ch

## Abstract

### Background

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).

### Results

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.

### Conclusions

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.

## Background

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 [1-6]. 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) [7-11].

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 [12,13]. 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 [9,14-21]. 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 [9]. 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 [22,23].

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 [24-27]. Genetic studies have demonstrated that *lab* is required for proper head formation [28], for the specification of cellular identity in the midgut [29] as well as in the embryonic brain [30]. 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 [31-35].

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 [36-39]. 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.

## Results

In this study, transgenic fly strains carrying the *lab* coding sequence under control of the heat inducible Hsp70 promotor were used [40]. 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). Wildtype 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 wildtype 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 [39]). 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 [39].

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 wildtype 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 wildtype 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 [41,42]. Since *lab* activity is only observed from gastrulation onwards [26], this suggests that high levels of widespread ectopic *lab* expression are able to activate genes which under wildtype conditions show non-overlapping spatio-temporal expression domains as compared to that of *lab*. Increases in the 5-10 fold range were seen for 6 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* [43]. 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 [43].

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 [9], 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 [30]. 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 [30]. *18 wheeler* has been identified as a binding site of the homeotic protein UBX in polytene chromosomes [18].

To verify the differences in expression level after heat shock induced overexpression of *lab* as compared to heat shocked wildtype 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 [23,44]. 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 of a given homeoprotein can have different functional consequences in terms of developmental specificity [29,45], 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 concomittant overexpression of cofactors, which are thought to act together with homeotic proteins to determine their in vivo target specificity [34,46]. It is noteworthy in this respect, that the gene *mirror*, which has been proposed to be an additional cofactor for homeoprotein specificity [47], 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 [1,2,9,15]. 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 are 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.

## Conclusions

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* [48-50]. Considering the evolutionary conservation of gene structure, expression and function [1,35], 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.

## Materials and methods

### Fly strains, embryo collections and heat shock regime

The wildtype 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 [40]. 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 [51]). Embryos younger than embryonic stage 10 were not used, since heat shock in these earlier stages results in lethality [52].

### 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 [53]. 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 [54]. The primary antibody was rabbit anti-LAB [55] 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 [36] 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 [39]). 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) [37]. *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 [56].

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 analyzed 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 [39], 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 wildtype embryos compared to non heat shocked wildtype controls, were excluded from further analysis (data not shown; [39]). 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 [39]. 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 [39]. 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).

**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.

## References


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

2. Manak JR, Scott MP: A class act: conservation of homeodomain protein functions. *Development* 1994, Suppl.: 61-71.

3. Krumlauf R: Hox genes in vertebrate development. *Cell* 1995, 78: 191-201.

4. Carroll SB: Homeotic genes and the evolution of arthropods and chordates. *Nature* 1995, 376: 479-485.

5. Maconochie M, Nonchev S, Morrison A, Krumlauf R: Paralogous Hox genes: function and regulation. *Annu Rev Genet* 1996, 30**:** 529-556.

6. Gellon G, McGinnis W: Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays* 1998, 20: 116-125.

7. Lawrence PA, Morata G: Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* 1994, 78: 181-189.

8. Duboule D, Morata G: Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet* 1994, 10: 358-364.

9. Graba Y, Aragnol D, Pradel J: *Drosophila* Hox complex downstream targets and the function of homeotic genes. *BioEssays* 1997, 19: 379-388.

10. Morata G, Sanchez-Herrero E: Patterning mechanisms in the body trunk and the appendages of *Drosophila. Development* 1999, 126: 2823-2828.

11. Mann RS, Morata G: The developmental and molecular biology of genes that subdivide the body of *Drosophila*. *Annu Rev Cell Dev Biol* 2000, 16: 243-271.

12. Garcia-Bellido A: Genetic control of wing disk development in *Drosophila*. *Ciba Foundation Symp* 1975, 29**:** 161-182.

13. Pradel J, White RAH: From selectors to realizators. *Int J Dev Biol* 1998, 42**:** 417-421.

14. Andrew DJ, Scott MP: Downstream of the homeotic genes. *The New Biologist*1992, 4: 5-15.

15. Morata G: Homeotic genes of *Drosophila***.** *Curr Opin Genes Dev* 1993, **3:** 606-614.

16. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K: Homeodomain-DNA recognition. *Cell* 1994, **78:** 211-223.

17. Mastick GS, McKay R, Oligino T, Donovan K, Lopez AJ: Identification of target genes regulated by homeotic proteins in *Drosophila melanogaster* through genetic selection of Ultrabithorax protein binding sites in yeast. *Genetics* 1995, 139: 349-363.

Botas J, Auwers L: Chromosomal binding sites of Ultrabithorax homeotic proteins. *Mech Dev* 1996, 56: 129-138.

19. Mannervik M: Target genes of homeodomain proteins. *BioEssays* 1999, 21**:** 267-270.

20. Nasiadka A, Krause HM: Kinetic analysis of segmentation gene interactions in *Drosophila* embryos. *Development* 1999, 126: 1515-1526.

21. Nasiadka A, Grill A, Krause HM: Mechanisms regulating target gene selection by the homeodomain-containing protein Fushi tarazu. *Development* 2000, 127: 2965-2976.

22. Biggin MD, McGinnis W: Regulation of segmentation and segmental identity by *Drosophila* homeoproteins: the role of DNA binding in functional activity and specificity. *Development* 1997, 124: 4425-4433.
23. Liang Z, Biggin MD: *Eve* and *ftz* regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* 1998, 125: 4471-4482.
24. Mlodzik M, Fjose A, Gehring WJ: Molecular structure and spatial expression of a homeobox gene from the *labial* region of the *Antennapedia*-complex. *EMBO J* 1988, 7: 2569-2578.
25. Diederich RJ, Merrill VKL, Pultz MA, Kaufman TC: Isolation, structure and expression of *labial*, a homeotic gene of the *Antennapedia* complex involved in *Drosophila* head development. *Genes Dev* 1989, 3: 399-414.
26. Kaufman TC, Seeger MA, Olsen G: Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet* 1990, 27: 309-362.
27. Duboule D: *Guidebook to the Homeobox Genes*. Oxford: Oxford University Press, 1994
28. Merrill VKL, Diederich RJ, Turner FR, Kaufman TC: A genetic and developmental analysis of mutations in *labial*, a gene necessary for proper head formation in *Drosophila melanogaster*. *Dev Biol* 1989, 135: 376-391.
29. Hoppler S, Bienz M: Specification of a single cell type by a *Drosophila* homeotic gene. *Cell* 1994, 76: 689-702.
30. Hirth F, Hartmann B, Reichert H: Homeotic gene action in embryonic brain development of *Drosophila*. *Development* 1998, 125: 1579-1589
31. Pöpperl H, Bienz M, Studer M, Chan S-K, Aparicio S, Brenner S, Mann RS, Krumlauf R: Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell* 1995, 81: 1031-1042.
32. Lutz B, Lu H-C, Eichele G, Miller D, Kaufman TC: Rescue of *Drosophila labial* null mutant by chicken ortholog *Hoxb-1* demonstrates that the function of Hox genes is phylogenetically conserved. *Genes Dev* 1996, 10: 176-184.
33. Chan S-K, Pöpperl H, Krumlauf R, Mann RS: An *extradenticle*-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J* 1996, 15: 2476-2487.
34. Mann RS, Chan S-K: Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet* 1996, 12: 258-262.
35. Hirth F, Reichert H: Conserved genetic programs in insect and mammalian brain development. *BioEssays* 1999, 21: 677-684.
36. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996, 14**:** 1675-1680.
37. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: High density synthetic oligonucleotide arrays. *Nat Genet* 1999, 21: 20-24.
38. Rubin GM *et al*: Comparative genomics of the eukaryotes**.** *Science* 2000, 287: 2204-2215.

39. Leemans R, Egger B, Loop T, Kammermeier L, He H, Hartmann B, Certa U, Hirth F, Reichert H: Quantitative transcript imaging in normal and heat shocked *Drosophila* embryos by using high-density oligonucleotide arrays. *Proc Natl Acad Sci USA,* 2000, 97: 12138-12143.
40. Heuer JG, Kaufman TC: Homeotic genes have specific functional roles in the establisment of the *Drosophila* embryonic peripheral nervous system. *Development* 1992, 115: 35-47.
41. St. Johnston D, Nusslein-Volhard C: The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 1992, 68: 201-219.
42. Micklem DR: mRNA localisation during development. *Dev Biol* 1995, 172: 377-395.
43. Shen J, Wu H, Gudas LJ: Molecular cloning and analysis of a group of genes differentially expressed in cells which overexpress the *Hoxa1* homeobox gene. *Exp Cell Res* 2000, 259: 274-283.
44. Carr A, Biggin MD: A comparison of *in vivo* and *in vitro* DNA-binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila* embryos. *EMBO J* 1999, 18: 1589-1608.
45. Cribbs DL, Benassayag C, Randazzo FM, Kaufman TC: 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* 1995, 14: 767-778.
46. Mann RS, Affolter M: Hox proteins meet more partners. *Curr Opin Genet Dev* 1998, 8: 423-429.
47. Mann RS: The specificity of homeotic gene function. *BioEssays* 1995, 17: 855-863.
48. Bryant Z, Subrahmanyan L, Tworoger M, LaTray L, Liu C-R, Li M-J, Van den Engh G, 49 Ruohola-Baker H: Characterization of differentially expressed genes in purified *Drosophila* follicle cells: Towards a general strategy for cell type-specific developmental analysis. *Proc Natl Acad Sci USA,* 1999, 96: 5559-5564.
White KP, Rifkin SA, Hurban P, Hogness DS: Microarray analysis of *Drosophila* development during metamorphosis. *Science* 1999, 286: 2179-2184.
Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG, Oliver B: Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res* 2000, 10: 2030-2043.
51. Campos-Ortega J, Hartenstein V: *The embryonic development of Drosophila melanogaster*. Heidelberg: Springer, 1997.
52. Walter MF, Petersen NS, Biessmann H: Heat shock causes the collapse of the intermediate filament cytoskeleton in *Drosophila* embryos. *Dev Genet* 1990, 11: 270-279.
53. Tautz D, Pfeifle C: 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* 1989, 98: 81-85.
54. Therianos S, Leuzinger S, Hirth F, Goodman CS, Reichert H: Embryonic development of the *Drosophila* brain: formation of commissural and descending pathways. *Development* 1995, 121: 3849-3860.
55. Grieder NC, Marty T, Ryoo H-D, Mann RS, Affolter M: Synergistic activation of a *Drosophila* enhacer by HOM/EXD and DPP signaling. *EMBO J* 1997, 16**:** 7402-7410.

56. Mahadevappa M, Warrington JA: A high-density probe array sample preparation method using 10- to 100- fold fewer cells. *Nat Biotechnol* 1999, 17: 1134-1136.

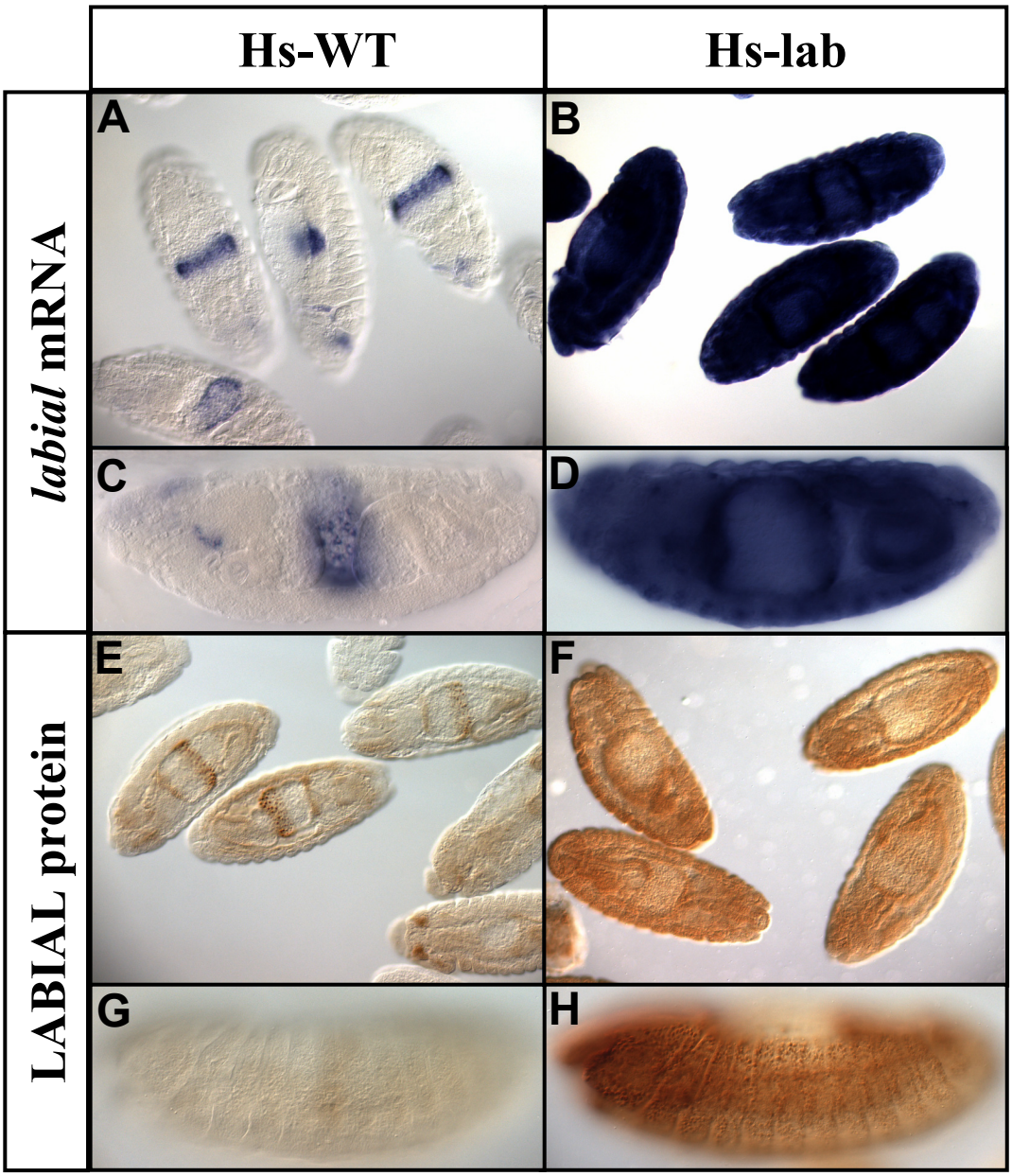



**Fig. 1.** Heat shock driven ubiquitous overexpression of *lab* monitored by *in situ* hybridization and immunocytochemistry. (A, B, C, D) RNA in situ hybridization; (E, F, G, H) immunocytochemical staining. *lab* gene expression is shown in heat shocked wild-type embryos (A, C, E, G) and in heat shocked 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.

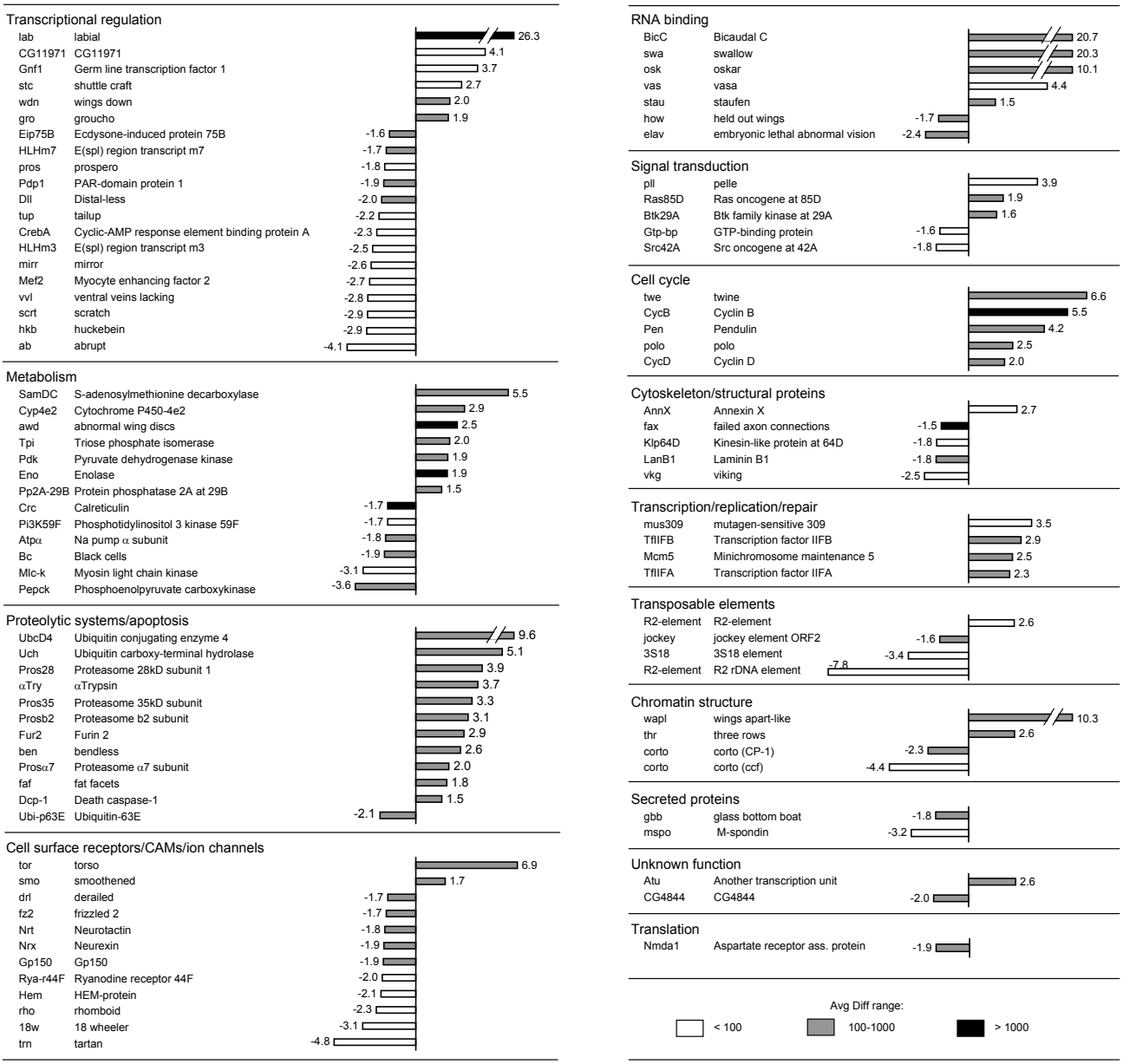



Fig. 2. Genes differentially expressed in response to heat shock induced overexpression of lab, grouped according to functional classes. Bars represent the fold-change between differentially expressed genes in heat shock treated wildtype embryos and heat shocked hs-lab embryos. Positive values indicate 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.

**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 |
| | ΣN = 1513 | Σn = 96 | | 48 | 48 |

Genes that are differentially expressed following heat induced ubiquitous overexpression of *lab* in stage 10-17 hs-*lab* embryos, grouped according to functional classses. (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 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 |

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.

# 5. Functional Equivalence of Hox Gene Products in the Specification of the Tritocerebrum during Embryonic Brain Development of *Drosophila*

Frank Hirth[1,*], Thomas Loop[1], Boris Egger[1], David F. B. Miller[2], Thomas C. Kaufman[2], and Heinrich Reichert[1]

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

[2]*Department of Biology, Howard Hughes Medical Institute, Indiana University, Jordan Hall A507, Bloomington, IN 47405-6801 USA*

*Correspondence: Dr. Frank Hirth, Institute of Zoology, Biocenter/Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland
Phone: (41-61) 2671617. Fax: (41-61) 2671613. e-mail: Frank.Hirth@unibas.ch

## Summary

Hox genes encode evolutionarily conserved transcription factors involved in the specification of segmental identity during embryonic development. This specification of identity is thought to be directed by differential Hox gene action, based on differential spatio-temporal expression patterns, protein sequence differences, interactions with cofactors, and regulation of specific downstream genes. During embryonic development of the Drosophila brain, the Hox gene labial is required for the regionalized specification of the tritocerebral neuromere; in the absence of labial the cells in this brain region do not acquire a neuronal identity and major axonal pathfinding deficits result. Here we use genetic rescue experiments to investigate the functional equivalence of the Drosophila Hox gene products in the specification of the tritocerebral neuromere. Using the Gal4-UAS system, we first demonstrate that the labial mutant brain phenotype can be rescued by targeted expression of the Labial protein under the control of CNS-specific labial regulatory elements. We then show that under the control of these CNS-specific regulatory elements all other Drosophila Hox gene products, except Abdominal-B, are able to efficiently replace Labial in the specification of the tritocerebral neuromere. We also observe a correlation between the rescue efficiency of the Hox proteins and the chromosomal arrangement of their encoding loci. Our results indicate that, despite considerably diverged sequences, most Hox proteins are functionally equivalent in their capability to replace Labial in the specification of neuronal identity. This suggests that in embryonic brain development, differences of Hox gene action rely mainly on cis-acting regulatory elements and not on Hox protein specificity.

## Introduction

The homeotic/Hox genes encode a network of evolutionarily conserved 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. This specification of identity is thought to be directed by differential Hox gene action, based on differential spatio-temporal expression patterns, protein sequence differences, interactions with cofactors, and regulation of specific downstream genes (Carroll, 1995; Graba et al., 1997; Gellon and McGinnis, 1998; Mann and Morata, 2000). The functional roles of Hox genes in insect development have been studied extensively in *Drosophila.* 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 in the clusters and their spatial and temporal expression pattern in the body; genes located towards the 3' end are expressed more anteriorly and earlier than genes towards the 5' end (spatial and temporal colinearity) (Manak and Scott, 1994; Duboule and Morata, 1994; Maconochie et al., 1996).

Hox genes are expressed in the developing brain and ventral nerve cord of *Drosophila* in an ordered set of domains. In the embryonic brain, specific Hox genes are expressed in the

posterior half of the tritocerebrum (and to a small extent in the deutocerebrum) as well as in the three subesophageal neuromeres. The tritocerebrum is the posterior neuromere of the supraesophageal ganglion and consists of two bilaterally symmetric hemiganglia that are bounded anteriorly by the deutocerebrum and are linked by the tritocerebral commissure that runs across the midline beneath the gut (Burrows, 1996; Reichert and Boyan, 1997). The tritocerebrum is connected to more posterior parts of the brain through longitudinal connectives and forms projections to the frontal ganglion via the frontal connectives. The Hox gene that is specifically expressed in the posterior half of the tritocerebral neuromere is *labial (lab)*. Loss-of-function *lab* mutations cause profound defects in the establishment of the tritocerebral neuromere (Hirth et al., 1998). In *lab* mutants, the tritocerebral commissure is missing and the longitudinal connectives are reduced or absent. Moreover, the cells in the *lab* mutant domain do not acquire a neuronal identity as exemplified by the lack of expression of neuronal markers indicating that *lab* is required for the specification of neuronal identity in the tritocerebrum. Comparable effects are seen in *Deformed* mutants, the only major difference being that these effects were observed in the mandibular and anterior maxillary brain neuromere which is the expression domain of *Deformed*. None of the other Hox gene mutants show comparable brain defects (Hirth et al., 1998).

Here we use genetic rescue experiments to investigate the functional equivalence of all of the *Drosophila* Hox genes in specifying the neuronal identity in the tritocerebral neuromere. For this we use the Gal4-UAS system (Brand and Perrimon, 1993) for targeted misexpression of Hox genes in the posterior tritocerebral domain (in which *lab* is normally expressed) of *lab* null mutants. As expected, we find that the *lab* mutant brain phenotype can be rescued by targeted expression of the Lab protein under the control of CNS-specific *lab* regulatory elements. We then demonstrate that under the control of these CNS-specific regulatory elements most of the other *Drosophila* Hox gene products are also able to replace the Lab protein in the specification of the tritocerebral neuromere. Only the Abdominal-B protein does not efficiently rescue the *lab* mutant phenotype in the brain. For the other Hox proteins, we observe a correlation between their efficiency of rescue the *lab* mutant brain phenotype and the chromosomal arrangement of their encoding loci. Our results indicate that, despite considerably diverged sequences, most Hox proteins are functionally equivalent in their capability to replace Labial in the specification of neuronal identity in the brain. This suggests that differences of Hox gene action in brain development rely mainly on cis-acting regulatory elements and not on Hox protein specificity.

## Materials and Methods

### Fly Strains and Genetics

The P{w+ lab::Gal4}K5J2 driver was generated by cloning a genomic fragment from labial that extends from the HindIII site 3.6 Kb upstream of the transcriptional start site downstream to the BssHII site at +10 bp (Chouinard and Kaufman, 1991). The downstream site was

converted using a HindIII linker creating a 3.6 Kb HindIII fragment that was cloned into this site in pGaTN (Brand and Perrimon, 1993). This plasmid was then cut with NotI to remove the lab::Gal4 cassette, cloned into pCosperNot (supplied by John Tamkun) and used to generate the transgenic line P{w$^+$ *lab*::Gal4}K5J2.

For *lab*::Gal4 specific targeted misexpression of *proboscipedia (pb), Deformed (Dfd), abdominal-A (abd-A),* and *Abdominal-B (Abd-B)* in *lab* mutant embryos, the following UAS::Hox responder lines were used: p[*UAS*::*pb* 49.1] homozygous on chromosome II (Aplin and Kaufman, 1997); p[*UAS*::*Dfd*] homozygous on chromosome II (Brown et al.,1999); p[*UAS*::*abd*-A 21.6] homozygous on chromosome I (Greig and Akam, 1993), supplied by M. Akam; p[*UAS*::*Abd-B*m] homozygous on chromosome II (Castelli-Gair et al., 1994) driving the expression of the Abd-Bm form (Casanova et al. 1986; Zavortink and Sakonju 1989), supplied by M. Akam.

For *lab*::Gal4 specific targeted misexpression of *labial (lab), Sex combs reduced (Scr), Antennapedia (Antp),* and *Ultrabithorax (Ubx)* in *lab* mutant embryos, p[*UAS*::*lab*], p[*UAS*::*Scr*], p[*UAS*::*Antp*], and p[*UAS*::*Ubx*] responder lines were generated (see also Miller et al., in press). The respective Hox cDNAs were cloned into a polylinker downstream from a minimal *hsp70* promoter of the Gal4 responder plasmid pUAST (Brand and Perrimon, 1993) which contains a P-element with the *white* mini-gene as a marker. The *hsp70* promoter is activated in the presence of Gal4 due to five upstream Gal4 binding sites (UAS). For generating p[*UAS*::*lab*], a 2.1 kb cDNA derived from a 2.4a minigene (including the second intron; Chouinard and Kaufman, 1991) encompassing the entire *lab* coding region, was digested with the SspI to generate the 2.1Kb cDNA which was inserted into pBlueScriptKS+ (Stratagene) at the EcoRV site. The cDNA was subsequently removed with EcoRI(5') and KpnI(3') for insertion into pUAST at the same sites. For generating p[*UAS*::*Scr*], the 1.2 Kb BamHI(5') and MluI(3') truncated Scr L3 cDNA (Mahaffey and Kaufman, 1987) was inserted into pSE280 (Invitrogen) using the same sites. A partial Scr cDNA was then removed from pSE280 with NcoI(5'), blunted with Klenow, and then released with XhoI. This modified cDNA was inserted into pUAST at the Klenow blunted EcoRI and XhoI sites. For generating p[*UAS*::*Antp*], the entire Antp G1100 cDNA (Scott et al., 1983) was inserted into pUAST at the EcoRI site. For generating p[*UAS*::*Ubx*], the previously reported Ubx NAB3 cDNA containing isoform 1S, which is the predominant embryonic cDNA (O'Connor et al., 1988), was inserted into pUAST at the EcoRI site. All strains as well as all experimental genotypes were maintained in standard laboratory cultures at 25°C.

Control experiments verified that the P{w$^+$ *lab*::Gal4}K5J2 driver is expressed in a spatial pattern, which corresponds to that of endogenous *lab* in the procephalon, and in the tritocerebral neuromere. UAS::transgene activation in the procephalon is delayed for 2.5hrs as compared to earliest presence of endogenous Lab protein (Kaufman et al., 1990), thus under the control of P{w$^+$ *lab*::Gal4}K5J2, UAS::responder activation starts at late stage 10 (5h-5.5h AEL; Campos-Ortega and Hartenstein, 1997). Phenotypic penetrance of the *lab* mutant brain phenotype was 88.6% (n =209) as determined with the *lab* null allele *lab*$^{vd1}$

(Merrill et al., 1989; Hirth et al., 1998) using flies of the genotype *lab*$^{vd1}$/TM6B-*UbxlacZ*. The ability of the Hox proteins to rescue the *lab* mutant brain phenotype was determined by crossing P{w$^{+}$ *lab*::Gal4}K5J2; *lab*$^{vd1}$/TM6B-*UbxlacZ* to either P{UAS::*lab*}, *lab*$^{vd1}$/TM3-*AntplacZ* or to flies of genotype P{UAS::*Hox*}, *lab*$^{vd1}$/ TM6B-*UbxlacZ* where *Hox* = *pb*, *Dfd*, *Antp*, and *Abd-B*; or to flies of genotype, P{UAS::*Hox*}/+; *lab*$^{vd1}$/+ for *Hox* = *Scr* , *Ubx* and *abd-A*. All rescue experiments reported here were carried out at 25°C; no significant differences in rescue efficency were obtained when rescue experiments were carried out at 28°C. To identify rescued *lab*$^{-/-}$ cells and their axonal projection pattern, *UAS::tau-lacZ* located on the X chromosome (Callahan and Thomas, 1994) was additionally crossed in.

### Immunocytochemistry and Genetic Rescue Analysis

Whole-mount immunocytochemistry and laser confocal microscopy was performed as previously described (Hirth et al., 1998). In genetic rescue experiments, P{w$^{+}$ *lab*::Gal4}K5J2 driven P{UAS::*Hox*} activity in homozygous *lab* null mutants *(lab*$^{vd1}$*/lab*$^{vd1}$*)* was confirmed by the absence of balancer-specific (TM6B-*UbxlacZ;* TM3-*AntplacZ*) β–gal and/or Labial immunoreactivity as well as by the presence of corresponding Hox immunoreactivity in the tritocerebral *lab* domain. The criteria used to judge *lab*$^{-/-}$ embryos as fully rescued were (1) the presence of the tritocerebral commissure linking the two tritocerebral hemiganglia, (2) the restoration of the longitudinal pathways between the supra- and subesophageal ganglia, and (3) the expression of neuron-specific molecular labels as assayed by anti-HRP and anti-Elav immunoreactivity (Hirth et al., 1998). Only when all three criteria were fulfilled, the tritocerebrum of a *lab*$^{-/-}$ mutant embryo was scored as rescued. Additionally, in embryos of the genotype *UAS::tau-lacZ/+; lab::Gal4/UAS::Hox; lab*$^{-/-}$, the specificity of rescue was also determined by the presence of correct axonal projections of rescued *lab*$^{-/-}$cells along the rescued tritocerebral commissure.

### Laser Confocal Microscopy

For laser confocal microscopy, a Leica TCS SP was used. Optical sections ranged from 0.4 to 2 µm recorded in line average mode with picture size of 512 x 512 pixels. Captured images from optical sections were arranged and processed using IMARIS (Bitplane). Figures were arranged and labeled using Adobe Photoshop.

## Results

In the embryonic brain of *Drosophila*, the *labial (lab)* gene is expressed in the posterior half of the tritocerebral neuromere (Fig. 1A-D). In *lab* loss-of-function mutants, regionalized axonal patterning defects occur in the *lab* domain which are due to both cell-autonomous effects and non cell-autonomous effects. Thus, in the absence of *lab*, mutant cells are generated and positioned corrrectly in the brain, but these cells do not extend axons.

Moreover, extending axons from other neighboring wild-type neurons stop at the mutant domains or project ectopically. As a result, dramatic defects in commissural and longitudinal axon pathways occur (Hirth et al., 1998); the tritocerebral commissure, which links the two tritocerebral hemiganglia, is absent and the longitudinal pathways between the supraesophageal and subesophageal ganglia are reduced or absent (Fig. 1E, F). Immunocytochemical analysis demonstrates that cells in the mutant domain do not express any of the numerous neuronal markers such as ELAV that positionally equivalent cells express in the wildtype, indicating a complete lack of neuronal identity in the *lab* mutant brain domain (Hirth et al., 1998). This strong mutant phenotype is apparent in 88.6% of the cases (n = 209). These data indicate that *lab* is involved in the specification of tritocerebral neuronal identity in the *Drosophila* brain.

In order to carry out a genetic rescue of the mutant brain phenotype in *lab* mutant embryos, we made use of the Gal4-UAS system (Brand and Perrimon, 1993). For this, a transgenic fly line carrying a Gal4 transcriptional activator under the control of the *lab* promoter together with CNS-specific upstream enhancer elements of the *lab* gene was used (see Methods). By crossing this *lab::Gal4* line to different UAS–responders it is possible to express the responder constructs in a pattern that corresponds to that of the endogenous *lab* gene. To verify this we first crossed the *lab::Gal4* line to transgenic lines carrying a UAS::*taulacZ* (Callahan and Thomas, 1994) reporter construct. The spatial expression domain of this reporter construct in the embryonic brain mimicked the endogeneous *labial* expression domain (Fig 2. A, B). Spatially localized expression domains were seen in the posterior parts of the tritocerebral neuromere. (Ectopic reporter expression was seen in a small number of individual cells in the deutocerebral and mandibular neuromeres.) Double immunostaining experiments using anti-lacZ and anti-Lab antibodies confirmed that lacZ expression occurred in the axons and cortical cytoskeleton of those cells that showed nuclear Lab expression (Fig. 2 C, D).

We next determined whether the *labial* mutant brain phenotype could be rescued by transgenic expression of the Lab protein in a *labial* null mutant background. For this a UAS::*lab* responder was driven by the *lab::Gal4* driver in the tritocerebral *lab* mutant domain. Using this approach, we obtained efficient rescue of all of the tritocerebral defects in the *lab* mutants. Thus, in these rescued embryonic brains the tritocerebral commissure was present, the longitudinal pathways between the supra- and subesophageal ganglia were restored, and cells in the mutant domain showed correct neuron-specific molecular labels as revealed by anti-Elav (not shown) and anti-HRP immunoreactivity (Fig. 3 A, B). A quantification of the rescue efficiency for Lab in these experiments is given in table 1. The fact that in these experiments Lab protein was indeed expressed specifically in the tritocerebral domain was demonstrated by carrying out anti-Lab immunostaining on these rescued brains (Fig. 3 C, D).

To determine whether other members of the Hox gene complex might also be able to rescue the *lab* mutant brain defects and, thus, be functionally equivalent to Lab in determining the

segmental identity of the tritocerebral neuromere, transgenic lines were used in which the coding sequence of each of the remaining 7 Hox genes was placed under UAS control (see Methods and Miller et al., 2001). As a control, we first determined whether *lab*::Gal4 driven misexpression of any of the 8 Hox proteins in a *lab*+ background had any effects on the development and specification of the tritocerebral lab domain. In none of these experiments did we detect any sign of morphological abnormalities in the tritocerebrum or in any other part of the embryonic brain. Thus, in *lab::*Gal4/UAS*::Hox; lab*+ embryos, all labeled structures in the tritocerebral *lab* domain were normal. Moreover*, lab::Gal4* driven *UAS::taulacZ* reporter gene expression in conjunction with Hox gene misexpression in a *labial*+ background revealed that the tritocerebral *lab*+ cells showed a wildtype-like axonal projection pattern.

Next, we expressed each of the remaining 7 UAS::*Hox* responders under the control of the *lab*::Gal4 driver in the *lab* mutant domain. We first investigated the Hox proteins of the Antennapedia-Complex, since in the wildtype, all five proteins of this complex are expressed in specific domains of the developing brain (Hirth et al., 1998). Surprisingly, all of the Antennapedia-Complex Hox proteins were able to rescue the *lab* mutant brain defects in these experiments. Examples of the ability of these Hox proteins to rescue the *labial* mutant brain phenotype are shown for Sex combs reduced (Scr) and Antennapedia (Antp) (Fig. 4). In both cases, an efficient rescue of the tritocerebral defects in the *lab* mutants was obtained; the tritocerebral commissure was present, the longitudinal pathways were restored, and cells in the mutant domain showed correct neuron-specific molecular labels. In addition to the *lab*::Gal4 driven ectopic expression of Scr and Antp in the tritocerebral *labial* mutant domain, the large endogenous expression domains of these genes were observed unchanged in the subesophageal ganglion for Scr, and in the subesophageal ganglion and ventral nerve cord for Antp (Fig. 4 C, F) (Hirth et al., 1998). A quantification of the rescue efficiency for all of the Antennapedia-Complex Hox proteins in these experiments is given in table 1.

We next investigated the rescue potential of the Hox proteins of the Bithorax-Complex in comparable experiments. In contrast to the Hox proteins of the Antennapedia-Complex, the Bithorax-Complex Hox proteins are not expressed in the developing brain of the wildtype, rather their expression domains are restricted to the ganglia of the ventral nerve cord (Hirth et al., 1998). Remarkably, as was the case for the Antennapedia-Complex proteins, both the Ultrabithorax (Ubx) and the Abdominal-A (Abd-A) gene products of the Bithorax-Complex were able to rescue the *lab* mutant brain defects in these experiments. Once again, an efficient rescue of the tritocerebral brain defects in the *lab* mutants was obtained; the tritocerebral commissure was present, the longitudinal pathways were restored, and cells in the mutant domain showed correct neuron-specific molecular labels. An example of the ability of these Hox proteins to rescue the *labial* mutant brain phenotype is shown for Ubx (Fig. 5 A-C). Note that, in addition to the *lab*::Gal4 driven ectopic expression of Ubx in the tritocerebral *labial* mutant domain, the endogenous Ubx expression domain in the ventral nerve cord is also seen (Fig. 5 C) (Hirth et al., 1998). In contrast to the other two Bithorax-Complex Hox proteins, use of the Abdominal-B (Abd-B) gene product did not result in an

efficient rescue of the tritocerebral defects in the *lab* mutants. In over 90% of the *lab*::Gal4/UAS::*Abd-B*; $lab^{-/-}$ mutant embryos, profound axonal projection deficits were observed in the brain; the tritocerebral brain commissure was absent, the longitudinal brain pathways were reduced or lacking, and cells in the mutant domain lacked correct neuron-specific molecular labels (Fig. 5 D-F). A quantification of the rescue efficiency for all of the Bithorax-Complex Hox proteins in these experiments is given in table 1.

The efficient rescue of the tritocerebral defects in the *lab* mutants, which is achieved by targeted misexpression of 7 out of 8 Hox genes is striking; in the rescued embryonic brains the tritocerebral commissure was present, the longitudinal pathways between the supra- and subesophageal ganglia were restored, and cells in the mutant domain showed correct neuron-specific molecular labels. However, it is conceivable, that the rescue of all of these neuronal structures might be due to a restoration of generic neuronal properties in the cells of the lab mutant domain and not due to the rescue of specific neuronal identities in these cells. To investigate this, we determined whether the rescued cells in the $lab^{-/-}$ domain project their axons correctly accross the rescued tritocerebral commissure, as is the case for lab-expressing neurons in the wildtype brain. For this, we coexpressed a *UAS::taulacZ* reporter gene with each UAS::*Hox* responder in the tritocerebral *lab* mutant domain using the *lab::Gal4* driver. This coexpression makes it possible to visualize both the cell bodies and the axonal projections of the rescued $lab^{-/-}$ cells. For all of the Hox gene products except AbdB these experiments demonstrated that the rescued tritocerebral $lab^{-/-}$ cells were again able to extend axons that projected correctly along the rescued tritocerebral commissure (Fig. 6).

As is indicated in Table 1, the relative efficiency of rescue of the brain phenotype in *lab* mutants varied systematically for the different Hox proteins. The *lab* responder achieved the best rescue efficiency while the other Hox responders had slightly lower rescue efficiencies. Figure 7 shows the rescue efficiency of all other Hox proteins relative to the rescue efficiency of Lab, which was taken as 100%. Interestingly, the decline in relative rescue efficiency for these other Hox proteins appears to be co-linear (Lab>Pb>Dfd>Scr>Antp>Ubx>Abd-A) in that it reflects the proximal to distal arrangement of their encoding loci on the chromosome.

## Discussion

Our findings indicate that Proboscipedia (Pb), Deformed (Dfd), Sex combs reduced (Scr), Antennapedia (Antp), Ultrabithorax (Ubx), and Abdominal-A (Abd-A), but not Abdominal-B (Abd-B), are able to efficiently substitute for Lab in determining the segmental identity of the *Drosophila* brain. Morphological evidence for a homeotic transformation of the tritocerebral neuromere into one of a different segmental identity was not observed in any of these rescue experiments. This suggests, that all of the Hox proteins, with the exception of Abd-B, are to a large degree functionally equivalent to Lab in this aspect of embryonic brain development. This surprising functional equivalence contrasts with the general notion, derived from experiments on the specification of other body parts in *Drosophila*, that Hox proteins assign different identities along the anterior-posterior body axis by acting as specific selectors of different, alternative developmental pathways (Garcia-Bellido, 1975; Lawrence and Morata, 1994; Gellon and McGinnis, 1998; Mann and Morata, 2000).

There are several possible explanations for this discrepancy. First, the functional role of Lab in the specification of neuronal identity in the brain may differ from the role of other Hox proteins in other parts of the CNS. For example, Hirth et al. (1998) find that the loss-of-function phenotype of *lab* (and *Dfd*) in the embryonic CNS differs from that of the remaining Hox genes. Moreover, in contrast to other domains in the embryonic CNS, there is an absence of overlapping expression with other Hox proteins, so that there is no genetic "backup" in the tritocerebrum. Similar oberservations on Lab have been made in epidermal structures (reviewed by Kaufman et al., 1990; McGinnis and Krumlauf, 1992; Morata, 1993). Second, it is conceivable that all Hox proteins can specify neuronal identity and the generic formation of commissural and longitudinal connections in the CNS. However, this seems unlikely, since the morphology and innervation of the triocerebral neuromere is unique and highly specific and unlike that of any other neuromere in the CNS (Burrows, 1996). Similarly, the morphology, mode of formation and gut-specific association of the developing tritocerebral commissure is clearly different from that of the other ganglionic commissures in the embryonic CNS (e.g. Wildemann et al, 1997). Third, Hox proteins may indeed be to a larger degree functionally interchangable in the CNS than hitherto expected. In this respect, two sets of recent functional complementation experiments carried out on mammalian *Hoxa3/Hoxd3* genes and on mammalian *Hox11a/Hox11d* genes are noteworthy since they indicate that paralogous gene products can carry out identical biological functions if they are placed unter the control of the appropriate cis-acting regulatory elements (Zakany et al., 1996; Greer et al., 2000). Our results extend this notion of functional equivalance of Hox genes from the level of paralogous genes to the level of the entire Hox gene cluster, excepting *Abd-B*. This, in turn, suggests that almost all of the Hox proteins can carry out identical biological functions in the *Drosophila* brain, if they are under the control of the same cis-acting regulatory elements.

In our experiments, all of the Hox responders were expressed in the *lab* mutant under the control of the identical, *lab*-specific regulatory elements. Under these circumstances, the Lab

responder achieved the best rescue efficiency, while the other Hox responders (with the exception of Abd-B) had somewhat lower rescue efficiencies which ranged from 86%-59% of the rescue values achieved by Lab (see Fig. 7). Interestingly, the relative rescue efficiency of the Hox gene products Lab, Pb, Dfd, Scr, Antp, Ubx and Abd-A reflect their proximal to distal arrangement of their encoding loci on the chromosome. It is conceivable, that this co-linear correlation of rescue efficiency among theses Hox gene products is due to the variability in the Gal4-UAS system, to positional effects of transgene insertions, or to differences in transgene expression levels. However, a more reasonable explanation is, that the decline in relative rescue efficiency among these Hox proteins, as well as the qualitative difference between Abd-B and the other Hox proteins in their ability to rescue the *lab* mutant brain phenotpye, is due primarily to Hox protein sequence differences. Hox proteins do indeed show sequence differences in several respects. The most notable differences reside in the homeodomain, the hexapeptide motif (lacking in Abd-B), and the linker lengths between the homeodomain and the hexapeptide motif (Gehring et al., 1994; Duboule, 1994; Mann, 1995; Chan et al., 1996; Mann and Chan, 1996; Piper et al., 1999; Passner et al., 1999).

We posit that the findings reported here have implications for understanding Hox gene function and evolution. The functional equivalence of almost all of the Hox proteins in brain neuromere specification implies that the specificity of Hox gene action is achieved mainly through regulatory elements that control position, timing and level of Hox gene expression and only to a lesser degree through Hox protein sequence differences. Similar findings have been obtained in studies on Pax gene interchangeability in *Drosophila* (Li and Noll, 1994). Thus, the genes *paired* and *gooseberry*, which have distinct developmental roles in embryogenesis and which have considerably diverged coding sequences, can exert the same conserved function in genetic rescue experiments. Comparable findings have recently been reported in mammals (Bouchard et al., 2000), corroborating the notion put forward by Noll that the essential difference among these developmental regulatory genes of the same family may reside in their cis-regulatory regions.

The fact that the expression of different Hox genes in the *lab* mutant domain does not cause homeotic transformation of tritocerebral identity, suggests that Hox proteins act as "mediators" rather than as "selectors" within the developmental pathway that specifies segmental neuronal identity in the *Drosophila* brain. Recent experiments using both loss- and gain-of-function mutations suggest that this also applies to the specification of other structures along the antero-posterior body axis of *Drosophila.* For example, in haltere development, *abd-A* and to some extent *Abd-B* can substitute for *Ubx* gene action (Casares et al., 1996). Moreover, a comparable lack of Hox gene specificity has been observed in gonad development (Greig and Akam, 1995).

Finally, the high degree of functional interchangeability of Lab and all of the other *Drosophila* Hox proteins, with the exception of Abd-B, is consistent with evolutionary studies which propose a common origin of all of the Hox genes from a single ancestral progenitor and an early singularity of *Abd-B*-like genes in the ancestral Hox gene cluster (Schubert et al.,

1993). Given the striking evolutionary conservation of structure, expression and brain-specific function of *lab* and its mammalian *Hox1* orthologs (Hirth and Reichert, 1999; Reichert and Simeone, 1999), it will now be important to determine whether functional equivalence among non-paralogous Hox gene products is also valid for vertebrate hindbrain development.

## Acknowledgements

We thank M. Affolter, M. Akam, D. Brower, the Bloomington stock center, S. Celniker, I. Duncan, W. J. Gehring, C. Goodman, A. Macias, N. & W. McGinnis, K. Matthews, G. Morata, U. Nussbaumer, J. Tamkun, J. B. Thomas, and R. White for flies and antibodies. Supported by the SNSF and the EU-BIOTECH program.

## References


**Aplin, A.C. and Kaufman, T.C.** (1997). Homeotic transformation of legs to mouthparts by *proboscipedia* expression in *Drosophila* imaginal discs. *Mech. Dev.* **62**, 51-60.

**Bouchard, M., Pfeffer, P., and Busslinger, M.** (2000). Functional equivalence of the transcription factors Pax2 and Pax5 in mouse development. *Development* **127**, 3703-13.

**Brand, A. H. and Perrimon, N.** (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.

**Brown, S., Holtzman, S., Kaufman, T. C., and Denell, R.** (1999). Characterization of the Tribolium Deformed ortholog and its ability to directly regulate Deformed target genes in the rescue of a *Drosophila* Deformed null mutant. *Dev. Genes Evol.* **209**, 389-398.

**Burrows, M.** *The Neurobiology of an Insect Brain* (Oxford University Press, 1996).

**Callahan, C. A. and Thomas, J. B.** (1994). Tau-beta-galactosidase, an axon-targeted fusion protein. *Proc. Natl. Acad. Sci. USA* **91**, 5972-5976.

**Campos-Ortega, J. A. and Hartenstein, V.** *The Embryonic Development of Drosophila melanogaster* (2nd edition, Springer, Heidelberg, 1997).

**Carroll, S. B.** (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479-485.

**Casanova, J., E. Sanchez-Herrero, and G. Morata** (1986). Identification and characterisation of a parasegment specific regulatory element of the *Abdominal-B* gene of *Drosophila*. *Cell* **47**, 627-636.

**Casares, F., Calleja, M., and Sanchez-Herrero, E.** (1996). Functional similaritiy in appendage specification by *Ultrabithorax* and *abdominal-A Drosophila* HOX genes. *EMBO J.* **15**, 3934-3942.

**Castelli-Gair, J., Greig, S., Micklem, G., and Akam, M.** (1994). Dissecting the temporal requirements for homeotic gene function. *Development* **120**, 1983-1995.

**Chan, S.-K., Pöpperl, H., Krumlauf, R., and Mann, R. S.** (1996). An *extradenticle*-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J.* **15**, 2476-2487.

**Chouinard, S. and Kaufman, T. C.** (1991). Control of expression of the homeotic *labial* (*lab*) locus of *Drosophila* melanogaster: evidence for both positive and negative autogenous regulation. *Development* **113**, 1267-1280.

**Duboule, D.** *Guidebook to the Homeobox Genes* (Oxford University Press, 1994).

**Duboule, D. and Morata, G.** (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358-364.

**Garcia-Bellido, A.** (1975). Genetic control of wing disk development in *Drosophila*. *Ciba Foundation Symp.* **29**, 161-182.

**Gehring, W. J.** *et al*. (1994). Homeodomain-DNA recognition. *Cell* **78**, 211-223.

**Gellon, G. and McGinnis, W.** (1998). Shaping animal body plans in development and evolution by modulation of *Hox* expression patterns. *BioEssays* **20**, 116-125.

**Graba, Y., Aragnol, D., and Pradel, J.** (1997). *Drosophila* Hox complex downstream targets and the function of homeotic genes. *BioEssays* **19**, 379-388.

**Greer, J. M., Puetz, J., Thomas, K. R., and Capecchi, M. R.** (2000). Maintenance of functional equivalence during paralogous *Hox* gene evolution. *Nature* **403**, 661-665.
**Greig, S. and Akam, M.** (1993). Homeotic genes autonomously specify one aspect of pattern in the *Drosophila* mesoderm. *Nature* **362**, 630-632.
**Greig, S. and Akam, M.** (1995). The role of homeotic genes in the specification of the *Drosophila* gonad. *Curr. Biol.* **5**, 1057-1062.
**Hirth, F., Hartmann, B., and Reichert, H.** (1998). Homeotic gene action in embryonic brain development of *Drosophila*. *Development* **125**, 1579-1589.
**Hirth, F. and Reichert, H.** (1999). Conserved genetic programs in insect and mammalian brain development. *BioEssays* **21**, 677-684.
**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**, 309-362.
**Lawrence, P. and Morata, G.** (1994). Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* **78**, 181-189.
**Li, X. and Noll, M.** (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* **367**, 83-87.
**McGinnis , W., and Krumlauf, R.** (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302.
**Maconochie, M., Nonchev, S., Morrison, A., and Krumlauf, R.** (1996). Paralogous *Hox* genes: function and regulation. *Annu. Rev. Genet.* **30**, 529-556.
**Mahaffey, J. W. and Kaufman, T. C.** (1987). Distribution of the *Sex combs reduced* gene products in *Drosophila melanogaster*. *Genetics* **117**, 51-60.
**Manak, J. R. and Scott, M. P.** (1994). A class act: conservation of homeodomain protein functions. *Development Suppl*., 61-71.
**Mann, R.** (1995). The specificity of homeotic gene function. *BioEssays* **17**, 855-863.
**Mann R. S. and Chan, S.-K.** (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet.* **12**, 258-262.
**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**, 243-271.
**Merrill, V. K. L., Diederich, R. J., Turner, F. R., and Kaufman, T. C.** (1989). A genetic and developmental analysis of mutations in *lab*, a gene necessary for proper head formation in *Drosophila melanogaster*. *Dev. Biol.* **135**, 376-391.
**Miller, D. F. B., Rogers, B. T., Kalkbrenner, A., Hamilton, B., Holtzman, S. L., and Kaufman, T. C.** (2000). Cross-regulation of Hox genes in the *Drosophila melanogaster* Embryo. *Mech. Dev,* 102; 3-16.
**Morata, G.** (1993). Homeotic genes of *Drosophila*. *Curr. Opin. Genet. Dev.* **3**, 606-614.
**O'Connor, M. B., Benari, R., Perkins, L. A., and Bender, W.** (1988). Alternative RNA products from the *Ultrabithorax* domain of the bithorax complex. *EMBO J.* **7**, 435-445.
**Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K.** (1999). Structure of a DNA-bound Ultrabithroax-Extradenticle homeodomain complex. *Nature* **397**, 714-719.
**Piper, D. E., Batchelor, A. H., Chang, C.-P., Cleary, M. L., and Wolberger, C.** (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: Role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96**, 587-597.

**Reichert, H. and Boyan, G.** (1997). Building a brain: developmental insights in insects. *Trends Neurosci.* **20**, 258-263.

**Reichert, H. and Simeone, A.** (1999). Conserved usage of gap and homeotic genes in patterning the CNS. *Curr. Opin. Neurobiol.* **9**, 589-595.

Scott, M. P., Weiner, A. J., Hazelrigg, T. I., Polisky, B. A., Pirrotta, V., Scalenghe, F., and Kaufman, T. C. (1983). The molecular organization of the *Antennapedia* locus of *Drosophila*. *Cell* 35, 763-776.

**Schubert, F. R., Nieselt-Struwe, K., and Gruss, P.** (1993). The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl. Acad. Sci. USA* **90**, 143-147.

**Wildemann, B., Reichert, H., and Bicker, G.** (1997). Embryonic brain tract formation in *Drosophila melanogaster*. *Dev. Genes Evol.* **206**, 536-540.

**Zakany, J., Gerard, M., Favier, B., Potter, S. S., and Duboule, D.** (1996). Functional equivalence and rescue among group 11 Hox gene products in vertebral patterning. *Dev. Biol.* **176**, 325-328.

**Zavortink, M. and Sakonju, S.** (1989). The morphogenetic and regulatory functions of the *Drosophila Abdominal-B* gene are encoded in overlapping RNAs transcribed from separate promoters. *Genes Dev.* **3**, 1969-1981

**Table 1**

**Rescue of brain defects in *lab* mutants by Hox transgene expression**

| Hox protein | Lab | Pb | Dfd | Scr | Antp | Ubx | AbdA | AbdB |
|---|---|---|---|---|---|---|---|---|
| N examined | 132 | 145 | 145 | 142 | 149 | 134 | 138 | 165 |
| N rescued | 79 | 77 | 73 | 68 | 69 | 59 | 55 | 12 |
| % rescued | 59.8 | 53.1 | 50.3 | 47.8 | 46.3 | 44 | 39.8 | 7.2 |
| % corrected | 48.4 | 41.7 | 38.9 | 36.4 | 34.9 | 32.6 | 28.4 | 0 |

Quantitative rescue efficiency of *lab* mutant brain defects by the Hox gene products Lab, Pb, Dfd, Scr, Antp, Ubx, Abd-A, and Abd-B expressed in the *lab* mutant under the control of the same *lab*-specific cis-acting regulatory elements. The number of embryos examined (N examined), the number of examined embryos showing a complete rescue of the tritocerebral brain defects (N rescued), the percentage of embryos showing a complete rescue of the tritocerebral brain defects (% rescued), and the corrected percentage values for a rescue of the tritocerebral brain defects (% corrected) are shown. Percentage values were corrected in order to take account of the phenotypic penetrance of the lab mutation in tritocerebral development (88.6%). Thus, the corrected percentage values were calculated by substracting 11.4% from the uncorrected percentage values.

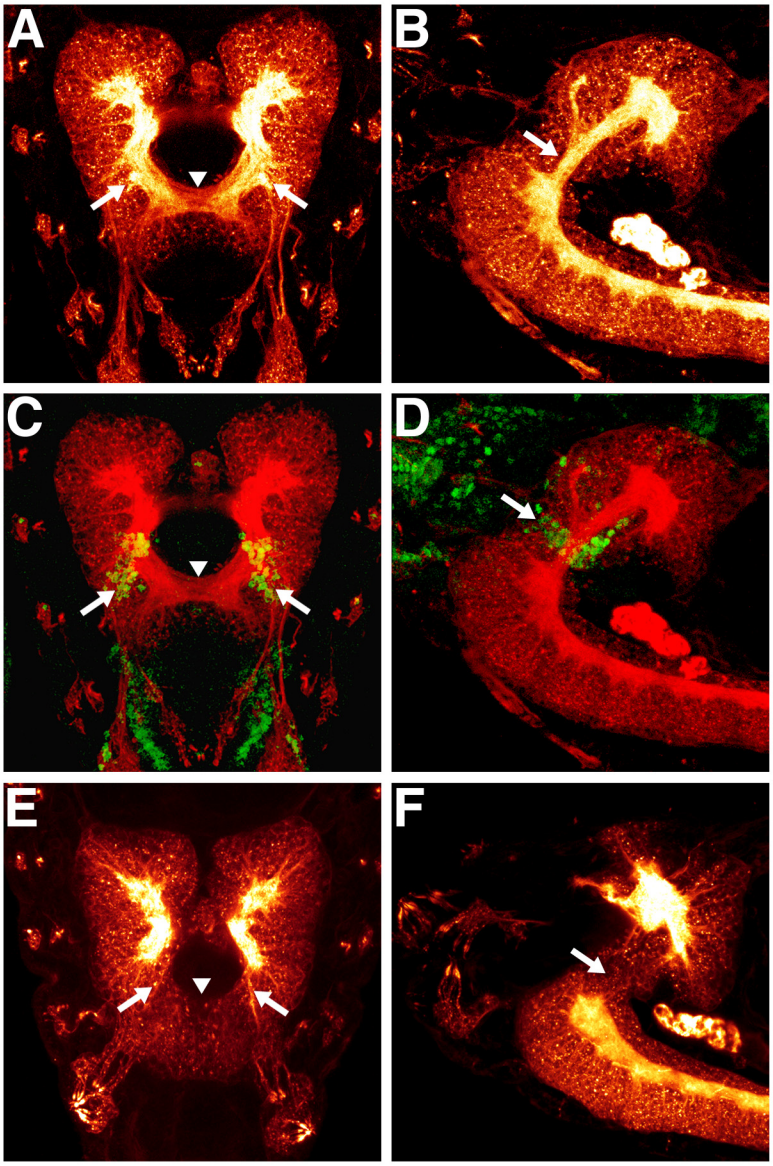



**Figure 1**
Expression of Labial and *labial* loss-of-function phenotype in the *Drosophila* embryonic brain. Laser confocal microscopy of stage 15 embryos, reconstructions of optical sections. (A, C, E) frontal views, (B, D, F) lateral views. (A, B) Wildtype embryonic brain. Anti-HRP immunolabeling. Arrows indicate circumesophageal connectives, arrowhead indicates tritocerebral commissure. (C, D) Wildtype embryonic brain. Double immunolabeling with anti-HRP (red) and anti-Lab (green). Arrows indicate Lab expression domain, arrowhead indicates tritocerebral commissure. Same embryo as in (A, B). (E,F) *lab* loss-of-function mutant embryonic brain. Anti-HRP immunolabeling. Arrows indicate missing circumesophageal connectives, arrowhead indicates missing tritocerebral commissure.

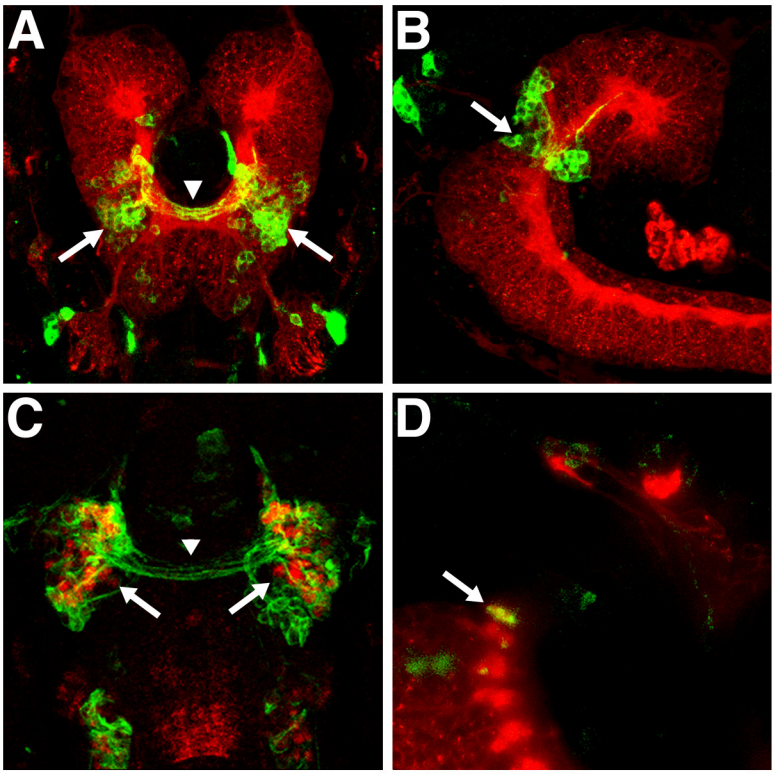



**Figure 2**
Reporter gene expression phenotype in the *Drosophila* embryonic brain. Characterization of brain-specific *lab*::Gal4 driver K5J2 using P{w+; *lab*::Gal4} K5J2 driven *UAS::taulacZ* reporter gene expression (Callahan and Thomas (1994). Laser confocal microscopy of stage 15 embryos, reconstructions of optical sections. (A, C) frontal views, (B) lateral view, (D) midline crossection. (A, B, D) Double immunolabeling with anti-HRP (red) and anti-β-Gal (green). P{w+ *lab*::Gal4}K5J2 driven *UAS::taulacZ* reporter gene expression is seen in the cortical cytoskeleton and axons of cells in the endogenous tritocerebral Lab expression domain (arrows in A, B) of the wildtype embryonic brain. Arrowhead indicates tritocerebral commissure. Ectopic reporter gene expression is seen in a small number of cells in the deutocerebral and mandibular neuromeres. (C) Double immunolabeling with anti-Lab (red) and anti- β-Gal (green) shows that reporter gene expression occurrs in the cortical cytoskeleton of the cells that also show nuclear Lab expression (arrows in C) as well as in their axons projecting along the tritocerebral commissure (arrowhead indicates the tritocerebral commissure). (D) Reporter gene expression is seen in the midline crossection of the tritocerebral commissure (arrow).

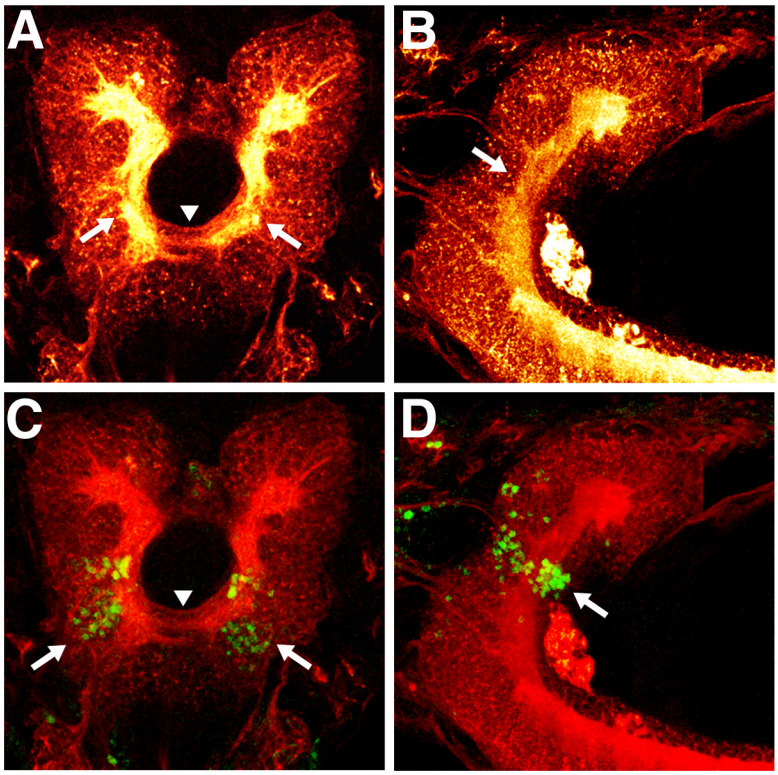



**Figure 3**
Genetic rescue of the *lab* mutant brain phenotype by transgenic expression of the Lab protein in a *lab* null mutant background. Laser confocal microscopy of stage 15 embryos, reconstructions of optical sections. (A, C) frontal views, (B, D) lateral views. (A, B) are from the same preparation; (C, D) are from the same preparation. (A, B) Anti-HRP immunolabeling. Arrows indicate circumesophageal connectives, arrowhead indicates tritocerebral commissure. (C, D) Double immunolabeling with anti-HRP (red) and anti-Lab (green). Arrows indicate targeted misexpression domain of Lab in the *lab* mutant embryonic brain (equivalent to the endogeneous expression domain of Lab in the wildtype embryonic brain). Arrowhead indicates tritocerebral commissure.

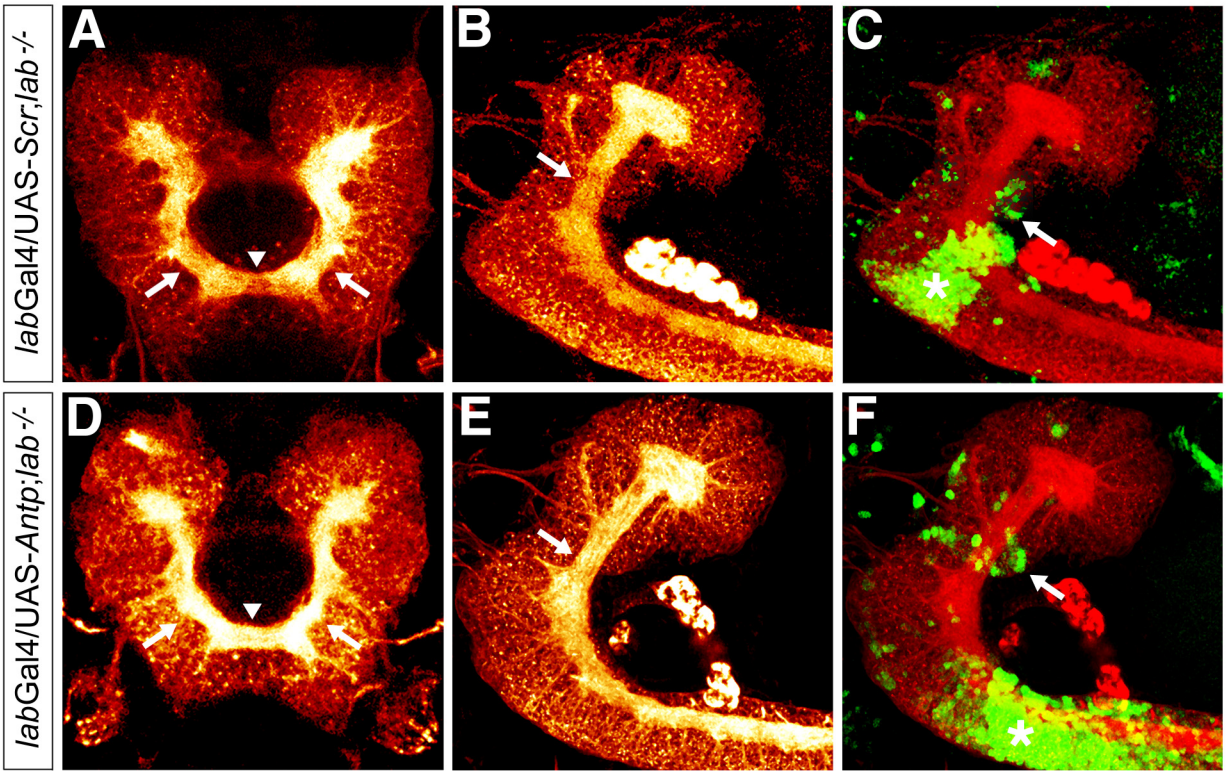



**Figure 4**
Genetic rescue of the *lab* mutant brain phenotype by transgenic expression of the Scr protein (A-C) or the Antpprotein (D-F) in a *lab* null mutant background. Laser confocal microscopy of stage 15 embryos, reconstructions of optical sections. (A, D) frontal views, (B, C, E, F) lateral views. (B, C) are from the same preparation; (E, F) are from the same preparation. (A, B, D, E) Anti-HRP immunolabeling. Arrows indicate circumesophageal connectives, arrowhead indicates tritocerebral commissure. (C) Double immunolabeling with anti-HRP (red) and anti-Scr (green). Arrow indicates targeted misexpression domain of Scr in the *lab* mutant embryonic brain (equivalent to the endogeneous expression domain of Lab in the wildtype embryonic brain). Asterisk labels the endogenous Scr expression domain in the subesophageal ganglion. (F) Double immunolabeling with anti-HRP (red) and anti-Antp (green). Arrow indicates targeted misexpression domain of Antp in the *lab* mutant embryonic brain (equivalent to the endogeneous expression domain of Lab in the wildtype embryonic brain). Asterisk labels the endogenous Antp expression domain in the subesophageal ganglion and ventral nerve cord.

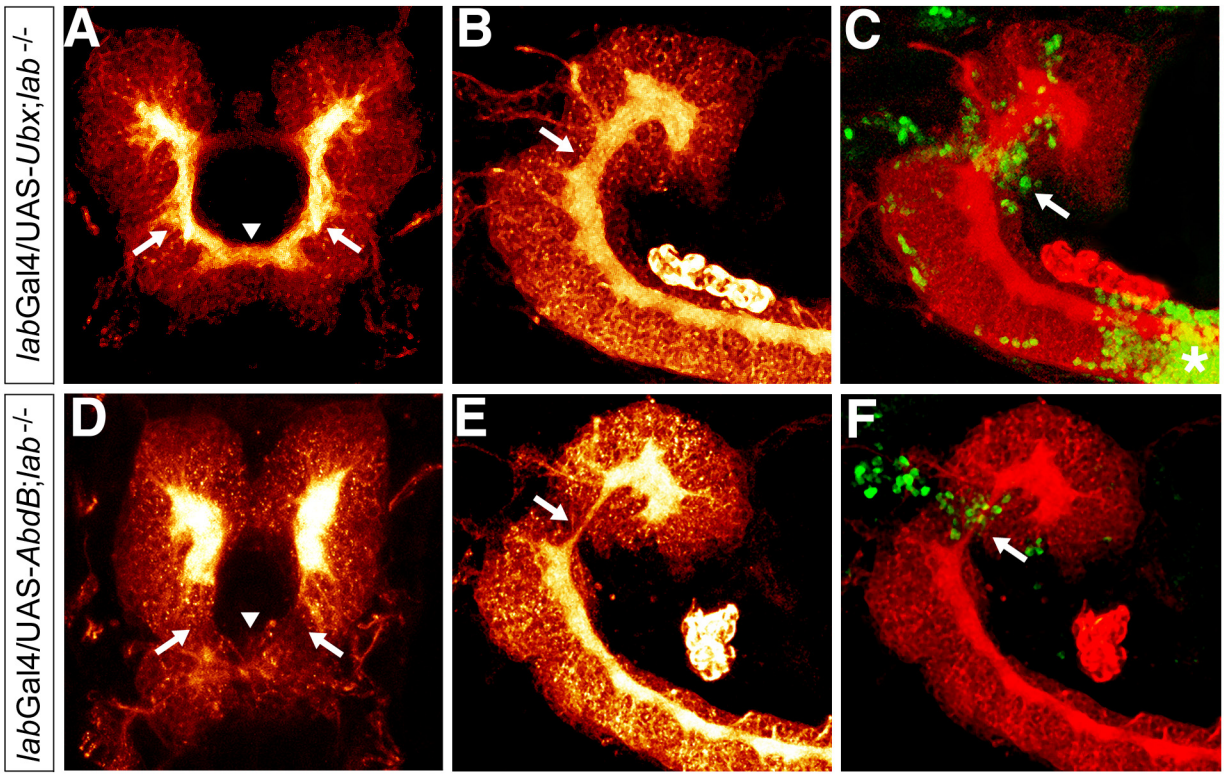



**Figure 5**
Genetic rescue of the *lab* mutant brain phenotype by transgenic expression of the Ubx protein (A-C), and failure of genetic rescue of the *lab* mutant brain phenotype by transgenic expression of the Abd-B protein (D-F) in a *lab* null mutant background. Laser confocal microscopy of stage 15 embryos, reconstructions of optical sections. (A, D) frontal views, (B, C, E, F) lateral views. (B, C) are from the same preparation; (E, F) are from the same preparation. (A, B, D, E) Anti-HRP immunolabeling. Arrows indicate location of circumesophageal connectives, arrowhead indicates location of tritocerebral commissure. (C) Double immunolabeling with anti-HRP (red) and anti-Ubx (green). Arrow indicates targeted misexpression domain of Ubx in the *lab* mutant embryonic brain (equivalent to the endogeneous expression domain of Lab in the wildtype embryonic brain). Asterisk labels part of the endogenous Ubx expression domain in the ventral nerve cord. (D) Double immunolabeling with anti-HRP (red) and anti-Abd-B (green). Arrow indicates targeted misexpression domain of Abd-B in the *lab* mutant embryonic brain (equivalent to the endogeneous expression domain of Lab in the wildtype embryonic brain). The endogenous Abd-B expression domain in the ventral nerve cord is located in posterior neuromeres that are not shown.

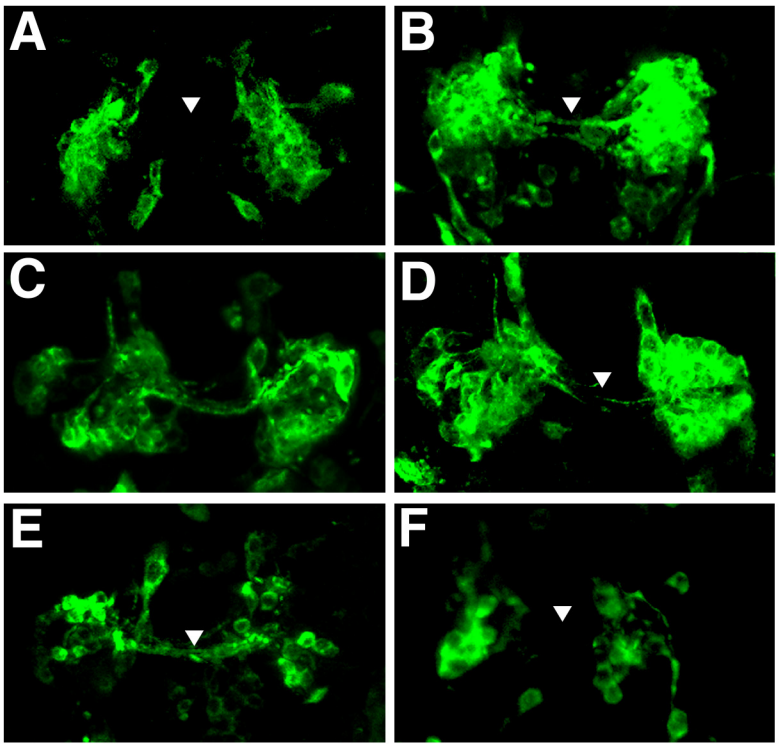



**Figure 6**
Reporter gene expression shows genetic rescue of commissural axonal projections in the *lab-/-* cells of the tritocerebrum by transgenic expression of the Lab, Dfd, Antp, and Ubx proteins (B-E), and failure of rescue by transgenic expression of the Abd-B protein (F); also shown is the absence of commissural axonal projections in the tritocerebral *lab* null mutant domain (A). Laser confocal microscopy of stage 13-15 embryos, reconstructions of optical sections, frontal views. Immunolabeling with anti-β-Gal (green). *UAS::taulacZ* reporter gene expression is seen in the cortical cytoskeleton and axons of cells in tritocerebral lab mutant. Arrowheads indicate presense or absence of commissural axons of the *lab-/-* cells. In (A), visualization of cell bodies and axonal projections was by *lab*::Gal4 driven *UAS::taulacZ* reporter gene expression in the tritocerebral *lab* mutant domain. In (B-F), visualization of cell bodies and and genetic rescue of axonal projections of the lab-/- cells was through coexpression of *UAS::taulacZ* reporter with UAS::*Hox* responders in the tritocerebral lab mutant domain by the lab::Gal4 driver. For all of the Hox gene products except AbdB these experiments demonstrated that the rescued tritocerebral *lab-/-* cells were able to extend axons that projected correctly along the rescued tritocerebral commissure.

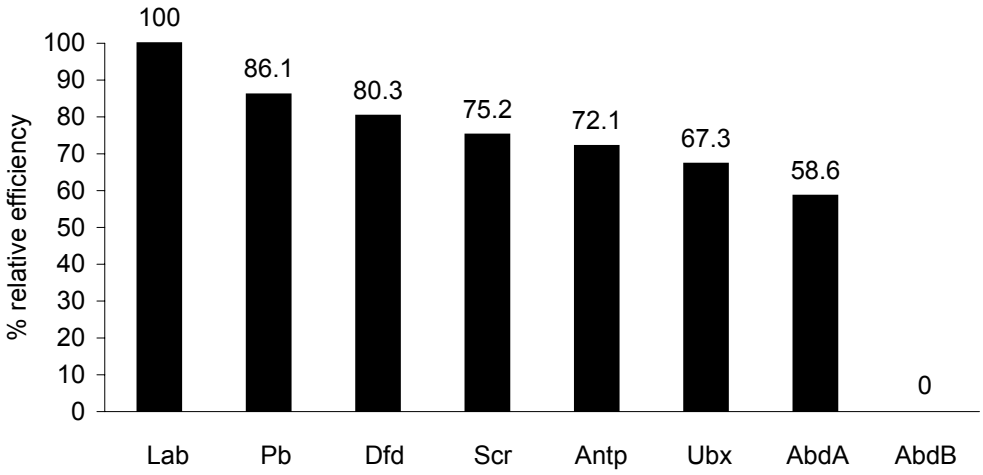



**Figure 7**

Relative rescue efficiency of Hox gene products as related to Lab. The relative efficiency of rescue of the tritocerebral brain defects in *lab* null mutants is shown for the Hox gene products Lab, Pb, Dfd, Scr, Antp, Ubx, Abd-A, and Abd-B expressed in the lab mutant under the control of the same *lab*-specific cis-acting regulatory elements (see Table 1). The rescue efficiency for Lab is taken as 100% and the rescue values (the relative percentage of embryos showing a complete rescue of the tritocerebral brain defects) of the other Hox gene products are shown in percentage relative to this. The relative rescue efficiency of the Hox gene products (Lab>Pb>Dfd>Scr>Antp>Ubx>Abd-A) reflects the proximal to distal arrangement of their encoding loci on the chromosome.

# 6. 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], Michael Primig[3], Dario Acampora[4], Antonio Simeone[4] and Heinrich Reichert.

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

**[1]Roche Bioinformatics and [2]Roche Genetics Pharmaceuticals Division, F.Hoffmann-La Roche, Ltd., CH-4070 Basel, Switzerland**

[3]Genomics-Proteomics, Biocenter, Klingelbergstrasse 70 University of Basel, CH-4056 Basel, Switzerland

**[4]MRC Center for Developmental Neurobiology, King's College London, Guy's Campus, New Hunt's House, London SE1 1UL, UK**

Corresponding Author: Heinrich Reichert, Institute of Zoology, Biocenter/Pharmacenter, Klingelbergstrasse 50, University of Basel, CH-4056 Basel, Switzerland. Tel.: +41-61-267-1612; Fax; +41-61-267-1613; email: heinrich.reichert@unibas.ch

## Abstract

**Background:** The homeobox genes of the *orthodenticle (otd)/Otx* family play conserved roles in embryogenesis of the head and brain. 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 restore defects in cephalic and brain development. This suggests 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.

**Results:** 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 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 identified showed differential expression changes following overexpression of either *otd* or *Otx2,* and for 90 of these, comparable changes were observed under both experimental conditions. We postulate that these transcripts are common downstream targets of the fly *otd* gene and the human *Otx2* gene in *Drosophila.*

**Conclusions:** Our experiments indicate that approximately one third of the *otd*-regulated transcripts also respond to overexpression of the human *Otx2* gene in *Drosophila*. 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*.

## Background

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[1-5]. 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[6-9]. 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[6, 10-15]. 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[6, 14, 16]. 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[17]. 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[18-21]. *Otx1* null mice are viable but show spontaneous epileptic seizures and abnormalities affecting the dorsal telencephalic cortex [22].

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 [3, 23-25]. 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 partially rescue the brain and cephalic defects in *Drosophila* although *Otx2* rescues at a lower frequency than *otd*, and *Otx1* rescues less efficiently still[24, 25]. 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[23, 26].

*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[25, 27]. 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[27]. 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[27, 28].

In order to address this issue at a genome-wide level we have combined cross-phylum overexpression experiments with expression profiling using oligonucleotide arrays. We sought to identify the common downstream target genes of fly *otd* and human *Otx2* in *Drosophila*. To this end, we used transgenic flies which carried either the fly *otd* gene or the human *Otx2* gene under the control of a heat-inducible promoter [29-33]. These experiments identified 287 annotated genes that showed highly significant ($p \leq 0.001$) 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. In summary, approximately one third of the candidate *otd* downstream target genes in *Drosophila* also respond to overexpression of the human *Otx2* gene homolog and nearly all of them display identical patterns of either up- or downregulation under both experimental conditions. 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 similarity 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[24]. 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 (see Material and Method). 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 (data not shown). Wildtype control flies were subjected to the identical heat shock conditions.

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 wildtype 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[34]. For each experimental condition, several replicates were carried out (see Materials and Methods). The degree of reproducibility within individual 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 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 $p \leq 0.001$, a total of 287 genes were found to be differentially regulated following *otd* overexpression, as compared to heat shocked wildtype control embryos. This corresponds to 2.1% of the genes represented on the array. At a significance level of $p \leq 0.01$, a total of 762 genes were found to be differentially regulated following *otd* overexpression as compared to heat shocked wildtype 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. At a significance level of $p \leq 0.001$, a total of 682 genes were found to be differentially expressed following *Otx2* overexpression, as compared to heat shocked wildtype control embryos. This corresponds to 5.1% of the genes represented on the array. At a significance level of $p \leq 0.01$, 1395 genes were found to be differentially expressed following *Otx2* overexpression as compared to heat shocked wildtype 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 $p \leq 0.001$, 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 $p \leq 0.01$, 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 $p \leq 0.001$ 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; [35]) 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 2 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 *forkhead domain 96cb (fd96Cb)* which encodes a nuclear binding protein, for *patched (ptc)* which encodes a protein involved in signal transduction, for *picot*, which encodes a transporter, and for *cortactin* and *Regulator of cyclin A1 (Rca1)*, which encode gene products of currently unknown molecular function. Only two transcripts showed increases in the 5-10-fold range, namely *sugar transporter1 (sut1)* encoding a protein involved in sugar transportation, and *scraps (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 *eyegone (eyg)*, encoding a transcription factor known to be involved in eye development.

Figure 3 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, *retained (retn), SMC2, licorne (lic), Rtc1, Hairless (H)* and *deadhea (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* also respond to overexpression of the human *Otx2* gene homolog. Figure 4 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 to *otd* and to *Otx2*. Two marked exceptions were *pimple (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* are controlled in a comparable manner by the human *Otx2* gene homolog.

There are a number of interesting genes among these common candidate genes. The four known transcripts in class 'ligand binding or carrier', *scra, Kinesin-like protein at 61F (Klp61F), alpha-Spectrin (alpha-spec)* and *Centrosomal protein 190kD (Cen190)*, are all involved in actin or microtubule binding or movement[36-39]. This finding is intriguing since one of the *Otx2* downstream genes identified in the mouse is a tropomyosin gene, which also encodes actin binding protein[40]. Among the four known transcripts in the class 'nucleic acid binding' are the genes *Minichromosome maintenance 7 (Mcm7)* and *Suppressor of variegation 205 (Su(var)205)* [41, 42] which encode chromatin binding proteins and the genes *eyg* and *HLH54F* which encode transcription factors[43, 44]. The four known transcripts in the functional class 'enzymes' are *Lysozyme D (LysD)*, *cdc2*, *Rpd3*, and *BcDNA:LD08534[45-48]*. Although the *cdc2* gene product is classified as 'enzyme', it also acts at the G2/M transition of mitotic cell cycle[47]. Moreover, *Rpd3* encodes a histone deacetylase which is involved in the chromatin structure[46]. In the class "transporter" the SNAP receptor encoding *n-synaptobrevin (n-syb)* gene is involved in synaptic vesicle docking and fusion and is expressed in the embryonic CNS[49]. In the class 'signal transducer', the gene *EG: 30B8.6* encodes a putative GABA-B receptor[50]. Finally, the gene *Segregation distorter (Sd)* classified as 'enzyme regulator' encodes a RAN GTPase activator[51]. Among the transcripts of known genes are several genes, whose precise functional role is not well defined at the molecular level. These are the *Bx34* and *MRG15* genes[52, 53] 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 [54]. The gene product of *Bub3* is localized to the kinetochore and may function in the mitotic check point[55]. *pim* is expressed in the embryonic CNS and encodes a protein implicated in mitotic sister chromatid separation[56].

**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 wildtype 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 *wunen (wun)* and *Scc1*, whose expression levels remained unchanged in response to *otd* or *Otx2* overexpression, were used 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 oligonucleotide 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[23-26]. 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 currently 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 appear to 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 ($p \leq 0.001$) 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 ($p \leq 0.01$) 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.

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 amino acid sequence between the OTD and OTX2 proteins, this may not be altogether surprising. The OTD and OTX2 proteins consist of 548 and 289 amino acids, respectively. Shared homology between them is restricted to the homeodomain and to a short domain immediately upstream of the homeodomain as well as a tripeptide at the amino terminus[25].

Moreover, since *Otx* genes cannot completely replace the *otd* gene in cross-phylum rescue experiments *in vivo*, a complete correspondence of *otd* downstream genes and common *otd*/*Otx* downstream genes might not be expected [3, 24, 25]. However, more than one third of the *otd*-regulated genes do respond also to *Otx2* overexpression. We suggest that these common downstream genes are likely to explain the overlapping roles of the *otd*/*Otx* genes in cross-phylum rescue experiments *in vivo*. These target genes reflect the evolutionarily conserved roles of the members of the *otd*/*Otx* gene family in *Drosophila*. To investigate this further, it will now be important and interesting to carry out similar functional genomic analyses of *otd* and *Otx* gene action in a mammalian system such as the mouse [27].

### *otd* overexpression: a genomic perspective of candidate downstream genes

The experiments reported here identify approximately 300 genes that showed highly significant ($p \leq 0.001$) 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 higher 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 the genes in the genome [30]. Indeed, in similar experiments in which the homeobox gene *labial* was overexpressed using the same heat shock protocol as described here, 6.4% of the genes represented on the array used were shown to be differentially regulated at a significance level of $p<0.01$ [30]. (It should however, be noted that the array used in these *labial* overexpression experiments represents only 10% of the genes in the fly genome.) Thus the relative number of putative *otd* targets appears to be 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 [57]. 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*) [58]. 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 minutes post heat shock. Assuming that *otd* expression follows a similar kinetic profile as *ftz*, it is likely that we have identified primary targets as well as genes whose response was caused by indirect effects requiring intermediate transcription.

Third, these results show that the primary consequence of *otd* overexpression in *Drosophila* is the upregulation of its 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 an approximately equal number of upregulated and downregulated target genes [30].

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 an *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 [59]. 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 identified. 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 concentration and stability. As different levels of a homeoprotein may 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 [60, 61]. 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[62].

**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[24, 25, 63-66]. 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 reported here has uncovered remarkable similarities in the activity of the human transgene in the fly as compared to that of its fly homolog. Thus, *otd* and *Otx2* both upregulate most of their target genes upon overexpression. Moreover, the target genes of both transcription factors fall into the same functional categories. 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 ($p \leq 0.001$) 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 Materials and Methods). Nevertheless, these data should be interpreted with caution, since several explanations, not mutually exclusive, are possible for the observation that more genes respond to overexpression of *Otx2*. First, only one single transgenic strain of *otd* and only one single transgenic strain of *Otx2* were used. Thus, strain differences or insertion effects might account for the fact that more genes show differential expression following overexpression of *Otx2* compared to overexpression of *otd.* Second, it is conceivable that overexpression of the *Otx2* gene affects 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. Similarly, the OTX2 protein might be more promiscuous than OTD in its interactions with the numerous cofactors that determine target specificity. Third, it has been shown that the DNA binding specificity of homeoproteins is low *in vitro*. But given that the homeodomain is conserved and *Otx2* rescues the *otd* phenotype, this suggests that they should recruit a similar subset of cofactors and regulate a common subset of downstream genes at least in those tissues where *otd* is endogenously expressed. Furthermore, the *Otx2* product, which is not a fly protein, 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 wildtype was *Drosophila melanogaster* Oregon-R. For overexpression of *otd*, we used the *hsp-otd* line 5A generated by Royet and Finkelstein [67]. For overexpression of human *Otx2*, we used the *hsp-Otx2* line generated by Leuzinger *et al* [24]. 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[29]. Embryos younger than embryonic stage 10 were not used, since heat shock in these earlier stages results in lethality[68]. 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 procedure [69]. 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 1 [34] 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 [32]. RNA was isolated, labeled and hybridized to the arrays as described [29, 30] 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, Redwood City, CA). For quantification of relative transcript abundance, the Average Difference value (Avg Diff) was used[32]. 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: hs*wt* vs. *wt*, hs*wt* vs. hs*otd* and hs*wt* vs. hs*Otx2*. For differential transcript imaging, only transcripts that had highly significant or significant changes in Avg Diff ($p \leq 0.001$ and $p \leq 0.01$, respectively) 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 policy).

**Reverse Transcription Polymerase Chain Reaction**

300 ng poly(A)$^+$ RNA was isolated from embryos of *wt*, hs*wt*, hs*otd* and hs*Otx2* (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 $1.5 \times 10^{-6}$ µg/µl and $3.6 \times 10^{-7}$ 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 also thank Reinhold Koch especially for contributions to standardize the description of the statistics. 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.

## References


1. Finkelstein R, Boncinelli E: **From fly head to mammalian forebrain: the story of otd and Otx**. *Trends Genet* 1994; **10**:310-315.
2. Thor S: **The genetics of brain development: conserved programs in flies and mice**. *Neuron* 1995; **15**:975-977.
3. Sharman AC, Brand M: **Evolution and homology of the nervous system: cross-phylum rescues of otd/Otx genes**. *Trends Genet* 1998; **14**:211-214.
4. Holland LZ, Holland ND: **Chordate origins of the vertebrate central nervous system**. *Curr Opin Neurobiol* 1999; **9**:596-602.
5. Galliot B, Miller D: **Origin of anterior patterning. How old is our head?** *Trends Genet* 2000; **16**:1-5.
6. Finkelstein R, Smouse D, Capaci TM, Spradling AC, Perrimon N: **The orthodenticle gene encodes a novel homeo domain protein involved in the development of the Drosophila nervous system and ocellar visual structures**. *Genes Dev* 1990; **4**:1516-1527.
7. Simeone A, Gulisano M, Acampora D, Stornaiuolo A, Rambaldi M, Boncinelli E: **Two vertebrate homeobox genes related to the Drosophila empty spiracles gene are expressed in the embryonic cerebral cortex**. *Embo J* 1992; **11**:2541-2550.
8. Simeone A, Acampora D, Mallamaci A, Stornaiuolo A, D'Apice MR, Nigro V, Boncinelli E: **A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo**. *Embo J* 1993; **12**:2735-2747.
9. Acampora D, Simeone A: **The TINS Lecture. Understanding the roles of Otx1 and Otx2 in the control of brain morphogenesis**. *Trends Neurosci* 1999; **22**:116-122.
10. Gao Q, Wang Y, Finkelstein R: **Orthodenticle regulation during embryonic head development in Drosophila**. *Mech Dev* 1996; **56**:3-15.

11. Cohen SM, Jurgens G: **Mediation of Drosophila head development by gap-like segmentation genes**. *Nature* 1990; **346**:482-485.

12. Grossniklaus U, Cadigan KM, Gehring WJ: **Three maternal coordinate systems cooperate in the patterning of the Drosophila head**. *Development* 1994; **120**:3155-3171.

13. Schmidt-Ott U, Gonzalez-Gaitan M, Jackle H, Technau GM: **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* 1994; **91**:8363-8367.

14. Wieschaus E, Perrimon N, Finkelstein R: **orthodenticle activity is required for the development of medial structures in the larval and adult epidermis of Drosophila**. *Development* 1992; **115**:801-811.

15. Finkelstein R, Perrimon N: **The orthodenticle gene is regulated by bicoid and torso and specifies Drosophila head development**. *Nature* 1990; **346**:485-488.

16. Klambt C, Jacobs JR, Goodman CS: **The midline of the Drosophila central nervous system: a model for the genetic analysis of cell fate, cell migration, and growth cone guidance**. *Cell* 1991; **64**:801-815.

17. Simeone A, Acampora D, Gulisano M, Stornaiuolo A, Boncinelli E: **Nested expression domains of four homeobox genes in developing rostral brain**. *Nature* 1992; **358**:687-690.

18. Suda Y, Matsuo I, Kuratani S, Aizawa S: **Otx1 function overlaps with Otx2 in development of mouse forebrain and midbrain**. *Genes Cells* 1996; **1**:1031-1044.

19. Matsuo I, Kuratani S, Kimura C, Takeda N, Aizawa S: **Mouse Otx2 functions in the formation and patterning of rostral head**. *Genes Dev* 1995; **9**:2646-2658.

20. Ang SL, Jin O, Rhinn M, Daigle N, Stevenson L, Rossant J: **A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain**. *Development* 1996; **122**:243-252.

21. Acampora D, Mazan S, Lallemand Y, Avantaggiato V, Maury M, Simeone A, Brulet P: **Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation**. *Development* 1995; **121**:3279-3290.

22. Acampora D, Mazan S, Avantaggiato V, Barone P, Tuorto F, Lallemand Y, Brulet P, Simeone A: **Epilepsy and brain abnormalities in mice lacking the Otx1 gene**. *Nat Genet* 1996; **14**:218-222.

23. Acampora D, Avantaggiato V, Tuorto F, Barone P, Reichert H, Finkelstein R, Simeone A: **Murine Otx1 and Drosophila otd genes share conserved genetic functions required in invertebrate and vertebrate brain development**. *Development* 1998; **125**:1691-1702.

24. Leuzinger S, Hirth F, Gerlich D, Acampora D, Simeone A, Gehring WJ, Finkelstein R, Furukubo-Tokunaga K, Reichert H: **Equivalence of the fly orthodenticle gene and the human OTX genes in embryonic brain development of Drosophila**. *Development* 1998; **125**:1703-1710.

25. Nagao T, Leuzinger S, Acampora D, Simeone A, Finkelstein R, Reichert H, Furukubo-Tokunaga K: **Developmental rescue of Drosophila cephalic defects by the human Otx genes**. *Proc Natl Acad Sci U S A* 1998; **95**:3737-3742.

26. Boyl PP, Signore M, Acampora D, Martinez-Barbera JP, Ilengo C, Annino A, Corte G, Simeone A: **Forebrain and midbrain development requires epiblast-restricted Otx2 translational control mediated by its 3' UTR**. *Development* 2001; **128**:2989-3000.

27. Reichert H, Simeone A: **Developmental genetic evidence for a monophyletic origin of the bilaterian brain**. *Philos Trans R Soc Lond B Biol Sci* 2001; **356**:1533-1544.

28. Boncinelli E, Morgan R: **Downstream of Otx2, or how to get a head**. *Trends Genet* 2001; **17**:633-636.

29. Leemans R, Egger B, Loop T, Kammermeier L, He H, Hartmann B, Certa U, Hirth F, Reichert H: **Quantitative transcript imaging in normal and heat-shocked Drosophila embryos by using high-density oligonucleotide arrays**. *Proc Natl Acad Sci U S A* 2000; **97**:12138-12143.

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

31. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring by hybridization to high-density oligonucleotide arrays**. *Nat Biotechnol* 1996; **14**:1675-1680.

32. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays**. *Nat Genet* 1999; **21**:20-24.

33. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann Wea: **Comparative genomics of the eukaryotes**. *Science* 2000; **287**:2204-2215.

34. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, al. e: **The genome sequence of Drosophila melanogaster**. *Science* 2000; **287**:2185-2195.

35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000; **25**:25-29.

36. Barbosa V, Yamamoto RR, Henderson DS, Glover DM: **Mutation of a Drosophila gamma tubulin ring complex subunit encoded by discs degenerate-4 differentially disrupts centrosomal protein localization**. *Genes Dev* 2000; **14**:3126-3139.

37. Lee JK, Coyne RS, Dubreuil RR, Goldstein LS, Branton D: **Cell shape and interaction defects in alpha-spectrin mutants of Drosophila melanogaster**. *J Cell Biol* 1993; **123**:1797-1809.

38. Heck MM, Pereira A, Pesavento P, Yannoni Y, Spradling AC, Goldstein LS: **The kinesin-like protein KLP61F is essential for mitosis in Drosophila**. *J Cell Biol* 1993; **123**:665-679.

39. Field CM, Alberts BM: **Anillin, a contractile ring protein that cycles from the nucleus to the cell cortex**. *J Cell Biol* 1995; **131**:165-178.

40. Zakin L, Reversade B, Virlon B, Rusniok C, Glaser P, Elalouf JM, Brulet P: **Gene expression profiles in normal and Otx2-/- early gastrulating mouse embryos**. *Proc Natl Acad Sci U S A* 2000; **97**:14388-14393.

41. Wakimoto BT: **Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila**. *Cell* 1998; **93**:321-324.

42. Feger G: **Identification and complete cDNA sequence of the missing Drosophila MCMs: DmMCM3, DmMCM6 and DmMCM7**. *Gene* 1999; **227**:149-155.

43. Treisman JE: **A conserved blueprint for the eye?** *Bioessays* 1999; **21**:843-850.

44. Ledent V, Vervoort M: **The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis**. *Genome Res* 2001; **11**:754-770.

45. Regel R, Matioli SR, Terra WR: **Molecular adaptation of Drosophila melanogaster lysozymes to a digestive function**. *Insect Biochem Mol Biol* 1998; **28**:309-319.

46. Wallrath LL: **Unfolding the mysteries of heterochromatin**. *Curr Opin Genet Dev* 1998; **8**:147-153.

47. Su TT, O'Farrell PH: **Size control: cell proliferation does not equal growth**. *Curr Biol* 1998; **8**:R687-689.

48. **The FlyBase database of the Drosophila Genome Projects and community literature. The FlyBase Consortium**. *Nucleic Acids Res* 1999; **27**:85-88.

49. Saitoe M, Schwarz TL, Umbach JA, Gundersen CB, Kidokoro Y: **Absence of junctional glutamate receptor clusters in Drosophila mutants lacking spontaneous transmitter release**. *Science* 2001; **293**:514-517.

50. Benos PV, Gatt MK, Murphy L, Harris D, Barrell B, Ferraz C, Vidal S, Brun C, Demaille J, Cadieu E, *et al*: **From first base: the sequence of the tip of the X chromosome of Drosophila melanogaster, a comparison of two sequencing strategies**. *Genome Res* 2001; **11**:710-730.

51. Crow JF: **Unmasking a cheating gene**. *Science* 1999; **283**:1651-1652.

52. Zimowska G, Aris JP, Paddy MR: **A Drosophila Tpr protein homolog is localized both in the extrachromosomal channel network and to nuclear pore complexes**. *J Cell Sci* 1997; **110**:927-944.

53. Bertram MJ, Pereira-Smith OM: **Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif**. *Gene* 2001; **266**:111-121.

54. Prokopenko SN, He Y, Lu Y, Bellen HJ: **Mutations affecting the development of the peripheral nervous system in Drosophila: a molecular screen for novel proteins**. *Genetics* 2000; **156**:1691-1715.

55. Dobie KW, Hari KL, Maggert KA, Karpen GH: **Centromere proteins and chromosome inheritance: a complex affair**. *Curr Opin Genet Dev* 1999; **9**:206-217.

56. Philip AV: **Mitotic sister-chromatid separation: what Drosophila mutants can tell us**. *Trends Cell Biol* 1998; **8**:150.

57. Kablar B, Vignali R, Menotti L, Pannese M, Andreazzoli M, Polo C, Giribaldi MG, Boncinelli E, Barsacchi G: **Xotx genes in the developing brain of Xenopus laevis**. *Mech Dev* 1996; **55**:145-158.

58. Nasiadka A, Krause HM: **Kinetic analysis of segmentation gene interactions in Drosophila embryos**. *Development* 1999; **126**:1515-1526.

59. Bate M, Martinez-Arias A: **The Development of *Drosophila Melanogaster***. *Cold Spring Harbor, Cold Spring Harbor Laboratory Press* 1993.

60. Hoppler S, Bienz M: **Specification of a single cell type by a Drosophila homeotic gene**. *Cell* 1994; **76**:689-702.

61. Cribbs DL, Benassayag C, Randazzo FM, Kaufman TC: **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* 1995; **14**:767-778.

62. Mann RS, Affolter M: **Hox proteins meet more partners**. *Curr Opin Genet Dev* 1998; **8**:423-429.

63. Hartmann B, Hirth F, Walldorf U, Reichert H: **Expression, regulation and function of the homeobox gene empty spiracles in brain and ventral nerve cord development of Drosophila**. *Mech Dev* 2000; **90**:143-153.

64. Yang X, Li DM, Chen W, Xu T: **Human homologue of Drosophila lats, LATS1, negatively regulate growth by inducing G(2)/M arrest or apoptosis**. *Oncogene* 2001; **20**:6516-6523.

65. Gunawardena S, Goldstein LS: **Disruption of axonal transport and neuronal viability by amyloid precursor protein mutations in Drosophila**. *Neuron* 2001; **32**:389-401.

66. Parkes TL, Elia AJ, Dickinson D, Hilliker AJ, Phillips JP, Boulianne GL: **Extension of Drosophila lifespan by overexpression of human SOD1 in motorneurons**. *Nat Genet* 1998; **19**:171-174.

67. Royet J, Finkelstein R: **Pattern formation in Drosophila head development: the role of the orthodenticle homeobox gene**. *Development* 1995; **121**:3561-3572.

68. Walter MF, Petersen NS, Biessmann H: **Heat shock causes the collapse of the intermediate filament cytoskeleton in Drosophila embryos**. *Dev Genet* 1990; **11**:270-279.

**69. Tautz D, Pfeifle C:** 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* 1989**; 98:**81-85**.

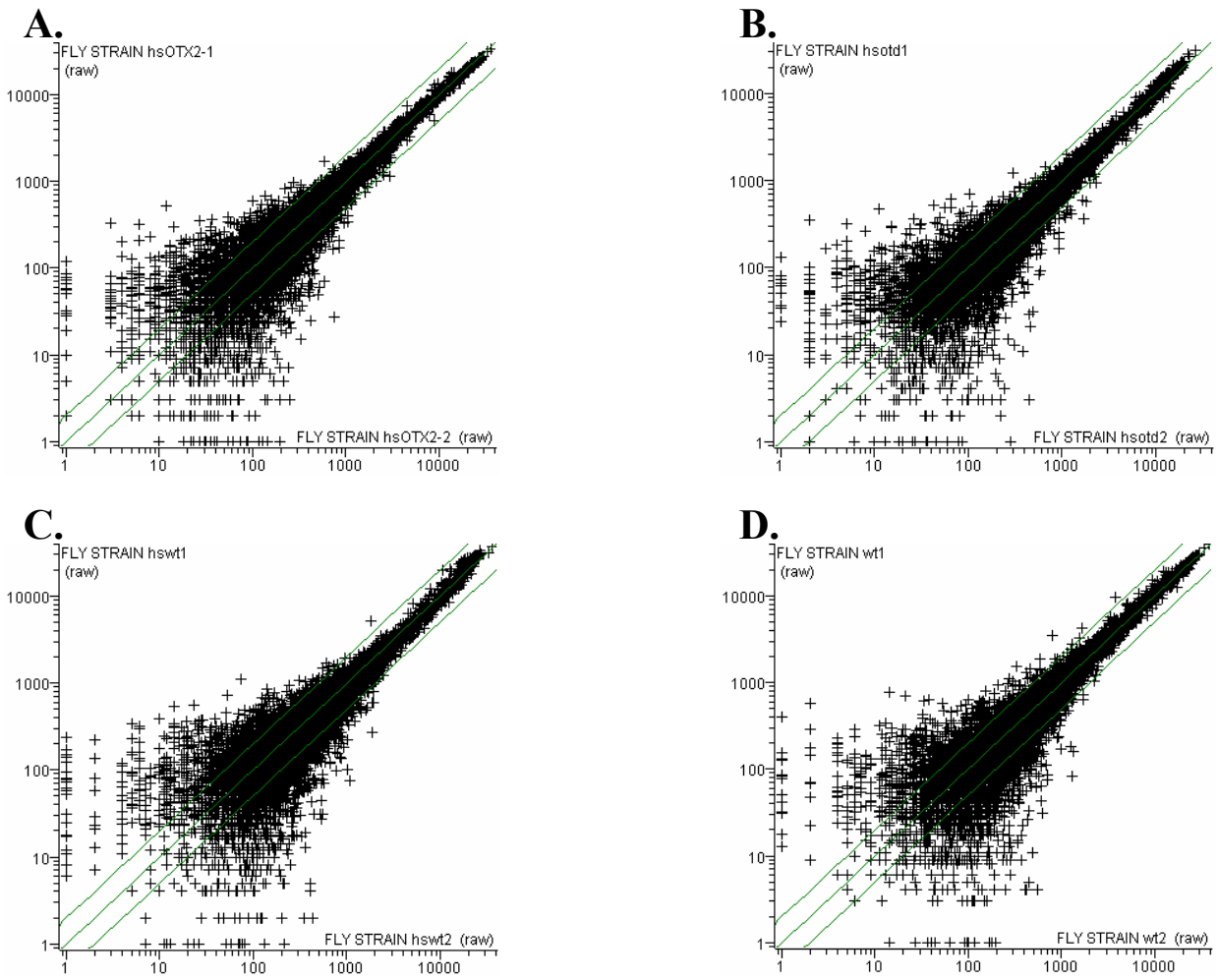



**Fig. 1**. Normalized Average Difference (Avg Diff) of one pair of replicate arrays for each experimental condition in a log(base 10) scale. A. hs*Otx2*; B. hs*otd*; C. hswt; D. wt. Only probe sets with positive values in both arrays are used. The central line is y = x, and the flanking lines indicate the difference of a factor of two.

**Fig.2 Known transcripts differentially expressed in response to overexpression of *otd***

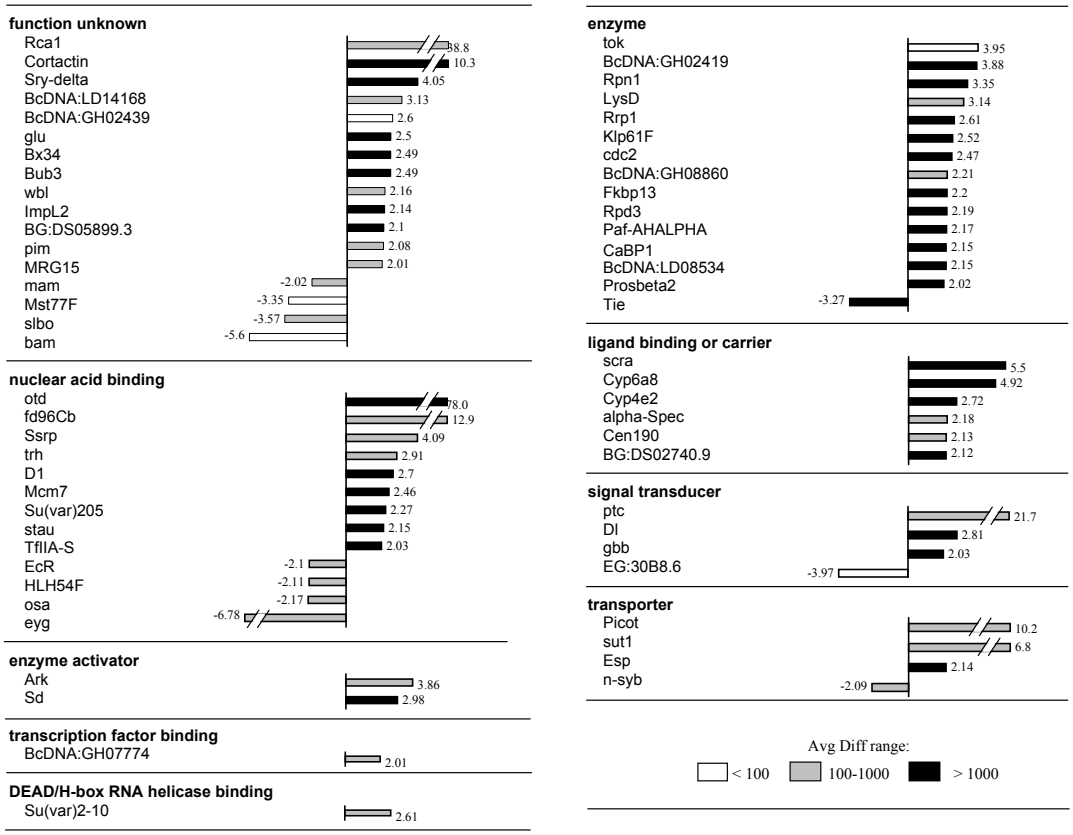



**Fig. 2.** Known transcripts differentially expressed in response to overexpression of *otd* , grouped according to functional classes. Bars represent the fold change between differentially expressed transcripts in hswt embryos and hs*otd* embryos. Positive values indicate that the relative expression 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***

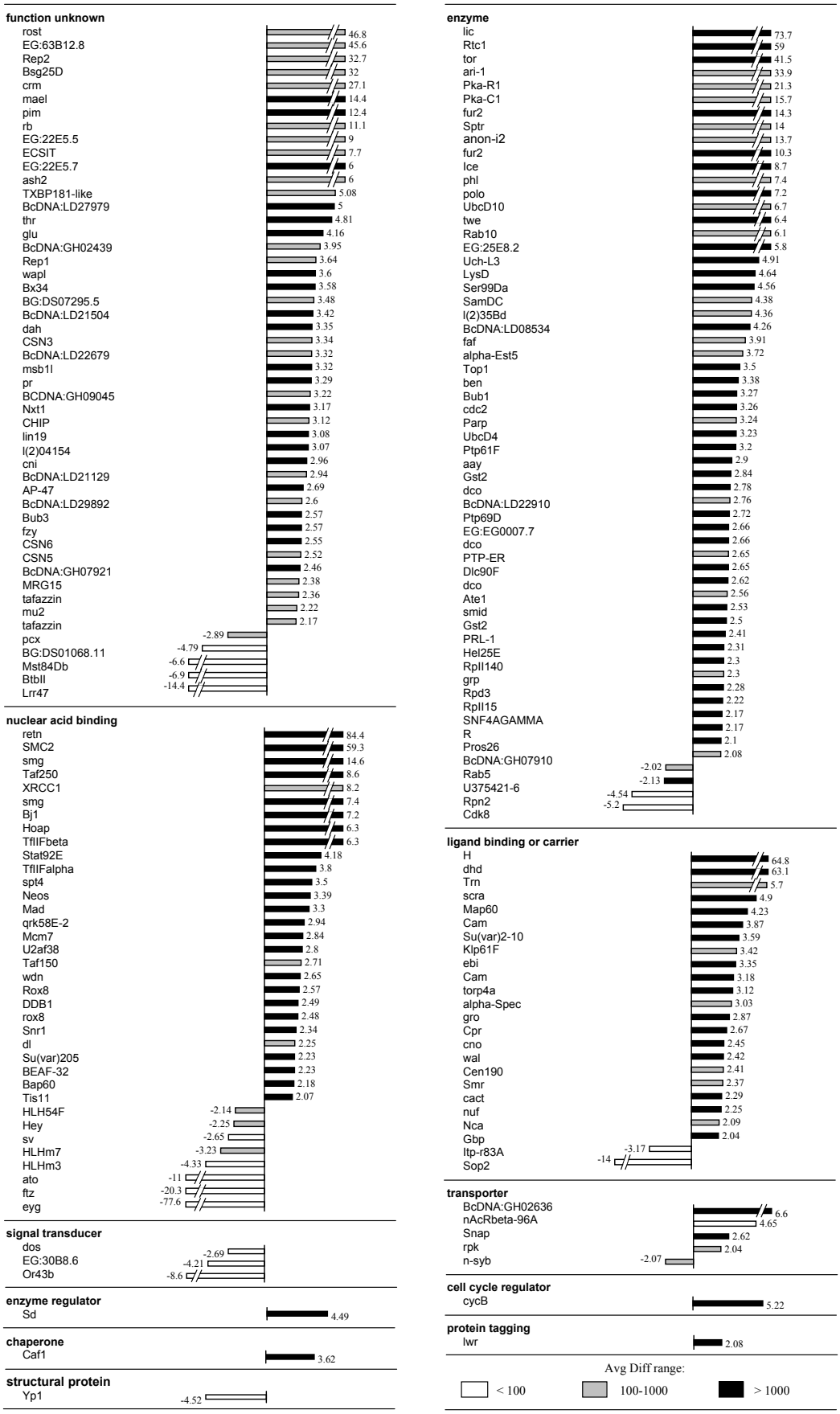



**Fig. 3.** Known transcripts differentially expressed in response to overexpression of *Otx2*, grouped according to functional classes. Bars represent the fold change between differentially expressed transcripts in hswt embryos and hs*Otx2* embryos. Positive values indicate that the relative expression 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.

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

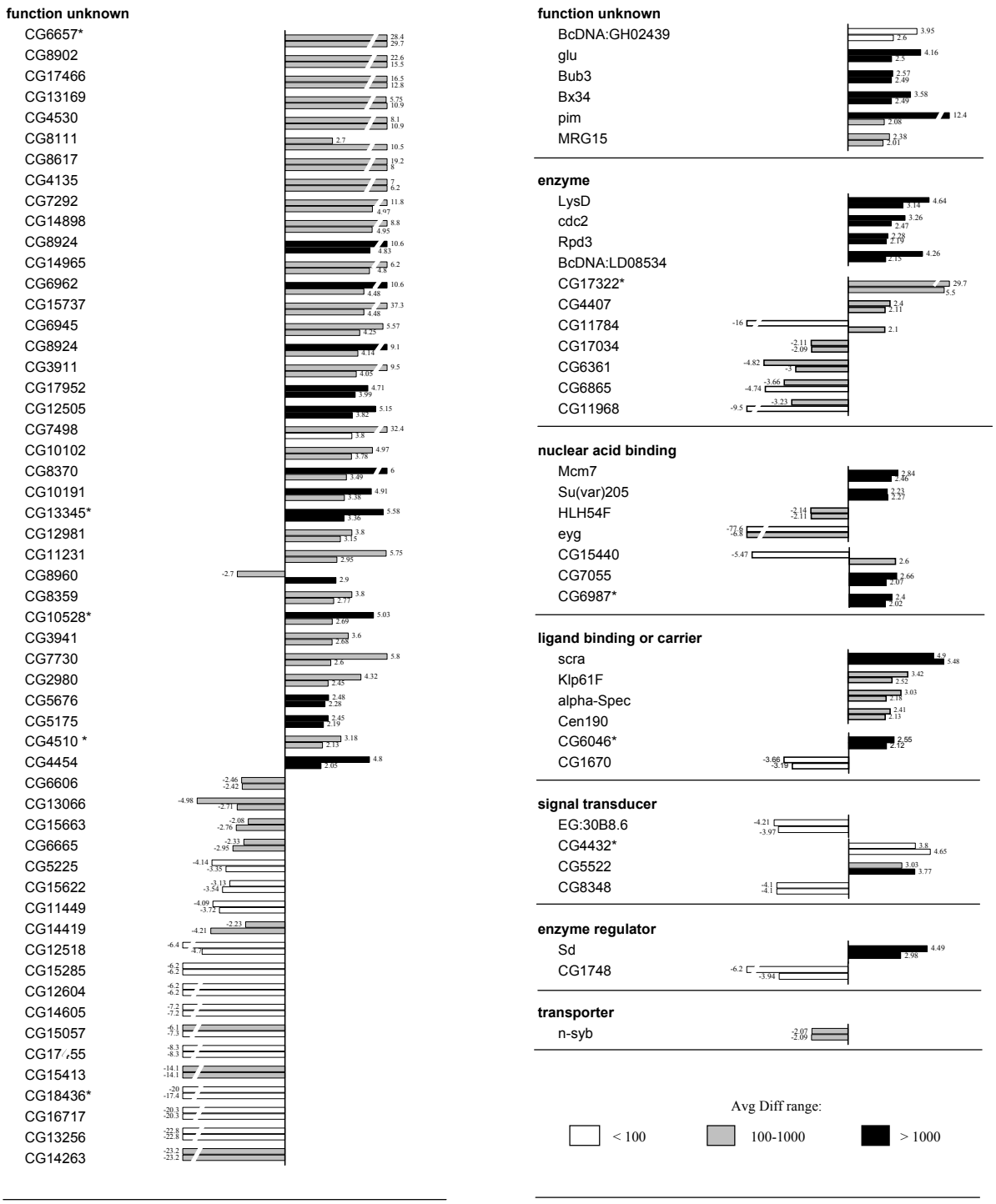



**Fig. 4.** Transcripts differentially expressed in response to overexpression of *otd* and in response to overexpression of human *Otx2*, grouped according to functional classes. Bars represent the fold change between differentially expressed transcripts in hswt embryos and hs*otd* or hs*Otx2* embryos. 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.

**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 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 $p \leq 0.001$. B. Number of transcripts that were differentially expressed at the significance level of $p \leq 0.01$.

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

| Functional class | $n^{otd}$ | $n^{otd}$/N (%) | $n^{otd}$/M (%) | $n^{Otx2}$ | $n^{Otx2}$/N (%) | $n^{Otx2}$/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 |
| 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 |

**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 ($p \leq 0.001$); 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.

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

| | 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 |

**Table 3.** Comparison of change folds between oligonucleotide arrays and RT-PCR RT-PCR was performed on cDNA derived from hswt, 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. wun is used as a control for the comparison of the RT-PCR data between hswt an hs*Otx2*. Scc1 is used as a control for the comparison of the RT-PCR data between hswt an hs*otd*.

# 7. Gliogenesis in *Drosophila*: Genome-Wide Analysis of Downstream Genes of *glial cells missing* in the Embryonic Nervous System

Boris Egger[1], Ronny Leemans[1], Thomas Loop[1], Lars Kammermeier[1], Yun Fan[1], Tanja Radimerski[1], Martin C. Strahm[2], Ulrich Certa[3], and Heinrich Reichert[1,4]

[1]Biozentrum/Pharmazentrum, University of Basel, CH-4056 Basel, Switzerland.
[2]Roche Bioinformatics and [3]Genetics Pharmaceuticals Division, F. Hoffmann-La Roche, Ltd, CH-4070 Basel, Switzerland.

[4]Corresponding author:
Heinrich Reichert, Institute of Zoology, Biozentrum/Pharmazentrum
University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland.
Tel.: 0041-61-267 16 12 Fax: 0041-61-267 16 13 e-mail: Heinrich.Reichert@unibas.ch

## Summary

**In *Drosophila*, the *glial cells missing (gcm)* gene encodes a transcription factor that controls the determination of glial versus neuronal fate. In *gcm* mutants, presumptive glial cells are transformed into neurons and, conversely, when *gcm* is ectopically misexpressed, presumptive neurons become glia. Although *gcm* is thought to initiate glial cell development through its action on downstream genes which execute the glial differentiation program, little is known about the identity of these genes. To identify *gcm* downstream genes in a comprehensive manner, we used genome-wide oligonucleotide arrays to analyse differential gene expression in wild-type embryos versus embryos in which *gcm* is misexpressed throughout the neuroectoderm. Transcripts were analysed at two defined temporal windows during embryogenesis. During a first period of initial *gcm* action on determination of glial cell precursors, over 400 genes were differentially regulated. Among these are numerous genes that encode other transcription factors, underscoring the master regulatory role of *gcm* in gliogenesis. During a second later period when glial cells had already differentiated, over 1200 genes were differentially regulated. Most of these genes, including many genes for chromatin remodeling factors, and cell cycle regulators, were not differentially expressed at the early stage indicating that the genetic control of glial fate determination is largely different from that involved in maintenance of differentiated cells. At both stages, glial-specific genes were upregulated and neuron-specific genes were downregulated supporting a model whereby *gcm* promotes glial development by activating glial genes while simultaneously repressing neuronal genes. Also at both stages, numerous genes that were not previously known to be involved in glial development were differentially regulated and, thus, identified as potential new downstream targets of *gcm*. For a subset of the differentially regulated genes, tissue-specific in vivo expression data were obtained which confirmed the transcript profiling results. This first genome-wide analysis of gene expression events downstream of a key developmental transcription factor presents a novel level of insight into the repertoire of genes that initiate and maintain cell fate choices in CNS development.**

## Introduction

During CNS development, two major cell types are generated, namely neurons and glial cells. These can be generated either by common precursors (neuroglioblasts) or by precursors that are specialized to produce either neurons (neuroblasts) or glial cells (glioblasts) (Akiyama-Oda et al., 1999; Bernardoni et al., 1999; Gage, 2000; Malatesta et al., 2000; Qian et al., 2000). Neuroglial development has been studied in detail in *Drosophila*, where each embryonic neuromere consists of approximately 60 glial cells and 700 neurons (Klämbt and Goodman, 1991; Ito et al., 1995; Schmidt et al., 1997; Jones, 2001). In *Drosophila*, neuroblasts divide asymmetrically to produce ganglion mother cells, which divide once to produce two neurons (Doe and Skeath, 1996), whereas glioblasts produce only glial cells. Glial and neuronal cell lineages in *Drosophila* also derive from neuroglioblasts, which divide asymmetrically to produce a neuroblast and a glioblast.

In *Drosophila*, the *glial cells missing (gcm)* gene encodes a transcription factor that controls the determination of glial versus neuronal fate in neuroectodermally derived cells (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996; Wegner and Riethmacher, 2001). In *gcm* mutants, cells that normally develop into glia enter a neuronal differentiation pathway leading to a loss of glia and a gain of neurons. In contrast, targeted *gcm* misexpression in neural progenitors leads to an increase of glial cells at the expense of neurons. Neither a specific embryonic stage nor a neural "ground state" appear necessary for *gcm* action since misexpression of *gcm* in epidermis or mesoderm suppresses normal cell fate and causes cells to adopt a glial fate (Akiyama-Oda et al., 1998; Bernardoni et al., 1998). Mesectodermal midline glial cells do not require *gcm* function (Granderath and Klämbt, 1999).

The molecular mechanisms of *gcm* action in glial development are poorly understood. Clearly, *gcm* transcription factor action depends on its target genes, however, relatively few genes, such as the *reversed polarity (repo)* gene, are known to act downstream of *gcm* (Akiyama et al., 1996). To identify *gcm* downstream genes in a comprehensive manner, we carried out a novel functional genomic approach using genome-wide oligonucleotide arrays. These arrays are used to analyze the transcripts in wild-type embryos versus embryos in which *gcm* is misexpressed throughout the CNS. Tissue-specific misexpression was achieved by using a *scabrous-GAL4* (*sca-GAL4*) line (Klaes et al., 1994) to drive *gcm* expression throughout the embryonic neuroectoderm. Transcripts were analysed at two defined temporal windows during embryogenesis. First, during a period of initial *gcm* action on determination of glial cell precursors and second during a later period when glial cells have already differentiated. In both cases, we find significant changes in transcript abundance for hundreds of identified genes following *gcm* misexpression. Remarkably, over half of these genes has not yet been studied in any in vivo context in *Drosophila*. All of these identified genes are potential direct or indirect downstream targets of *gcm* and may, thus, be involved in regulating glial cell fate.

## Materials and methods

### Flies

The wild-type was Oregon-R. For targeted misexpression of *gcm*, virgin females from *scabrous-GAL4* (Klaes et al., 1994) were crossed to *yw*; *UAS-gcm*; *UAS-gcm* males (Jones et al., 1995). For *gcm* loss-of-function studies the null allele *gcmΔP1* (Jones et al., 1995) was used balanced over *CyO-wglacZ*. Homozygous mutants were identified by absence of either anti-RK2/REPO or anti-βGAL staining. Stocks were kept on standard medium at 25°C. After a 1 hr pre-collection, wild-type and *sca-gcm* embryos were collected in parallel for 1 hr and staged to 6-7 hrs AEL (stage 11) or to 13-14 hrs AEL (late stage 15/early stage 16). Stages are according to (Campos-Ortega and Hartenstein, 1997).

### Arrays and hybridization

A custom-designed *Drosophila* oligonucleotide array (roDROMEGAa, Affymetrix, Santa Clara, CA) was used (Montalta-He et al., 2002). It contains 14,090 sequences representing 13,369 single transcripts encoding *Drosophila* proteins deposited in SWISS-PROT/TrEMBL databases (Celera Genome/BDGP Release no. 1) (Adams et al., 2000) as well as prokaryotic and custom chosen control sequences. Each sequence is represented on the array by a set of 14 oligonucleotide probes of matching sequence and 14 probes with a single nucleotide mismatch. 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 array as described in (Leemans et al., 2001). Four replicates were performed for each experimental condition.

### Data analysis

Data acquisition and processing was as described in (Leemans et al., 2001). For quantification of relative transcript abundance, the Average Difference value (Avg Diff) was used. All arrays were normalized against the mean of the total sums of Avg Diff values across all 16 arrays. For differential transcript imaging, only transcripts that had significant changes in Avg Diff ($p \leq 0.01$; unpaired t-test) in the 1.5-fold and above range were considered, and then only if the mean Avg Diff for the transcript was $\geq 50$ in at least one condition.

### In situ hybridization and immunocytochemistry

In situ hybridization was according to (Tautz and Pfeifle, 1989). Embryos were mounted in Canada balsam (Serva) and photographed with a Prog/Res/3008 digital camera (Kontron, Zürich) on a Zeiss Axioskop microscope with differential interference contrast optics. Immunocytochemical experiments were done according to (Therianos et al., 1995; Leemans

et al., 2001). The primary antibodies were rat anti-RK2/REPO 1:1000 (Campbell et al., 1994), mouse anti-TEN-M 1:250 (Baumgartner et al., 1994), rabbit anti-EY 1:500 (Kammermeier et al., 2001), mouse anti-WRAPPER 1:5 (Noordermeer et al., 1998) and goat anti-HRP (FITC-conjugated) 1:20 (Jackson Immunoresearch). For fluorescent labellings secondary antibodies were Alexa568 and Alexa488 conjugated, all 1:150 (Molecular Probes). For laser confocal microscopy, a Leica TCS SP was used.

## Results

### Targeted misexpression of *gcm* in the embryonic neuroectoderm results in a switch from neuronal to glial fate.

For targeted misexpression of *gcm* in the neuroectoderm of *Drosophila* embryos, a *scaGAL4* enhancer trap line (Klaes et al., 1994) was crossed with an *UAS-gcm* responder line (*sca-gcm*) (Jones et al., 1995). This resulted in ectopic *gcm* expression in the embryonic CNS starting from embryonic stage 9 and diminishing, similar to endogenous *gcm* expression, at embryonic stage 15. Although misexpression of *gcm* starts at stage 9 in *sca-gcm* embryos, ectopic expression of the *repo* gene, a known direct target of *gcm,* was not seen before stage 11, similar to endogenous *repo* expression.

In order to identify genes that are either direct *gcm* target genes or among the initial set of downstream genes of *gcm*, we carried out a first genome-wide analysis of differential gene expression at embryonic stage 11 when the first glial marker, the direct *gcm* target gene *repo,* is expressed. In the wild-type during stages 10-11, two small groups of neuroectodermal cells per hemisegment transiently express *gcm*, and a single *gcm*-expressing glial precursor delaminates from each of these groups and expresses the *repo* gene (Fig. 1A,C) (Hosoya et al., 1995; Jones et al., 1995). In contrast, in *sca-gcm* embryos during stages 10-11, all of the cells in the neuroectoderm express *gcm* (Fig. 1B) and, in consequence, most of the neural precursor cells become REPO positive (Fig. 1D). With the exception of altered gene expression in cells of the neuroectoderm, neither gene expression changes outside of the neural lineage nor any obvious morphological changes are seen in these stage 11 *sca-gcm* embryos.

In order to identify also additional indirect downstream genes of *gcm* that act further along in the genetic cascade of *gcm* action, we carried out a second genome-wide transcriptional analysis at embryonic stage 15/16 when glial cells are differentiated. In the wild-type at stage 15/16, approximately 700 neurons and 60 glial cells per neuromere have differentiated, and the glial cells (with the exception of midline glia) are REPO positive (Fig. 1E) (Ito et al., 1995). In contrast, in stage 15/16 sca-*gcm* embryos, 80%-90% of the cells in the CNS express REPO protein (Fig. 1F) and have a glial morphology (Hosoya et al., 1995). Correspondingly,

the number of cells expressing the neuronal marker *embryonic lethal abnormal vision (elav)* in these *sca-gcm* embryos is reduced by approximately 90% (data not shown) (Hosoya et al., 1995), and a striking reduction of the CNS axon scaffold is observed. In addition to the pronounced changes in the number of glial versus neuronal cells, stage 15/16 *sca-gcm* embryos also show defects in ventral nerve cord condensation and in peripheral innervation. No other gross morphological changes were seen in these stage 15/16 *sca-gcm* embryos.

**Overview of differential gene expression following *gcm* misexpression**

Analysis of differential gene expression in stage 11 and stage 15/16 *sca-gcm* versus wild-type embryos was carried out with oligonucleotide arrays representing 13,369 annotated *Drosophila* genes. This corresponds to virtually all of the currently annotated genes of the *Drosophila* genome sequence (Adams et al., 2000). For each embryonic stage, 2 x 4 replicate oligonucleotide arrays were used to detect transcript levels in *sca-gcm* embryos as compared to wild-type controls. Only transcripts which show an expression level fold change (FC) $\geq$ 1.5 or $\leq -1.5$ at significance values of $p \leq 0.01$ (t-test) were considered as differentially expressed (see materials and methods). A complete list of all of these genes, as well as their quantitative fold change values is given at www.ncbi.nlm.nih.gov/geo/.

At stage 11, we detected 417 transcripts with differential expression values in *sca-gcm* embryos as compared to wild-type. This corresponds to ~3% of the transcripts on the array. Approximately the same number of transcripts have increased (n=219) and decreased (n=198) abundance levels, indicating that *gcm* causes both activation and repression of downstream gene transcription. At stage 15/16, we detected 1259 genes with differential expression values in *sca-gcm* embryos compared to wild-type. This corresponds to ~9% of the transcripts on the array. Thus, markedly more transcripts are differentially expressed at stage 15/16 than at stage 11. Again, approximately the same number of transcripts are upregulated (n=609) and downregulated (n=650).

For an overview, all differentially expressed genes of known or predicted molecular function were grouped into functional classes. At stage 11, 199 transcripts of known function belonging to 13 functional classes are differentially expressed in *sca-gcm* embryos (Table 1). The two functional classes with the largest number of differentially regulated transcripts are enzymes (78) and nucleic acid binding proteins (44) including 26 transcription factors. At stage 15/16, 614 transcripts of known function belonging to 15 functional classes are differentially expressed in *sca-gcm* embryos (Table 1). The two functional classes with the largest number of differentially regulated transcripts are again enzymes (249) and nucleic acid binding proteins (96) including 38 transcription factors. Strikingly, however, at both stages, the majority of the differentially expressed transcripts are of (currently) unknown function; 218 (52%) at stage 11 and 645 (51%) at stage 15/16.

**Differential expression of genes encoding transcription factors**

The fact that *gcm* acts as a fate switch and key regulator of gliogenesis suggests that *gcm* might control a number of other transcription factors which in turn would regulate the expression of their own downstream genes. Transcript profiling of *gcm* misexpression indicated that *gcm* does indeed control the expression of numerous other transcription factors. These transcription factor encoding genes and a quantification of their changes in expression level are shown in figure 2.

In stage 11 embryos, 26 genes encoding transcription factors are differentially regulated by targeted *gcm* misexpression (11 upregulated, 15 downregulated). The *gcm* gene has the highest expression level increase (8.6 fold), in accordance with our experimental procedure. (The *gcm* gene also has a high absolute level of expression at this stage; see color coding in figure. 2). The *repo* gene, a known direct target of *gcm* (Akiyama et al., 1996), has the second highest increase in expression level (4.8-fold). Many of the other upregulated transcription factor genes such as *zinc finger homeobox-2* (*zhf-2*), *u-shaped* (*ush*) and the Enhancer of Split complex-member *HLHm3* are known to act in different aspects of embryonic nervous system development (Delidakis and Artavanis-Tsakonas, 1992; Lundell and Hirsh, 1992; Cubadda et al., 1997). Genes of the Enhancer of split complex, for example, act during neural versus epidermal cell fate decision (Jennings et al., 1994), and in the mouse, Enhancer of split members *Hes1* and *Hes5* have been shown to enhance glial cell fate (Furukawa et al., 2000; Hojo et al., 2000). Among the transcription factors with decreased expression levels are *engrailed* (*en*) and *ventral veins lacking/drifter* (*vvl/drf*) which are expressed in a subset of neuronal precursor cells and are also involved in midline glial cell development, but not in lateral glial cell development (Condron et al., 1994; Anderson et al., 1995). Other genes encoding transcription factors with decreased expression levels are *sloppy paired 1* (*slp1*), *goosecoid* (*gsc*), and *forkhead domain 96Cb* (*fd96Cb)*, which are expressed in subsets of neural precursor cells (Hacker et al., 1992; Hahn and Jäckle, 1996; Bhat et al., 2000). Moreover, the *scratch* (*scrt*) transcription factor, a pan-neuronal gene encoding a zinc finger protein that promotes neuronal development and can induce additional neurons when ectopically expressed (Roark et al., 1995), also shows decreased expression levels.

In stage 15/16 embryos, 38 genes encoding transcription factors are differentially regulated by targeted *gcm* misexpression (18 upregulated, 20 downregulated). As expected, *gcm* has the highest expression level increase (18.2 fold). (The absolute level of expression of the *gcm* gene is now an order of magnitude lower at this stage than at stage 11; see color coding in figure 2.) In contrast to high REPO protein levels in stage 15/16 *sca-gcm* embryos (Fig 1F), significant expression of *repo* transcripts is not detected at this stage. Several genes encoding transcription factors, which are expressed in specific neurons, such as *eyeless* (*ey*) and *Ultrabithorax* (*Ubx*) (Hirth et al., 1998; Kammermeier et al., 2001), are downregulated. Moreover, several members of the Enhancer of split complex such as *HLHmbeta*, *HLHm7*, and *E(spl)*, are downregulated at stage 15/16, in contrast to stage 11; in addition to a role in

early neurogenesis, these genes continue to be expressed in the normal developing nervous system of the wild-type at later embryonic stages (Wech et al., 1999).

The marked increase in the number of affected transcripts at stage15/16 is due in part to the fact that numerous genes encoding transcription factors belonging to the basal transcription machinery are differentially regulated at this stage. Among these are *TfIIFbeta*, *Taf55*, *TfIIB*, *Taf60*, *Taf80*, and *Taf150* (Frank et al., 1995; Lee et al., 1997; Aoyagi and Wassarman, 2000). Moreover, among the upregulated genes encoding transcription factors several are involved in chromatin remodeling such as the brahma complex or associated genes, *dalao* (*dalao*), *Brahma associated protein 60 kp* (*Bap60*) *Snf5-related 1*(*Snr1*) and *absent, small or homeotic disc 2* (*ash2*) (Francis and Kingston, 2001). This suggests that the maintenance of glial cell differentiation at later embryonic stages involves chromatin remodeling as well as the regulation of global transcriptional processes.

In addition to the above mentioned genes for transcription factors involved in chromatin remodeling, a number of genes encoding other proteins which bind to DNA/chromatin are influenced by *gcm* misexpression. These genes and a quantification of their expression level changes are shown in figure 4A. In stage 11 embryos, 7 genes encoding chromatin binding proteins are differentially regulated (3 upregulated, 4 downregulated), and at stage 15/16 embryos, 26 genes encoding chromatin binding proteins are differentially regulated (17 upregulated, 9 downregulated). Prominent among the upregulated genes thought to be involved in chromatin condensation and segregation are *gluon* (*glu*) (Steffensen et al., 2001) and the two DNA replication factor genes *Mini chromosome maintenance 6* and *7* (*Mcm6*) and (*Mcm7*) (Ohno et al., 1998). Among the genes with downregulated expression are the three Sox-related genes *sox-like* (*sox-like*), *Sox box protein 14 (Sox 14)* and *Dichaete* (*D*) which encode DNA bending proteins. *D* is known to be expressed in neural precursor cells and in midline glial cells (Soriano and Russell, 1998; Sanchez-Soriano and Russell, 2000).

Only 4 genes encoding DNA binding proteins, including 2 that encode transcription factors, are differentially expressed in both early and late stage *sca-gcm* embryos. This represents only 4% of the genes for DNA binding proteins that are differentially expressed in these embryos. This finding, which in qualitative terms holds for all other functional classes of differentially expressed genes, indicates that the molecular genetic mechanisms of early glial fate determination are largely different from those involved in the later maintenance of differentiated glial cells.

**Differential expression of genes encoding kinases and phosphatases**

Cell-cell interactions between neuronal and glial cells are crucial for key cellular processes such as metabolic exchange, extrinsic signaling and electrical insulation. The switch from neuronal to glial fate caused by *gcm* misexpression is, therefore, likely to affect genes that encode proteins involved in cell-cell signaling. Transcript imaging analysis of *gcm*

misexpression indicates that *gcm* does indeed control numerous genes that encode kinases and phosphatases involved in signaling pathways. A list of these genes as well as a quantitative representation of their changes in expression levels is shown in figure 3. Once again, a marked increase in the number of affected transcripts was observed at stage 15/16 as compared to stage 11.

In stage 11 embryos, 13 genes encoding kinases or phosphatases are differentially regulated by *gcm* misexpression (8 upregulated, 5 downregulated). Among the genes with increases in transcript abundance is *heartless* (*htl*) which encodes a fibroblast growth factor (FGF) receptor expressed in lateral glial cells (Shishido et al., 1997). Conversely the *Epidermal growth factor receptor* (*Egfr*) shows a decrease in transcript abundance; the *Egfr* pathway is implicated in midline glial cell development (Scholz et al., 1997). Decreased expression is also observed for *shaggy* (*sgg*), which encodes a protein kinase, and for *skittles* (*sktl*), which encodes a putative phosphatidylinositol-4-phosphate 5-kinase. Cells in *sgg* mutant embryos cannot adopt early epidermal fates and instead develop characteristics of CNS cells (Bourouis et al., 1989). Mutations in *sktl* cause abnormal development in the PNS (Prokopenko et al., 2000).

In stage 15/16 embryos, 59 genes encoding kinases or phosphatases are differentially regulated by *gcm* misexpression (29 upregulated, 30 downregulated). A number of genes involved in cell proliferation and mitotic division are upregulated. These included *polo* (*polo*), *discs overgrown* (*dco*), *smallminded* (*smid*) and *Nek2* (*Nek2*) (Llamazares et al., 1991; Schultz et al., 1994; Long et al., 1998; Zilian et al., 1999). In contrast, genes involved in aspects of neuronal development such as axogenesis and synaptogenesis are downregulated. Among these are *derailed* (*drl*), *Neuron-specific kinase* (*Nrk*), and *Cdk5 activator-like protein* (*Cdk5alpha*). The *drl* gene is involved in axonal guidance including routing across the midline (Bonkowsky et al., 1999). *Nrk* is specifically expressed in the embryonic CNS (Oishi et al., 1997). *Cdk5alpha* controls multiple aspects of axon patterning (Connell-Crowley et al., 2000). The only gene in this class that is known to be involved in glial differentiation is *htl*, which is upregulated at stage 11, and remains upregulated in stage 15/16 embryos albeit at a lower level.

**Differential expression of genes involved in cell cycle regulation**

As mentioned above, a number of chromatin binding protein and kinase/phosphatase encoding genes involved in cellular proliferation and in mitotic division are upregulated by *gcm* misexpression. This suggests that other genes involved in proliferation and division may also be affected by *gcm* misexpression. Transcript profiling of *gcm* misexpression indicates that *gcm* does indeed influence genes that encode cell cycle regulators. These genes and a quantitative representation of their changes in expression levels are shown in figure 4B.

10 genes encoding cell cycle regulators are differentially regulated by *gcm* misexpression (7 upregulated, 3 downregulated) in stage 15/16 embryos. For example, increases in transcript

abundance are found for *Cyclin B* (*CycB*), *Cyclin A* (*CycA*), and *Cyclin D* (*CycD*). These genes encode regulators of cyclin dependent kinases that act in different phases during mitotic cell cycles (Follette and O'Farrell, 1997). In contrast, and rather unexpectedly, a marked decrease in transcript abundance is found for *Cyclin E (CycE)*. *Cyc E* is essential for S-phase progression and its downregulation leads to the arrest of cell proliferation (Knoblich et al., 1994). Remarkably, in the earlier embryonic stage 11, none of the genes in the class of cell cycle regulators are influenced by *gcm* misexpression.

**Differential expression of genes encoding cell adhesion molecules**

Several cases for *gcm*-dependent regulation of genes encoding cell adhesion molecules were observed. These genes as well as a quantitative representation of their expression level changes is shown in figure 4C. At stage 11, 4 genes encoding cell adhesion molecules are differentially regulated by *gcm* misexpression (2 upregulated, 2 downregulated). At stage 15/16, 19 genes encoding cell adhesion molecules are differentially regulated by *gcm* misexpression (4 upregulated, 15 downregulated).

A striking example for a gene with a marked increased transcript level (13.6 fold) in stage 15/16 embryos is *wrapper*. *wrapper* encodes a cell adhesion molecule that is expressed in midline glial cells and in late stages also in lateral glial cells (Noordermeer et al., 1998). Genes with decreased transcript levels in stage 15/16 embryos that are mainly expressed in neurons are *Tenascin major* (*Ten-m*), *Cadherin-N* (*CadN*) and *neuromusculin* (*nrm*). All three act during axogenesis and synaptogenesis (Kania et al., 1993; Baumgartner et al., 1994; Levine et al., 1994; Iwai et al., 1997). The fact that most of the affected genes in the cell adhesion class show *gcm*-dependent decreased transcript levels could reflect the large diversity of cell adhesion molecules expressed by neurons.

***gcm* misexpression may influence genes that act in the hemocyte lineage**

In addition to its key role in gliogenesis, *gcm* also functions in a mesodermal lineage that gives rise to hematopoietic cells (Bernardoni et al., 1997; Lebestky et al., 2000). When ectopically expressed in the early mesoderm, *gcm* can induce expression of *Peroxidasin* (*Pxn*) which is a marker for macrophage cells. Misexpression of *gcm* in cells of the neural lineage also gives rise to a few cells that express hemocyte markers although most cells differentiate into glia (Bernardoni et al., 1997). In accordance with these findings, transcript profiling of *gcm* misexpression embryos indicates that several genes encoding marker proteins for cells of the hemocyte lineage are differentially regulated.

In stage 15/16 embryos, differential expression levels are detected for *Pxn, serpent* (*srp*), and *the Scavenger receptor class C (type I)* gene (Pearson et al., 1995), all of which are expressed in hemocytes. Scavenger receptors play a crucial role in the phagocytosis of apoptotic cells and might also be able to mediate the direct recognition of microbial pathogens (Platt et al.,

1998). It is noteworthy that the genes encoding Lysozyme B, Lysozyme C, Lysozyme D, and Lysozyme E are all upregulated by *gcm* misexpression in stage 11 embryos. These four closely related lysozyme genes, clustered at locus 61F on the third chromosome, function as part of a system of inducible antibacterial immunity (Daffre et al., 1994). These findings support the notion that the glial cell lineage and the hemocyte lineage, which gives rise to cells involved in defense and immunity, may be molecularly related (Bernardoni et al., 1997).

## Analysis of spatial expression of candidate *gcm* downstream genes by in situ hybridization and immunocytochemistry

To complement the quantitative transcript profiling analysis with tissue-specific spatial expression data, in situ hybridization and immunostaining was carried out on a subset of the genes that are differentially regulated by *gcm* misexpression (Fig. 5). In all cases, the qualitative changes in tissue-specific gene expression revealed by in situ hybridization and immunocytochemistry reflect and confirm the changes in gene expression determined by transcript profiling.

Expression of the transcripts for *htl*, *scrt*, *bangles-and-beads* (*bnb*) and *elav* was examined by in situ hybridization. In stage 11 wild-type embryos, the *htl* gene is expressed in a distinct set of neural precursors in the CNS (Fig. 5A). Outside of the CNS, *htl* is also expressed in elements of the mesodermal lineage (Shishido et al., 1997). Following targeted misexpression of *gcm* in cells of the neuroectoderm in stage 11 *sca-gcm* embryos, the expression of *htl* is expanded in the CNS region to include virtually all neural precursors (Fig. 5B). No changes in the expression of *htl* are seen outside of the CNS in these embryos. In stage 11 embryos, the pan-neural *scrt* gene is expressed in most or all neural precursors (Fig. 5C) (Roark et al., 1995). Following targeted misexpression of *gcm* in cells of the neuroectoderm in stage 11 *sca-gcm* embryos, the expression of *scrt* is diminished in most of the neural precursors, but is still apparent in a subset of these cells (Fig. 5D).

In stage 15/16 wild-type embryos, the *bnb* gene is expressed in lateral glial cells (Ng et al., 1989) (Fig. 5E,G). With the exception of a small group of cells near the anterior and posterior ends of the embryo, no other *bnb* expression is seen outside of the nervous system at this stage. In stage 15/16 *sca-gcm* embryos, the expression of *bnb* increases markedly and appears in virtually all of the cells of the nervous system (Fig. 5F,H). Expression of *bnb* outside of the nervous system does not appear to be influenced in these *sca-gcm* embryos. In stage 15/16 wild-type embryos, the *elav* gene is expressed in all neurons (Fig. 5I). In stage 15/16 in *sca-gcm* embryos, expression of *elav* is strongly reduced, but is still visible in some neurons of the brain as well as in some of the neurons that occupy the ventral-most cell layer in the ventral nerve cord (Fig. 5J).

Given that transcript abundance is not always reflected on the protein level (Keene, 2001), expression of three further candidate *gcm* downstream genes, *ey*, *Ten-m*, and *wrapper* was

investigated at the protein level by immunostaining in wild-type and *sca-gcm* embryos. In stage 15/16 wild-type embyos the EY protein is expressed in a segmentally reiterated subset of neurons in the CNS (Fig 5K). In stage 15/16 *sca-gcm* embryos the number of cells in the CNS that express the EY protein is dramatically reduced (Fig. 5L). In stage 15/16 wild-type embryos, the TEN-M protein is expressed on the axons that make up the longitudinal and commissural tracts of the CNS (Fig. 5M). This well defined axonal expression pattern of TEN-M protein is virtually abolished in stage 15/16 *sca-gcm* embryos (Fig. 5N). The TEN-M protein is also expressed outside of the nervous system (Baumgartner et al., 1994), but there is no obvious change in this non-neuronal expression of TEN-M in *sca-gcm* as compared to wild-type embryos. In stage 15/16 wild-type embryos, the WRAPPER protein is expressed in the midline and in some of the lateral glial cells, as well as in glial cells that support the chordotonal sensory organs in the PNS (Fig 5O)(Noordermeer et al., 1998). In stage 15/16 *sca-gcm* embryos a substantial increase of WRAPPER expression is seen in the CNS (Fig. 5P).

To control for possible effects of transgene insertion or of differences in genetic background, we repeated the tissue-specific spatial expression analysis for all of the above mentioned genes on embryos that contain either only the *sca-GAL4* construct or only the *UAS-gcm* constructs. In all cases, in situ hybridization and immunostaining results on these embryos were indistinguishable from results obtained on wild-type embryos (data not shown).

To determine if genes, that are influenced by *gcm* gain-of-function, might be influenced in an inverse way in *gcm* loss-of-function mutants, we studied tissue-specific spatial expression data of the candidate *gcm* downstream genes *repo, bnb, wrapper, elav, ey* and *Ten-m* in *gcm* null mutants using in situ hybridization and immunostaining. Expression of the genes *repo*, *bnb* and *wrapper* is upregulated in *gcm* gain-of-function embryos. In stage 15/16 *gcm* null mutant embryos, the expression of the *repo* gene in lateral glial cells, which is seen in the wild-type CNS, is strongly reduced (Fig. 5Q,R) (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996). Comparable findings are obtained for *bnb;* the expression of the *bnb* gene in lateral glial cells, which is seen in the wild-type CNS, disappears (Fig. 5Q,R). These findings contrast with stage 15/16 *sca-gcm* embryos, where the expression of *repo* and *bnb* appears in virtually all of the cells of the CNS. In stage 15/16 *gcm* null mutant embryos, the expression of *wrapper* in the lateral glial cells and in PNS glial cells (but not the midline glial cell expression), which is observed in wild-type embryos, disappears (Fig. 5S,T). This contrasts with stage 15/16 *sca-gcm* embryos, where the expression of *wrapper* becomes more widespread in the CNS. Expression of the genes *elav*, *ey* and *Ten-m* is downregulated in *gcm* gain-of-function embryos. Expression of *elav* in *gcm* null mutants is seen in additional neuronal cells as compared to the wild-type (data not shown) (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996). In contrast, the number of CNS cells that express either *ey* or *Ten-m* is not altered in *gcm* null mutants as compared to the wild-type (data not shown). This is not unexpected since *ey* and *Ten-m* are not pan-neuronally expressed as is *elav* but are expressed only in a subset of neuronal cells in the wild-type CNS (Baumgartner et al., 1994; Kammermeier et al., 2001).

## Discussion

### Candidate gene identification through genome-wide transcript imaging

By analysing gene expression profiles following *gcm* misexpression in the embryonic CNS, genome-wide transcript images were obtained for two phases of glial development. The first transcript image reflects an embryonic CNS in which precursor cells that normally give rise to neurons have been genetically reprogrammed to give rise to glial cells. It was obtained at an early embryonic stage when the first glial-specific genes, such as the *repo* gene (which is a direct target of *gcm)*, start to become expressed. This transcript image is therefore likely to identify genes that act in CNS precursors and are involved in the determination of glial versus neuronal cell lineage. Approximately 400 genes were found to be differentially expressed at this developmental stage, corresponding to 3% of the annotated genes in the fly genome. We posit that the genes that are differentially regulated at this early stage are either direct *gcm* target genes, such as *repo*, or among the initial set of downstream genes of *gcm*.

The second transcript image, obtained at a later embryonic stage when glial and neuronal cells are normally differentiated, reflects an embryonic CNS in which 80-90% of the normal number of neuronal cells have been genetically replaced by glial cells due to *gcm* action (Hosoya et al., 1995). This transcript image is therefore likely to identify genes that are involved in the maintenance of differentiated glial versus neuronal cells. Approximately 1300 genes were differentially expressed at this stage, corresponding to 9% of the annotated genes in the fly genome. We postulate that most of these differentially expressed genes are no longer direct or initial downstream targets of *gcm*, but are rather indirect downstream genes that act further along in the genetic cascade of *gcm* action.

The difference in total number of differentially expressed genes at the early stage versus the late stage is striking and, in qualitative terms, also holds for each of the major functional gene classes. Moreover, the overlap between the genes that are expressed at the two stages is restricted; only 93 (7%) of the 1259 genes that are differentially expressed at the late stage, are also differentially expressed at the early embryonic stage (Fig. 6). This suggests that the gene regulatory elements that control determination of glial cell fate are largely different from those required for maintenance of glial cell differentiation.

The expression profiles presented here derive from gain-of-function experiments in which the *gcm* gene is misexpressed in the embryonic CNS. A comparison of these findings with expression profiles derived from loss-of-function experiments involving *gcm* null mutants will be an important step in the further analysis of *gcm* downstream genes. However, in *gcm* null mutants only about 60 presumptive glial cells per segment are transformed into neurons; the 700 neurons of each segment are not affected. With the current sensitivity of oligonucleotide microarrays, it is unlikely that significant measurements of gene expression

changes in such a small number of cells can be obtained using whole-mount embryos. Thus, these complementary *gcm* loss-of-function experiments must await the development of single-cell isolation techniques for the embryonic nervous system of *Drosophila*.

**Candidate genes implicated in the determination of glial versus neuronal cell lineage**

Although *gcm* expression is necessary and sufficient to induce glial cell fate in and outside of the nervous system, it normally acts in glial precursors in the wild-type (Hosoya et al., 1995; Jones et al., 1995; Vincent et al., 1996; Akiyama-Oda et al., 1998; Bernardoni et al., 1998). A current model of this action is that *gcm* controls the determination of glial cell fate in glial precursors by activating both genes that promote glial differentiation and genes that repress neuronal differentiation (Giesen et al., 1997). Application of this model to our experimental paradigm leads to the prediction that expression of glial precursor-specific genes should be upregulated and expression of neuron precursor-specific genes should be downregulated in stage 11 *sca-gcm* embryos. Our genome-wide expression data verifies this prediction.

We find upregulation of genes that are known to be expressed in glial precursor cells. Prominent among these is *repo*, which contains 11 GCM consensus binding sites in its upstream regulatory region and is the first identified direct target of *gcm* (Akiyama et al., 1996). Another upregulated gene that is first expressed in the CNS in glial precursors is *htl*, which encodes a FGF-receptor (Shishido et al., 1997). Downregulation is found for several genes that are known to be expressed in neuronal precursors. Prominent among these is the pan-neuronal gene *scratch* (*scrt*) which promotes neurogenesis and can induce additional neurons when ectopically expressed (Roark et al., 1995). Interestingly, we also observe downregulation for several genes that are involved in midline glial cell development such as *Egfr, vvl, en* and *D* (Condron et al., 1994; Anderson et al., 1995; Scholz et al., 1997; Soriano and Russell, 1998). This suggests that midline glial development may be suppressed in *sca-gcm* embryos and might also explain the otherwise unexpected downregulation of *tramtrack* (*ttk*) and *pointed (pnt)*, since these genes are not only expressed in lateral glial cells but also in midline glial cells (Granderath and Klämbt, 1999).

In addition to genes that are known to be involved in the gliogenesis/neurogenesis decision, we find a large number of genes that have not previously been implicated in this aspect of CNS development. Indeed, for the majority of the known differentially regulated genes identified here, this report represents the first indication for an involvement in gliogenesis and/or neurogenesis. This is also the case for the annotated genes of unknown function, which have not been studied in any in vivo context, and which make up the majority of the differentially expressed genes identified.

The effects of targeted misexpression of *gcm* in stage 11 *sca-gcm* embryos appear to be restricted to cells of the neuroectoderm. Moreover, these effects manifest themselves primarily in altered gene expression in cells of the neuroectoderm. No morphological

changes are seen in stage 11 *sca-gcm* embryos as compared to wild-type, and non-specific side effects of *gcm* misexpression such as growth abnormalities, defective morphogenesis, or increased apoptosis are not observed in these embryos. We therefore assume that the observed differential gene expression specifically reflects activation or repression of *gcm* downstream genes. It is, nevertheless, unlikely that our study uncovers all of the genes that act downstream of *gcm* to induce glial cell fate. This is because our early transcript image is restricted to a specific time point in early gliogenesis development, and *gcm* may influence other targets at other stages. Moreover, the genetic overexpression of *gcm* may create an artificial situation in vivo, in which not all of the candidate downstream genes show changes in magnitude and direction of expression that correspond to their responses to *gcm* action under normal conditions. For example, whereas *gcm* expression in a mesodermal lineage induces genes involved in hemocyte cell development, overexpression of *gcm* in neuroectodermal cells causes a downregulation of the hemocyte marker genes *Pxn* and *srp*. (Downregulation of *Pxn* may, however, also be due to the fact that this gene is also expressed in the nervous system.) Finally, it is conceivable that some of the gene expression changes seen in *sca-gcm* embryos as compared to wild-type are due to insertional effects of the transgenes or to differences in genetic background. While we find no evidence for such effects among the 10 *gcm* candidate genes that we characterized by in situ and immunocytochemical experiments, we cannot rule out such effects for all of the candidate genes identified in this report. In consequence, a full appreciation and verification of all of these candidate *gcm* downstream genes and a comprehensive understanding of their roles in determination of glial versus neuronal cell lineage will require a careful gene-by-gene analysis in mutant embryos. This also applies to the genes that are differentially expressed in stage 15/16 embryos.

### Candidate genes implicated in the maintenance of differentiated glial versus neuronal cells

In stage 15/16 *sca-gcm* embryos most of the neurons in the embryonic nervous system are genetically replaced by glial cells, and differential gene expression in these embryos as compared to wild-type embryos reflects this fact. While the transcript image obtained at this stage will, therefore, identify genes that are involved in the maintenance of differentiated glial versus neuronal cells, non-specific side effects of *gcm* misexpression on differential gene expression cannot be ruled out. This is because the marked loss of neurons in stage 15/16 *sca-gcm* embryos results in morphological changes such as defective condensation of the CNS or reduction of peripheral innervation, and these morphological alterations may be accompanied by changes in gene expression.

Nevertheless, given that the strongest phenotype of stage 15/16 *sca-gcm* embryos is the gain of glial cells at the expense of neurons, we postulate that most of the observed differential gene expression at this stage is directly related to the replacement of differentiated neurons by differentiated glial cells. This is supported by the fact that a number of genes that are known

to be expressed in differentiated neurons such as *elav*, *lark*, *Ten-m* and *CadN* are downregulated while genes that are expressed in differentiated glial cells such as *htl*, *wrapper* and *bnb* are upregulated in stage 15/16 *sca-gcm* embryos. In several cases, however, genes encoding markers for lateral or peripheral glia were not judged to be upregulated by our data analysis. For example, for the genes *repo*, *locomotion defects* (*loco)*, and *gliotactin (gli)*, the normalized expression levels, the fold change levels or the statistical significance levels were below our threshold filter values, so that these genes were not considered to be upregulated in our microarray experiments.

Cell fate determination is often controlled at the transcriptional level by key regulatory factors that are expressed transiently, whereas the gene expression patterns that they establish persist. Maintenance of the transcriptional state in differentiated cells is then achieved by control elements involved in chromatin remodeling and modification (Francis and Kingston, 2001). Accordingly, in our analysis of stage 15/16 *sca-gcm* versus wild-type embryos, we identified several differentially expressed genes that are involved in chromatin remodeling such as *Bap60, dalao, Snr1* and *ash2*. In specific glial lineages, the onset of differentiation is thought to require cell cycle progression (Akiyama-Oda et al., 2000). In our analysis, differential expression of genes encoding cell cycle regulators or proteins involved in chromatin condensation and segregation during mitosis was also observed. Examples of this are cyclin encoding genes such as *CycB, CycA, CycD* and *CycE* which are differentially expressed in stage 15/16 embryos. The differential expression of these genes following *gcm* misexpression provides further support for the general notion that cell cycle regulators are key elements in cellular differentiation processes (Ohnuma et al., 2001).

In summary, this study combines in vivo transgenic analysis with genome-wide expression analysis based on oligonucleotide arrays to identify genes that are downstream of *gcm*, a key transcriptional control element in gliogenesis. The results of this study should be helpful in obtaining a comprehensive view of the molecular mechanisms of cell fate specification and cell type maintenance in the developing nervous system.

## Acknowledgments

We thank A. Roth, M. Tessier, M. Wilhelm-Seiler, C. Broger, M. Neeb, R. Brem, A. Giangrande, K. Takizawa, B. W. Jones, K. J. Schmid, M. Affolter and his group, A. Ebner, L. Michaut, A. Tomlinson, S. Baumgartner, U. Walldorf, R.A.H. White, J. N. Noordermeer, H. Montalta-He the Bloomington Stock Center for technical support, reagents and flies and M. Müller, F. Hirth, and M. Götz for helpful comments on the manuscript.

## References


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, 2185-2195.

Akiyama, Y., Hosoya, T., Poole, A. M. and Hotta, Y. (1996). The *gcm*-motif: a novel DNA-binding motif conserved in *Drosophila* and mammals. *Proc. Natl. Acad. Sci. U S A* 93, 14912-14916.

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. *Dev. Genes Evol.* 208, 578-585.

Akiyama-Oda, Y., Hosoya, T. and Hotta, Y. (1999). Asymmetric cell division of thoracic neuroblast 6-4 to bifurcate glial and neuronal lineage in *Drosophila*. *Development* 126, 1967-1974.

Akiyama-Oda, Y., Hotta, Y., Tsukita, S. and Oda, H. (2000). Distinct mechanisms triggering glial differentiation in *Drosophila* thoracic and abdominal neuroblasts 6-4. *Dev. Biol.* 222, 429-439.

Anderson, M. G., Perkins, G. L., Chittick, P., Shrigley, R. J. and Johnson, W. A. (1995). *drifter*, a *Drosophila* POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia. *Genes Dev.* 9, 123-137.

Aoyagi, N. and Wassarman, D. A. (2000). Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. *J. Cell. Biol.* 150, F45-50.

Baumgartner, S., Martin, D., Hagios, C. and Chiquet-Ehrismann, R. (1994). *ten$^{m}$*, a *Drosophila* gene related to tenascin, is a new pair-rule gene. *EMBO J.* 13, 3728-3740.

Bernardoni, R., Vivancos, B. and Giangrande, A. (1997). *glide*/*gcm* is expressed and required in the scavenger cell lineage. *Dev. Biol.* 191, 118-130.

Bernardoni, R., Miller, A. A. and Giangrande, A. (1998). Glial differentiation does not require a neural ground state. *Development* 125, 3189-3200.

Bernardoni, R., Kammerer, M., Vonesch, J. L. and Giangrande, A. (1999). Gliogenesis depends on *glide*/*gcm* through asymmetric division of neuroglioblasts. *Dev. Biol.* 216, 265-275.

Bhat, K. M., van Beers, E. H. and Bhat, P. (2000). Sloppy paired acts as the downstream target of Wingless in the *Drosophila* CNS and interaction between *sloppy paired* and *gooseberry* inhibits *sloppy paired* during neurogenesis. *Development* 127, 655-665.

Bonkowsky, J. L., Yoshikawa, S., O'Keefe, D. D., Scully, A. L. and Thomas, J. B. (1999). Axon routing across the midline controlled by the *Drosophila* Derailed receptor. *Nature* 402, 540-544.

Bourouis, M., Heitzler, P., el Messal, M. and Simpson, P. (1989). Mutant *Drosophila* embryos in which all cells adopt a neural fate. *Nature* 341, 442-444.

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, 2957-2966.
Campos-Ortega, J. and Hartenstein, V. (1997). The Embryonic Development of *Drosophila melanogaster*. Heidelberg: Springer.
Condron, B. G., Patel, N. H. and Zinn, K. (1994). Engrailed controls glial/neuronal cell fate decisions at the midline of the central nervous system. *Neuron* 13, 541-554.
Connell-Crowley, L., Le Gall, M., Vo, D. J. and Giniger, E. (2000). The cyclin-dependent kinase Cdk5 controls multiple aspects of axon patterning in vivo. *Curr. Biol.* 10, 599-602.
Cubadda, Y., Heitzler, P., Ray, R. P., Bourouis, M., Ramain, P., Gelbart, W., Simpson, P. and Haenlin, M. (1997). *u-shaped* encodes a zinc finger protein that regulates the proneural genes *achaete* and *scute* during the formation of bristles in *Drosophila*. *Genes Dev.* 11, 3083-3095.
Daffre, S., Kylsten, P., Samakovlis, C. and Hultmark, D. (1994). The lysozyme locus in *Drosophila melanogaster*: an expanded gene family adapted for expression in the digestive tract. *Mol. Gen. Genet.* 242, 152-162.
Delidakis, C. and Artavanis-Tsakonas, S. (1992). The Enhancer of split [E(spl)] locus of *Drosophila* encodes seven independent helix-loop-helix proteins. *Proc. Natl. Acad. Sci. U S A* 89, 8731-8735.
Doe, C. Q. and Skeath, J. B. (1996). Neurogenesis in the insect central nervous system. *Curr. Opin. Neurobiol.* 6, 18-24.
Follette, P. J. and O'Farrell, P. H. (1997). Cdks and the *Drosophila* cell cycle. *Curr. Opin. Genet. Dev.* 7, 17-22.
Francis, N. J. and Kingston, R. E. (2001). Mechanisms of transcriptional memory. *Nat. Rev. Mol. Cell. Biol.* 2, 409-421.
Frank, D. J., Tyree, C. M., George, C. P. and Kadonaga, J. T. (1995). Structure and function of the small subunit of TFIIF (RAP30) from *Drosophila* melanogaster. *J. Biol. Chem.* 270, 6292-6297.
Furukawa, T., Mukherjee, S., Bao, Z. Z., Morrow, E. M. and Cepko, C. L. (2000). rax, Hes1, and notch1 promote the formation of Muller glia by postnatal retinal progenitor cells. *Neuron* 26, 383-394.
Gage, F. H. (2000). Mammalian neural stem cells. *Science* 287, 1433-1438.
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, 2307-2316.
Granderath, S. and Klämbt, C. (1999). Glia development in the embryonic CNS of *Drosophila*. *Curr. Opin. Neurobiol.* 9, 531-536.
Hacker, U., Grossniklaus, U., Gehring, W. J. and Jäckle, H. (1992). Developmentally regulated *Drosophila* gene family encoding the fork head domain. *Proc. Natl. Acad. Sci. U S A* 89, 8754-8758.
Hahn, M. and Jäckle, H. (1996). *Drosophila goosecoid* participates in neural development but not in body axis formation. *EMBO J.* 15, 3077-3084.
Hirth, F., Hartmann, B. and Reichert, H. (1998). Homeotic gene action in embryonic brain development of *Drosophila*. *Development* 125, 1579-1589.

Hojo, M., Ohtsuka, T., Hashimoto, N., Gradwohl, G., Guillemot, F. and Kageyama, R. (2000). Glial cell fate specification modulated by the bHLH gene Hes5 in mouse retina. *Development* 127, 2515-2522.
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, 1025-1036.
Ito, K., Urban, J. and Technau, G. M. (1995). Distribution, classification, and development of *Drosophila* glial cells in the late embryonic and early larval ventral nerve cord. *Roux's Arch. Dev. Biol.* 204, 284-307.
Iwai, Y., Usui, T., Hirano, S., Steward, R., Takeichi, M. and Uemura, T. (1997). Axon patterning requires DN-cadherin, a novel neuronal adhesion receptor, in the *Drosophila* embryonic CNS. *Neuron* 19, 77-89.
Jennings, B., Preiss, A., Delidakis, C. and Bray, S. (1994). The Notch signalling pathway is required for Enhancer of split bHLH protein expression during neurogenesis in the *Drosophila* embryo. *Development* 120, 3537-3548.
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, 1013-1023.
Jones, B. W. (2001). Glial cell development in the *Drosophila* embryo. *Bioessays* 23, 877-887.
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, 71-78.
Kania, A., Han, P. L., Kim, Y. T. and Bellen, H. (1993). *neuromusculin*, a *Drosophila* gene expressed in peripheral neuronal precursors and muscles, encodes a cell adhesion molecule. *Neuron* 11, 673-687.
Keene, J. D. (2001). Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. *Proc. Natl. Acad. Sci. U S A* 98, 7018-7024.
Klaes, A., Menne, T., Stollewerk, A., Scholz, H. and Klämbt, C. (1994). The Ets transcription factors encoded by the *Drosophila* gene *pointed* direct glial cell differentiation in the embryonic CNS. *Cell* 78, 149-160.
Klämbt, C. and Goodman, C. S. (1991). The diversity and pattern of glia during axon pathway formation in the *Drosophila* embryo. *Glia* 4, 205-213.
Knoblich, J. A., Sauer, K., Jones, L., Richardson, H., Saint, R. and Lehner, C. F. (1994). Cyclin E controls S phase progression and its down-regulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation. *Cell* 77, 107-120.
Lebestky, T., Chang, T., Hartenstein, V. and Banerjee, U. (2000). Specification of *Drosophila* hematopoietic lineage by conserved transcription factors. *Science* 288, 146-149.
Lee, K. S., Oh, Y., Baek, G., Yoon, J., Han, K., Cho, N. and Baek, K. (1997). Analysis of the structure and expression fo the *TFIIB* gene in *Drosophila melanogaster*. *Mol. Cells* 7, 374-379.
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, research 0015.1-0015.9.
Levine, A., Bashan-Ahrend, A., Budai-Hadrian, O., Gartenberg, D., Menasherow, S. and Wides, R. (1994). *Odd Oz*: a novel *Drosophila* pair rule gene. *Cell* 77, 587-598.
Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* 21, 20-24.
Llamazares, S., Moreira, A., Tavares, A., Girdham, C., Spruce, B. A., Gonzalez, C., Karess, R. E., Glover, D. M. and Sunkel, C. E. (1991). *polo* encodes a protein kinase homolog required for mitosis in *Drosophila*. *Genes Dev.* 5, 2153-2165.
Long, A. R., Wilkins, J. C. and Shepherd, D. (1998). Dynamic developmental expression of *smallminded*, a *Drosophila* gene required for cell division. *Mech. Dev.* 76, 33-43.
Lundell, M. J. and Hirsh, J. (1992). The *zfh-2* gene product is a potential regulator of neuron-specific dopa decarboxylase gene expression in *Drosophila*. *Dev. Biol.* 154, 84-94.
Malatesta, P., Hartfuss, E. and Götz, M. (2000). Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development* 127, 5253-5263.
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, research 0015.1-0015.15.
Ng, S. C., Perkins, L. A., Conboy, G., Perrimon, N. and Fishman, M. C. (1989). A *Drosophila* gene expressed in the embryonic CNS shares one conserved domain with the mammalian GAP-43. *Development* 105, 629-638.
Noordermeer, J. N., Kopczynski, C. C., Fetter, R. D., Bland, K. S., Chen, W. Y. and Goodman, C. S. (1998). Wrapper, a novel member of the Ig superfamily, is expressed by midline glia and is required for them to ensheath commissural axons in *Drosophila*. *Neuron* 21, 991-1001.
Ohno, K., Hirose, F., Inoue, Y. H., Takisawa, H., Mimura, S., Hashimoto, Y., Kiyono, T., Nishida, Y. and Matsukage, A. (1998). cDNA cloning and expression during development of *Drosophila melanogaster* MCM3, MCM6 and MCM7. *Gene* 217, 177-185.
Ohnuma, S., Philpott, A. and Harris, W. A. (2001). Cell cycle and cell fate in the nervous system. *Curr. Opin. Neurobiol.* 11, 66-73.
Oishi, I., Sugiyama, S., Liu, Z. J., Yamamura, H., Nishida, Y. and Minami, Y. (1997). A novel *Drosophila* receptor tyrosine kinase expressed specifically in the nervous system. Unique structural features and implication in developmental signaling. *J. Biol. Chem.* 272, 11916-11923.
Pearson, A., Lux, A. and Krieger, M. (1995). Expression cloning of *dSR-CI*, a class C macrophage-specific scavenger receptor from *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U S A* 92, 4056-4060.
Platt, N., da Silva, R. P. and Gordon, S. (1998). Recognizing death: the phagocytosis of apoptotic cells. *Trends Cell Biol*. 8, 365-372.
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, 1691-1715.

Qian, X., Shen, Q., Goderie, S. K., He, W., Capela, A., Davis, A. A. and Temple, S. (2000). Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron* 28, 69-80.
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, 2384-2398.
Sanchez-Soriano, N. and Russell, S. (2000). Regulatory mutations of the *Drosophila* Sox gene *Dichaete* reveal new functions in embryonic brain and hindgut development. *Dev. Biol.* 220, 307-321.
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.
Scholz, H., Sadlowski, E., Klaes, A. and Klämbt, C. (1997). Control of midline glia development in the embryonic *Drosophila* CNS. *Mech. Dev.* 64, 137-151.
Schultz, S. J., Fry, A. M., Sutterlin, C., Ried, T. and Nigg, E. A. (1994). Cell cycle-dependent expression of *Nek2*, a novel human protein kinase related to the NIMA mitotic regulator of *Aspergillus nidulans*. *Cell Growth Differ.* 5, 625-635.
Shishido, E., Ono, N., Kojima, T. and Saigo, K. (1997). Requirements of DFR1/Heartless, a mesoderm-specific *Drosophila* FGF- receptor, for the formation of heart, visceral and somatic muscles, and ensheathing of longitudinal axon tracts in CNS. *Development* 124, 2119-2128.
Soriano, N. S. and Russell, S. (1998). The *Drosophila* SOX-domain protein Dichaete is required for the development of the central nervous system midline. *Development* 125, 3989-3996.
Steffensen, S., Coelho, P. A., Cobbe, N., Vass, S., Costa, M., Hassan, B., Prokopenko, S. N., Bellen, H., Heck, M. M. and Sunkel, C. E. (2001). A role for *Drosophila* SMC4 in the resolution of sister chromatids in mitosis. *Curr. Biol.* 11, 295-307.
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, 81-85.
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, 3849-3860.
Vincent, S., Vonesch, J. L. and Giangrande, A. (1996). *Glide* directs glial fate commitment and cell fate switch between neurones and glia. *Development* 122, 131-139.
Wech, I., Bray, S., Delidakis, C. and Preiss, A. (1999). Distinct expression patterns of different enhancer of split bHLH genes during embryogenesis of *Drosophila melanogaster*. *Dev. Genes Evol.* 209, 370-375.
Wegner, M. and Riethmacher, D. (2001). Chronicles of a switch hunt: gcm genes in development. *Trends Genet.* 17, 286-290.
Zilian, O., Frei, E., Burke, R., Brentrup, D., Gutjahr, T., Bryant, P. J. and Noll, M. (1999). *double-time* is identical to *discs overgrown*, which is required for cell survival, proliferation and growth arrest in *Drosophila* imaginal discs. *Development* 126, 5409-5420.

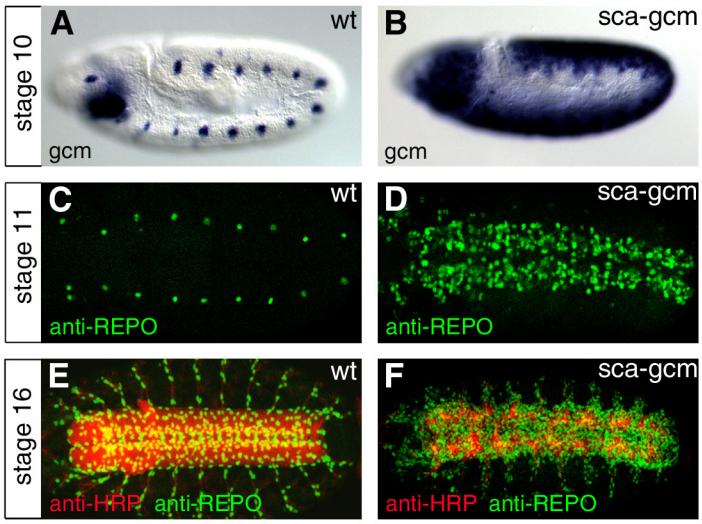
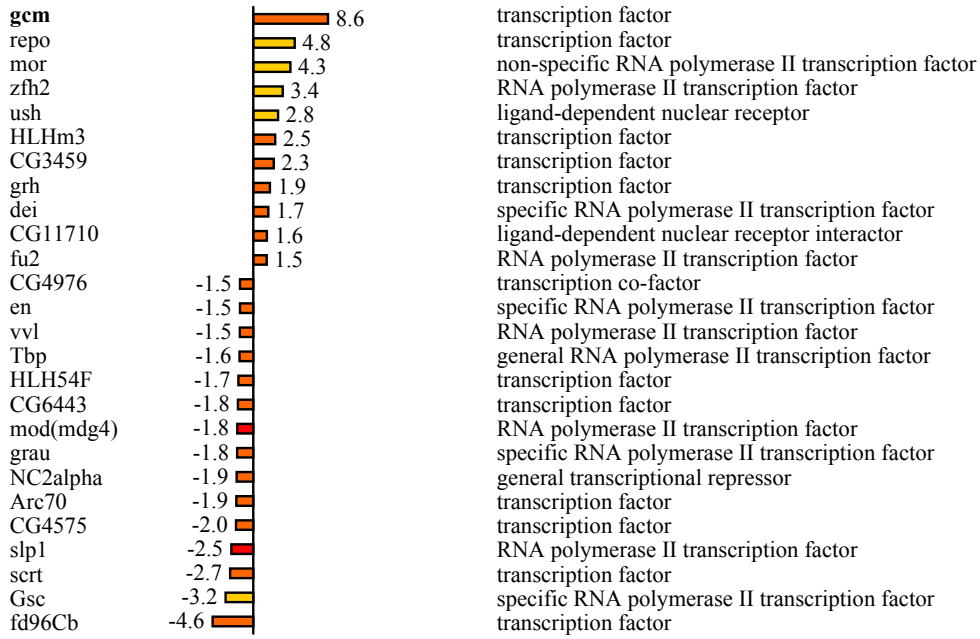



**Fig. 1**. Targeted misexpression of *gcm* results in gain of glial cells at the expense of neuronal cells. (A,B) In situ hybridisation of stage 10 embryos shows gcm expression in the wild-type (A) and in *sca-gcm* embryos (B); lateral views, anterior to the left. In the wild-type, small clusters of cells in the neuroectoderm express *gcm*; in *sca-gcm* embryos, all cells of the neuroectoderm express *gcm*. (C,D) Immunostaining with anti-REPO in the wild-type (C) and in *sca-gcm* embryos (D); laser confocal microscopy of stage 11, ventral views of the VNC, anterior is to the left. In the wild-type, single *gcm*-expressing glial precursors in each hemisegment express the *repo* gene. In *sca-gcm* embryos, virtually all of the neuronal and glial precursor cells are REPO positive. (E,F) Double immunostaining with anti-REPO (green), and anti-HRP (red) in the wild-type (E) and in *sca-gcm* embryos (F); laser confocal microscopy of stage 15/16 embryos. In the wild-type, neurons and glial cells are differentiated and correctly positioned, and all lateral glial cells express *repo*. In *sca-gcm* embryos, 80%-90% of the cells in the CNS express *repo*.

## Transcription factor

**stage 11**




**stage 15/16**

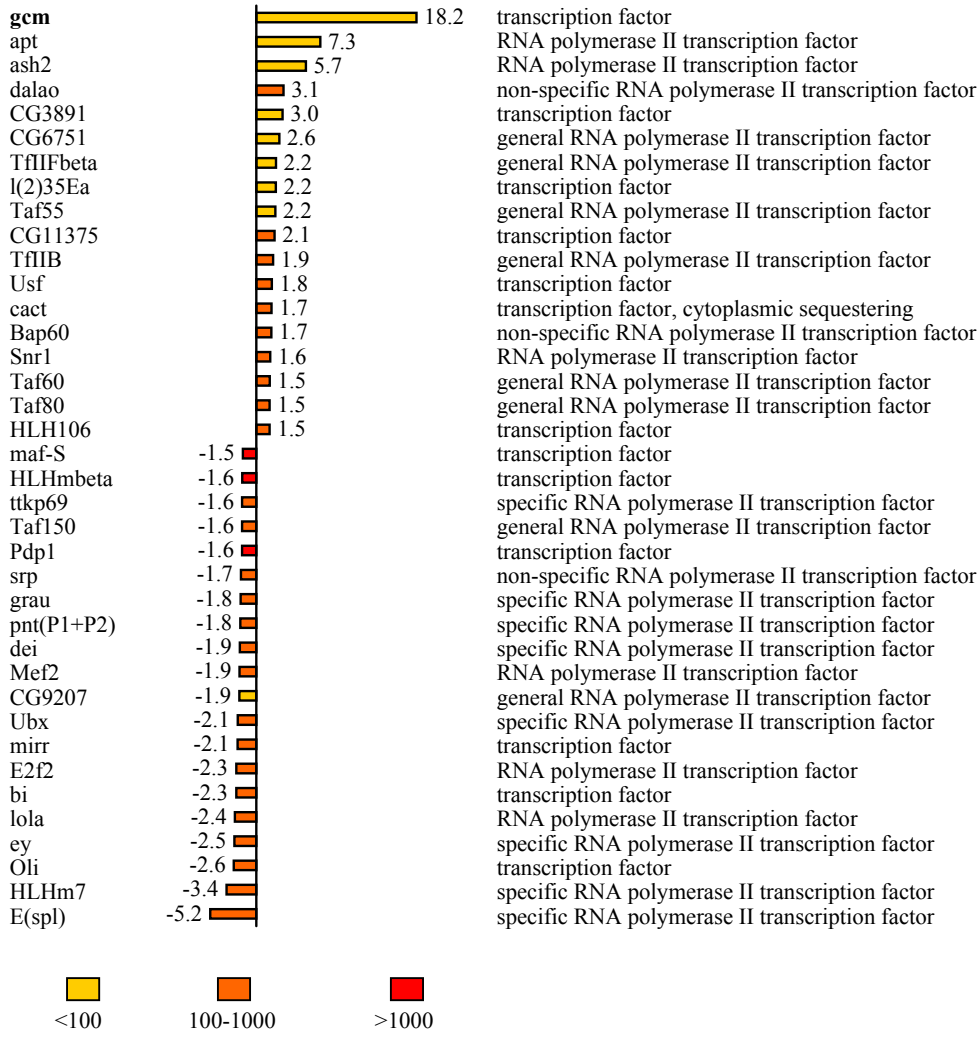
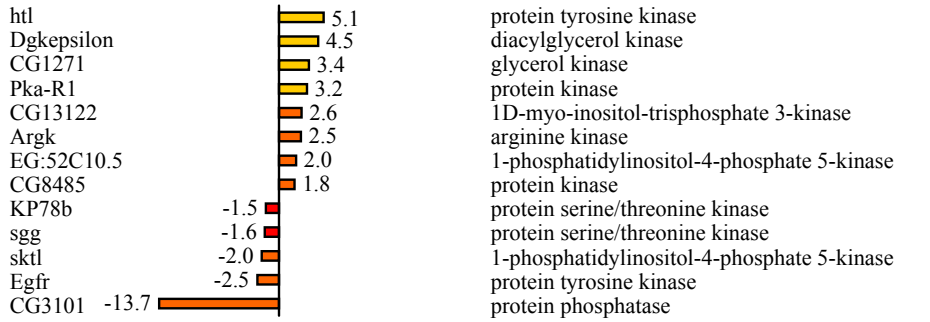



**Fig. 2.** Changes in transcript levels of the genes encoding transcription factors following *gcm* misexpression. Bars represent the fold changes in gene expression levels between wild-type embryos and *sca -gcm* embryos. Positive values indicate that the relative expression level of a gene is increased (upregulation) and negative values indicate a decrease (downregulation). Normalized average difference (Avg Diff) values are given for the wild-type condition as follows: yellow bars represent Avg Diff <100, orange bars represent Avg Diff ranging from 100-1000, and red bars represent Avg Diff >1000.

## Kinase/phosphatase

**stage 11**




**stage 15/16**

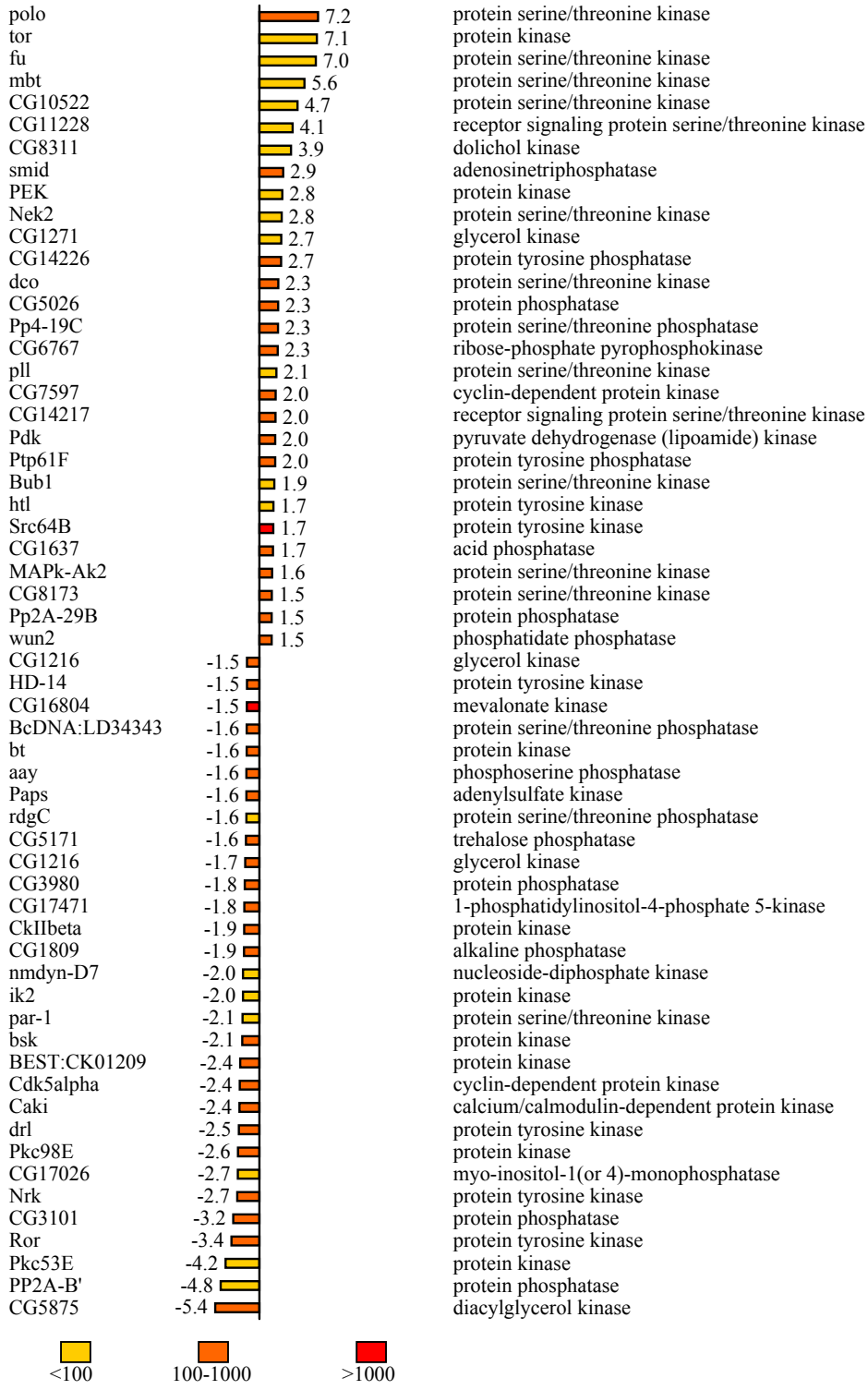
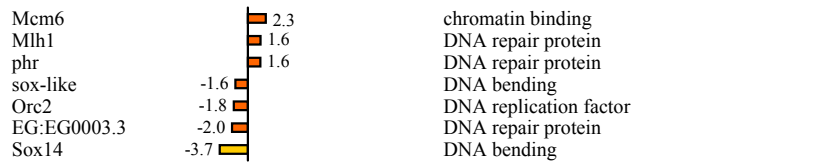




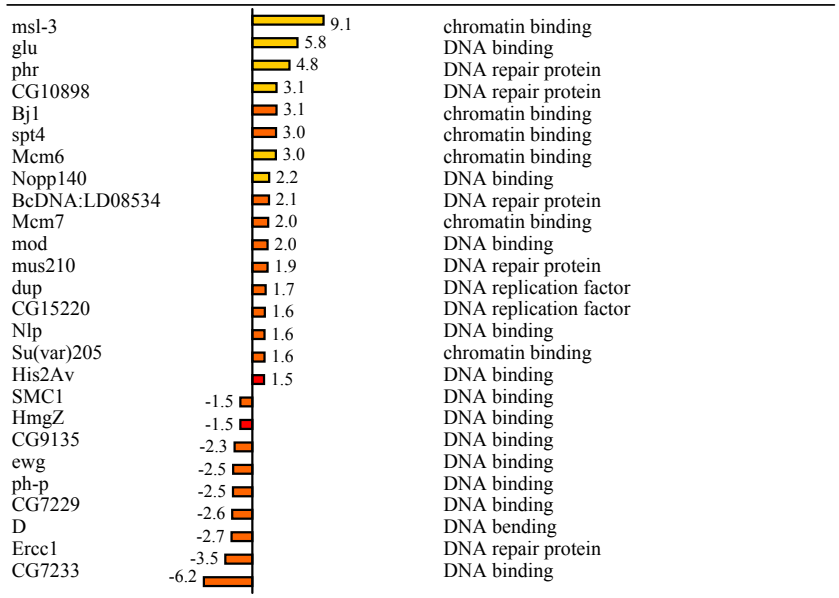

**Fig. 3.** Changes in transcript levels for the genes encoding protein kinases and phosphatases following *gcm* misexpression. Data presentation as in figure 2.

## A DNA binding/chromatin binding

**stage 11**

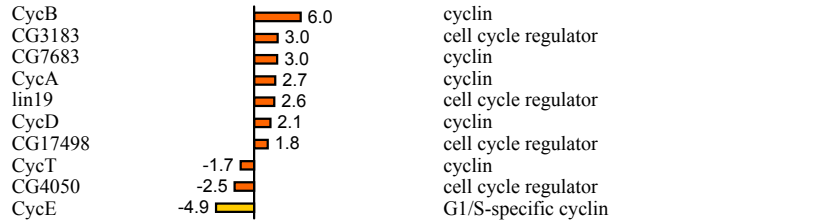



**stage 15/16**



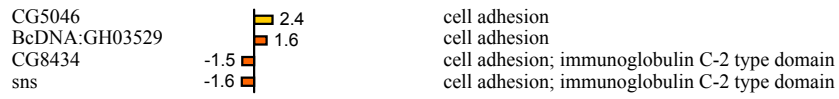

## B Cell cycle regulator

**stage 15/16**




## C Cell adhesion

**stage 11**




**stage 15/16**

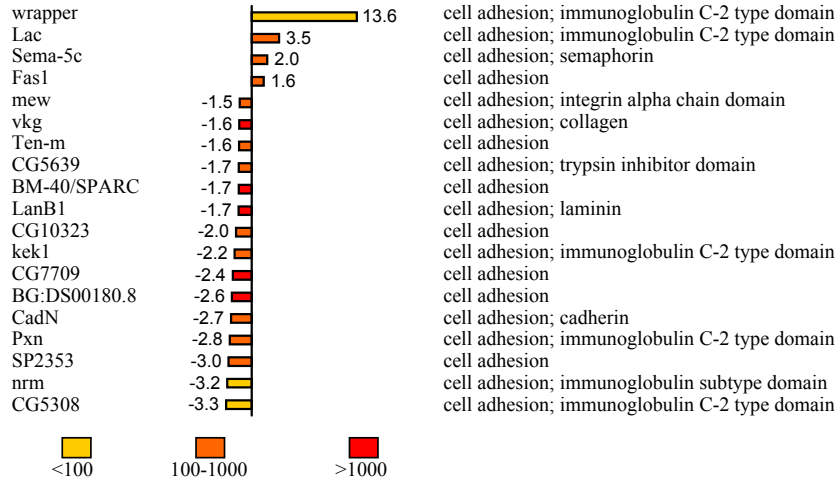



**Fig. 4.** Changes in transcript levels for the genes encoding DNA/chromatin binding proteins, cell cycle regulators or cell adhesion molecules following *gcm* misexpression.
(A) Genes encoding DNA/chromatin binding proteins. (B) Genes encoding cell cycle regulators. (C) Genes encoding cell adhesion molecules. Data presentation as in figure 2.

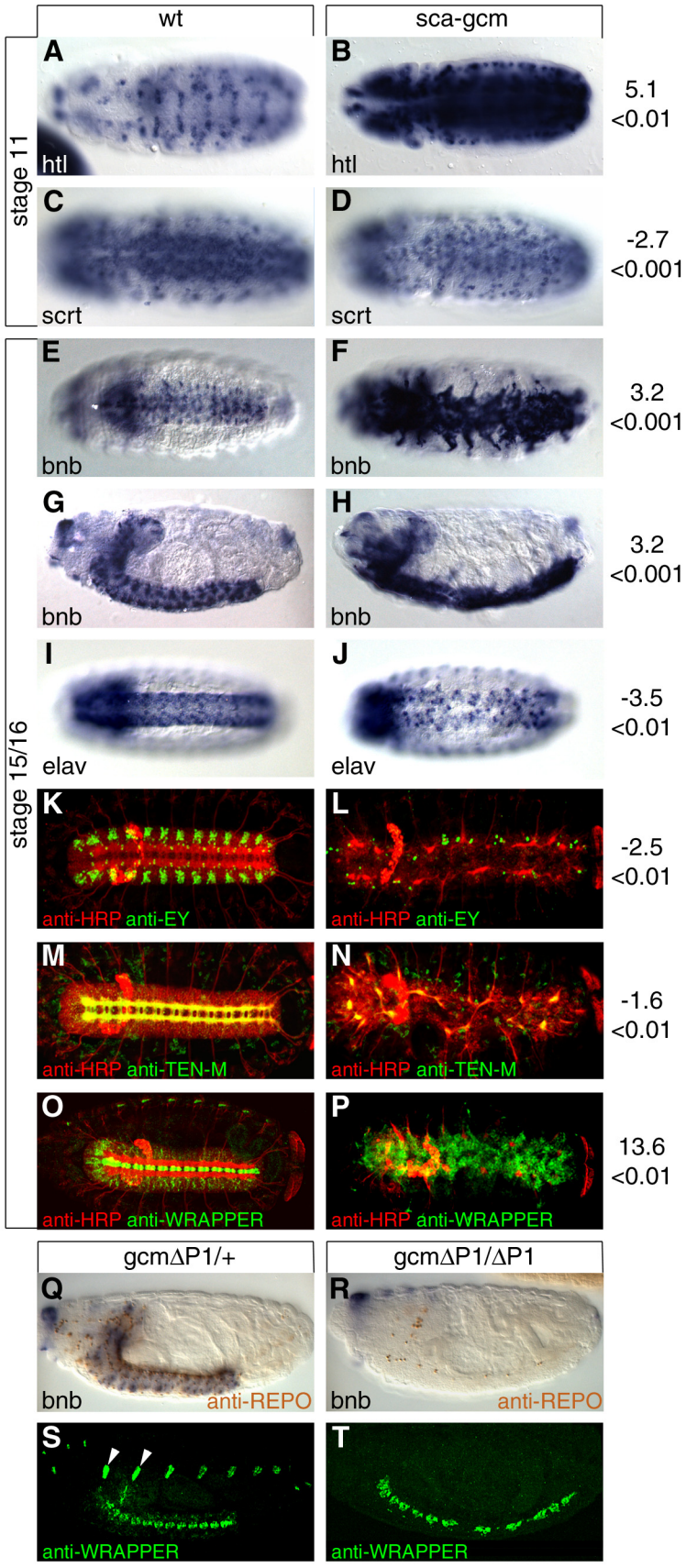

wt
sca-gcm
stage 11
stage 15/16
A
htl
B
htl
5.1
<0.01
C
scrt
D
scrt
-2.7
<0.001
E
bnb
F
bnb
3.2
<0.001
G
bnb
H
bnb
3.2
<0.001
I
elav
J
elav
-3.5
<0.01
K
anti-HRP anti-EY
L
anti-HRP anti-EY
-2.5
<0.01
M
anti-HRP anti-TEN-M
N
anti-HRP anti-TEN-M
-1.6
<0.01
O
anti-HRP anti-WRAPPER
P
anti-HRP anti-WRAPPER
13.6
<0.01
gcmΔP1/+
gcmΔP1/ΔP1
Q
bnb
anti-REPO
R
bnb
anti-REPO
S
anti-WRAPPER
T
anti-WRAPPER

**Fig. 5**. Spatial expression of selected candidate *gcm* downstream genes by in situ hybridization and immunocytochemistry. Whole mount in situ hybridization (A-J,Q,R) and immunostaining (K-T) show expression of differentially regulated genes in wild-type, *sca-gcm* and *gcm* mutant embryos. Ventral views of stage 11 (A-D) and stage 15/16 (E,F,I,J,K-P) embryos and lateral views of stage 15/16 embryos (G,H,Q-T), anterior is to the left. Fold changes and p-values are indicated to the right. (A,B) Expression of *htl* in stage 11 wild-type embryos is visible in a distinct set of neural precursors; in sca-*gcm* embryos, *htl* is expressed throughout the neurogenic region. (C,D) In stage 11 embryos, the *scrt* gene is expressed in neural precursors; in stage 11 *sca-gcm* embryos, the expression of *scrt* is diminished in most of the neural precursors, but is still apparent in a subset of these cells. (E-H) In stage 15/16 wild-type embryos, *bnb* gene is expressed in lateral glial cells; in stage 15/16 *sca-gcm* embryos, the expression of *bnb* increases markedly and appears virtually in all of the cells of the nervous system. (I,J) In stage 15/16 wild-type embryos, the *elav* gene is expressed in all neurons; in stage 15/16 in *sca-gcm* embryos, expression of *elav* is strongly reduced in most of the neurons. (K,L) In stage 15/16 wild-type embyos the EY protein is expressed in a segmentally reiterated subset of neurons in the CNS; in stage 15/16 *sca-gcm* embryos the number of EY expressing cells in the CNS is dramatically reduced. (M,N) In stage 15/16 wild-type embryos, the TEN-M protein is expressed on the axons that make up the longitudinal and commissural tracts of the CNS; this axonal expression of TEN-M is virtually abolished in stage 15/16 *sca-gcm* embryos. (O,P) In stage 15/16 wild-type embryos, the WRAPPER protein is expressed in midline glial cells, in some lateral glial cells and in glial cells supporting the chordotonal sensory organs; this expression has spread to the complete CNS region in stage 15/16 *sca-gcm* embryos. (Q,R) In late stage embryos REPO (brown) is expressed in all and *bnb* (blue) is expressed in a subset of lateral glial cells; in *gcm* mutants REPO expression is reduced to a few cells, and *bnb* expression is completely absent in the CNS. (S,T) In late stage embryos WRAPPER is expressed in midline glial cells, in some lateral glial cells and in glial cells supporting chordotonal sensory organs (arrowheads); in *gcm* mutant embryos WRAPPER expression in lateral glia (CNS) and in chordotonal sensory organs (PNS) is absent whereas expression in midline glial cells remains.

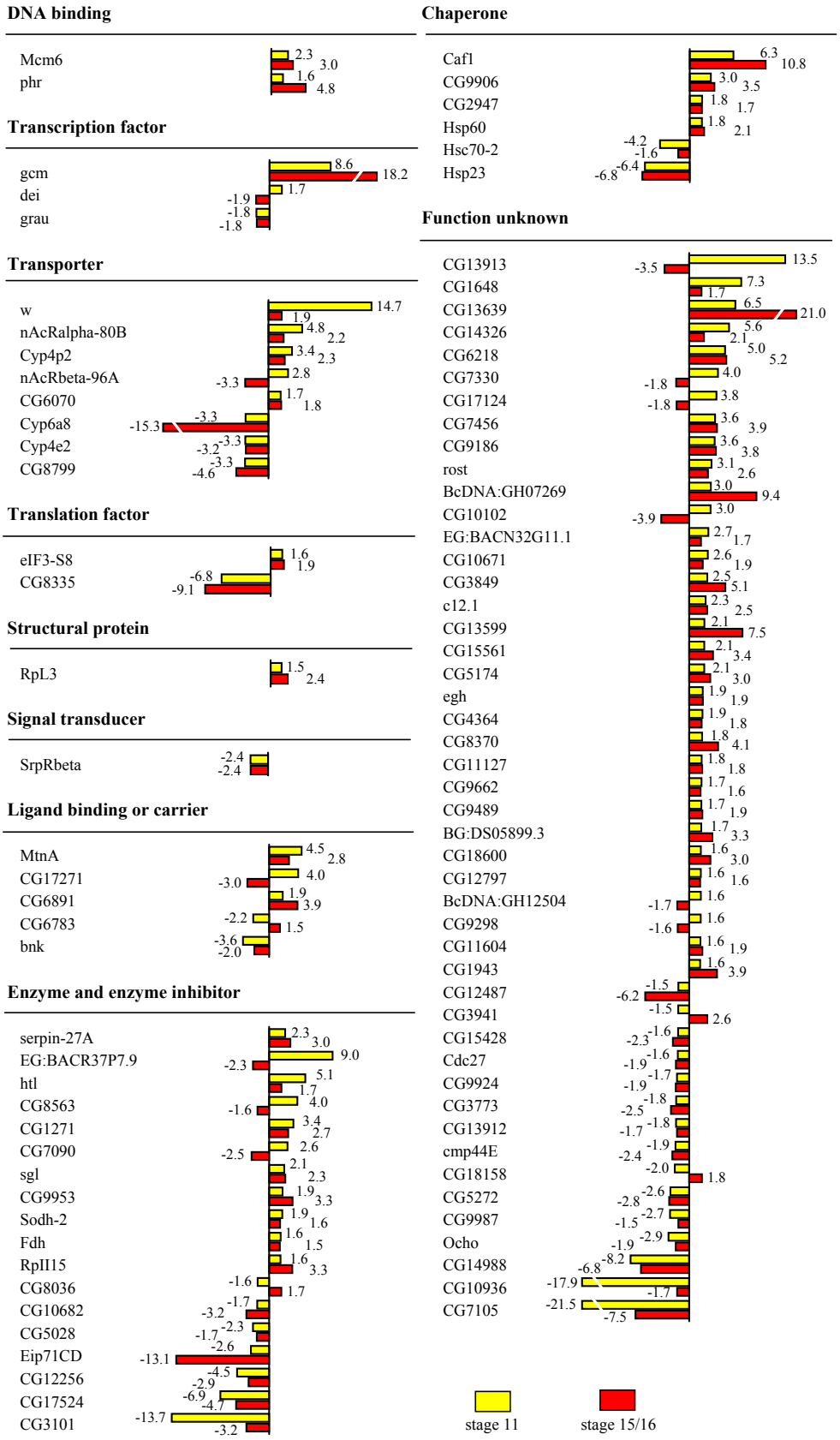



**Fig. 6.** Changes in transcript levels of the genes with differential expression in both early and late embryonic stages following *gcm* misexpression. 93 genes show significant changes in expression levels in response to *gcm* misexpression at stage 11 (yellow) as well as at stage 15/16 (red). Bars represent the fold changes in gene expression levels between wild-type embryos and *sca-gcm* embryos. Positive values indicate that the relative expression level of a gene is increased (upregulation) and negative values indicate a decrease (downregulation).

**Table.1 Differential gene expression in functional classes following *gcm* misexpression**

| Molecular Function | Number of transcripts | |
|---|---|---|
| | stage 11 | stage 15/16 |
| **Nucleic acid binding** | 44 | 96 |
| DNA binding | 33 | 64 |
| Transcription factor | 26 | 38 |
| RNA binding | 6 | 27 |
| Translation factor | 5 | 4 |
| Ribonucleoprotein | / | 1 |
| **Cell cycle regulator** | / | 10 |
| **Chaperone** | 7 | 13 |
| **Motor protein** | 4 | 7 |
| **Defense/immunity protein** | 4 | 3 |
| **Enzyme** | 78 | 249 |
| Kinase/phosphatase | 13 | 59 |
| **Enzyme activator** | / | 3 |
| **Enzyme inhibitor** | 3 | 7 |
| **Apoptosis regulator** | / | 2 |
| **Signal transducer** | 12 | 50 |
| **Cell adhesion** | 4 | 19 |
| **Structural protein** | 3 | 39 |
| **Transporter** | 25 | 51 |
| **Ligand binding or carrier** | 13 | 63 |
| **Antioxidant** | 1 | 2 |
| **Tumor suppressor** | 1 | / |
| **Function unknown** | 218 | 645 |
| **Total** | **417** | **1259** |

## 8. Transcriptional signature of an adult brain tumor in *Drosophila*

Thomas Loop[1], Ronny Leemans[1], Urs Stiefel[1], Leandro Hermida[2], Boris Egger[1], Fukang Xie[1], Michael Primig[2], Ulrich Certa[3], Karl-Friedrich Fischbach[4], Heinrich Reichert[1], and Frank Hirth[1*]

[1]Institute of Zoology, Biocenter/Pharmacenter, University of Basel, Klingelbergstr. 50, CH-4056 Basel, Switzerland; [2]Biocenter, University of Basel, Klingelbergstr. 70, CH-4056 Basel, Switzerland; [3]Roche Genetics Pharmaceuticals Division, F. Hoffmann-La Roche Ltd, CH-4070 Basel, Switzerland; [4]Institute of Biology III, University of Freiburg, Schaenzlestrasse 1, D-79104 Freiburg, Germany

E-mail addresses (in the order of appearance on the manuscript): Thomas.Loop@unibas.ch, Ronny.Leemans@unibas.ch, Urs.Stiefel@unibas.ch, Leandro.Hermida@unibas.ch, Boris.Egger@welc.cam.ac.uk, frankxie@gzsums.edu.cn, Michael.Primig@unibas.ch, Ulrich.Certa@Roche.com, kff@uni-freiburg.de, Heinrich.Reichert@unibas.ch, Frank.Hirth@unibas.ch

*Corresponding author:
Dr. Frank Hirth, Institute of Zoology, Biocenter/Pharmacenter, University of Basel, Klingelbergstr. 50, CH-4056 Basel; Tel.: 0041-61-2671617, Fax: 0041-61-2671613, e-mail: Frank.Hirth@unibas.ch

## Abstract

### Background

Mutations and gene expression alterations in brain tumors have been extensively investigated, however the causes of brain tumorigenesis are largely unknown. Animal models are necessary to correlate altered transcriptional activity and tumor phenotype and to better understand how these alterations cause malignant growth. In order to gain insights into the *in vivo* transcriptional activity associated with a brain tumor, we carried out genome-wide microarray expression analyses of an adult brain tumor in *Drosophila* caused by homozygous mutation in the tumor suppressor gene *brain tumor (brat)*.

### Results

Two independent genome-wide gene expression studies using two different oligonucleotide microarray platforms were used to compare the transcriptome of adult wildtype flies with mutants displaying the adult $brat^{k06028}$ mutant brain tumor. Cross-validation and stringent statistical criteria identified a core transcriptional signature of $brat^{k06028}$ neoplastic tissue. We find significant expression level changes for 321 annotated genes associated with the adult neoplastic $brat^{k06028}$ tissue indicating elevated and aberrant metabolic and cell cycle activity, upregulation of the basal transcriptional machinery, as well as elevated and aberrant activity of ribosome synthesis and translation control. One fifth of these genes show homology to known mammalian genes involved in cancer formation.

### Conclusion

Our results identify for the first time the genome-wide transcriptional alterations associated with an adult brain tumor in *Drosophila* and reveal insights into the possible mechanisms of tumor formation caused by homozygous mutation of the translational repressor *brat*.

## Background

Cancer is a multistep process, which involves loss of cell proliferation control, resistance to cell death, and invasion as well as metastasis. This process is often associated with multiple and progressive genetic alterations. Proto-oncogenes are activated in a dominant fashion by mutation, chromosome translocation or gene amplification, whereas tumor suppressor genes are inactivated or lost by mutation, chromosome loss, mitotic recombination or gene conversion [1-3].

Mutations and gene expression alterations in brain tumors have been extensively investigated. For example, data has been accumulated for the cerebellar tumor medulloblastoma which is the most common malignant brain tumor in children [4]. In addition to chromosome loss, medulloblastomas overexpress certain genes, including c-myc, pax5, and zic, all of which encode transcription factors [5-7]. Thus, animal models of brain

tumors are required to gain insights into the correlation of genetic alterations and tumor phenotype and to better understand how genetic alterations cause malignant neoplasm [4, 8]. To better understand the genetic alterations associated with brain tumors *in vivo*, the fruitfly *Drosophila melanogaster* is an excellent model system for which unprecedented genetic and transgenic technologies as well as entire genome sequence information are available [9, 10]. Indeed, more than two thirds of the known human cancer genes are thought to have *Drosophila* homologs [11, 12], and various studies have shown that genetic inactivation or misregulation of *Drosophila* oncogenes or tumor suppressor genes lead to neoplasms that display characteristic features of malignant growth like in human cancer [13-15]. The level of gene and pathway conservation, the similarity of cellular processes, and the emerging evidence of functional conservation of *Drosophila* genes that are homologous to mammalian oncogenes or tumor suppressor genes, indicate that studies of tumorigenesis in flies can directly contribute to the understanding of human brain tumors [16, 17].

One of the earlier studies on mutations causing malignant neoplasms in Drosophila examined 12 recessive lethal mutations affecting the optic neuroblasts and ganglion mother cells in the larval brain, the imaginal discs, and the hematopoietic organs [13]. Among others, it could be shown, that a mutation in the lethal (2) giant larvae (lgl) gene leads to a failure of the optic neuroblasts of the inner and outer formation centers and the ganglion mother cells (GMCs) to generate adult neurons. The neuroblasts and GMCs proliferated extensively, causing a dramatic enlargement of the larval brain hemispheres. Upon transplantation of neoplastic tissue into wildtype hosts, the transplanted tissues invaded the body cavities, the abdomen, the thorax, the ovaries, the gut, and the thoracic muscles [13].

Further investigation of mutations in the dlg1 and lgl genes revealed that the resulting phenotypes exhibited most of the characteristics defining neoplastic growth in vertebrates, including rapid growth both in situ and in culture, defective intercellular junctions, low adhesiveness in between cells, defective intercellular communication, absence of terminal differentiation, lack of response to ecdysone, invasiveness, and lethality to the host ([13], reviewed in [18]).

Furthermore, it is known that expression of an activated form of RAS1 (RAS1V12) in Drosophila imaginal discs is leading to ectopic cell proliferation and hyperplastic tissue growth [19]. The mutational activation of the human homolog of Ras is associated with a wide variety of human tumors (reviewed in [20]). When RASV12 expression is combined with inactivation of scribbled (scrib), lethal (2) giant larvae (lgl), discs large 1 (dlg1), bazooka (baz), stardust(sdt) and Cdc42 in Drosophila larvae, migration, invasion of tumor tissue and secondary tumor foundation can be found [15]. Another Dosophila gene possessing tumor suppressor properties is warts (wts). Flies homozygous for strong wts alleles show overproliferation in the imaginal discs and the central nervous system leading to pupal lethality. The observed cellular clones where irregular in shape, unpatterned and formed folded lobes. The cells to be found in wts mutant clones were larger than neighboring wildtype cells [21]. The mammalian homolog of wts, LATS1, is also known for causing soft-tissue sarcomas and ovarian stromal cell tumors when mutated in mice [22]. The current model of wts / LATS1 function suggests interaction with CyclinA / Cdc2, linking it directly to the cell cycle ([23], reviewed in [24]).

Mutations in the Drosophila gene salvador (sav) have been associated with increased cell growth and an increase in cell number. Also, Sav has been suggested to interact directly with Wts and thereby contributing to its function [25].

In this report, we have determined at the full genome level the transcriptional signature of an adult brain tumor in *Drosophila* caused by homozygous mutation in the tumor suppressor gene *brain tumor* (*brat*). This gene encodes a member of the conserved NHL family of proteins, which appear to regulate differentiation and growth. These factors have been directly implicated in human tumor formation [26, 27]. Based on the analysis of loss- and gain-of-function experiments, Brat acts in translation repression, ribosomal RNA synthesis, and negative regulation of cell growth [28-30]. Structural analysis indicates that the NHL beta-propeller is the essential domain for both the translational repression and cell growth inhibitory activities of Brat [31]. Inactivation of *brat* results in neoplastic overgrowth and tumor formation in the larval brain and generally causes lethality in the larval third instar and pupal stages [32]. It has been hypothesized that in brat loss-of-function larvae the optic neuroblasts and ganglion mother cells are incapable of generating neurons and repesent the source of the brain tumor [33]. Tumor cells derived from homozygous *brat* larval brain tissue can grow not only *in situ*, but also after transplantation into adult host flies. The transplanted cells grow rapidly, forming metastases and secondary malignant tumors, and finally kill their host [34].

Here we focus on the strong neoplastic adult brain phenotype caused by homozygous mutation of the $brat^{k06028}$ allele. To analyze the transcriptional activity associated with this brain tumor, we carried out two independent genome-wide microarray expression studies using two different oligonucleotide array platforms to compare adult wildtype flies with flies displaying the adult $brat^{k06028}$ mutant brain tumor. Our experiments identified 321 annotated genes that showed highly significant ($p<0.0001$) changes in expression levels due to loss of the tumor suppressor activity of *brat*. One fifth of these genes show homology to known mammalian genes involved in cancer formation. We observed significant upregulation of genes involved in the control of asymmetric neuroblast division. We also found significant upregulation of genes involved in ribosome biogenesis, translation, and RNA processing, suggesting a link between the tumor suppressor activity of *brat* and its role as a translational repressor during *Drosophila* development.

## Results

### Adult homozygous $brat^{k06028}$ mutants show a strong neoplastic brain phenotype

The brat mutant allele $brat^{k06028}$ was generated by a Plac W transposon insertion in the non-coding Exon 4 of the *brat* locus lying immediately upstream of the transcriptional start site [28, 32]. Larvae homozygous for this insertion manifest a strong neoplastic brain phenotype.

We found that 15% (n=979) of homozygous *brat*$^{k06028}$ mutants eclose. These homozygous adult *brat*$^{k06028}$ mutants show limited motility compared to Oregon R wildtype flies and their lifespan is strongly reduced; most of the homozygous *brat*$^{k06028}$ mutants die after the first 10 days. Histological analysis showed that homozygous adult *brat*$^{k06028}$ mutants display a pronounced overproliferation in their brains as compared to Oregon R wildtype brains (Figure 1). The penetrance of this phenotype is 100% (n = 300). Most of the newly hatched *brat*$^{k06028}$ mutants showed a head capsule completely filled with neoplastic tissue. In addition, we observed ectopic neuropil-like structures; these structures appeared bilaterally in ventro-medial positions dorsal to the central brain complex and more ventro-laterally to the optic lobes (Figure 1B, arrowheads). In paraffin sections the size of the central brain structures and the optic lobes of *brat*$^{k06028}$ mutants appeared reduced as compared to the wildtype, however, the overall morphology of neuropil subdivisions in these brains was retained.

## Transcriptional signature of the adult *brat*$^{k06028}$ mutant brain tumor

In order to gain insights into the genetic alterations associated with a fully developed brain tumor *in vivo*, we determined at the full genome level the transcriptome of the neoplastic brain tumors caused by homozygous *brat*$^{k06028}$ mutation as compared to wildtype brain. To ensure cross-validation and increased significance of our results, we conducted two independent genome-wide gene expression studies to compare wildtype flies with homozygous *brat*$^{k06028}$ flies using two different oligonucleotide array platforms (Table1). In the first experiment, RNA was extracted from homozygous adult *brat*$^{k06028}$ fly heads and from isolated adult fly heads of Oregon R as control. The labeled cRNA of this experiment was hybridized to custom made full-genome GeneChips (roDROMEGa, experiment A). In the second experiment, RNA was extracted from dissected adult brains, and hybridization of labeled cRNA to commercially available full-genome Gene Chips involved a signal amplification step (DrosGenome1, experiment B; see materials and methods). Moreover, in this second experiment, the transcriptome of dissected adult brains of homozygous *brat*$^{k06028}$ mutants was compared to dissected adult brains of flies generated by transposon excision of the *brat*$^{k06028}$ P-element (termed *brat*$^{k06028}$ *jumpout*, see materials and methods). Precise transposon excision resulted in reversion of the *brat*$^{k06028}$ neoplastic brain phenotype to wildtype-like brain (data not shown).

Stringent quality control and filtering for the two resulting data sets was done independently (see materials and methods). After analysis, 1778 transcripts in the wildtype condition and 2955 transcripts in the *brat*$^{k06028}$ condition of experiment A were statistically judged as present. This represents 12.7% and 21% of all transcripts on the roDROMEGa array, respectively. In experiment B, 5063 transcripts were judged as present in the wildtype-like *brat*$^{k06028}$ *jumpout* condition and 4981 in the *brat*$^{k06028}$ condition. This represents 36.3% and 35.7% of all transcripts on the DrosGenome1 array, respectively. Transcripts in both datasets were considered as differentially expressed when the change between conditions was larger than 2-fold, their signal strength was above 10 in the condition with the higher expression

level, and the change had a significance value of $p \leq 0.0001$ (t-test). In experiment A 725 transcripts were judged as differentially expressed, whereas in experiment B 1888 transcripts were judged as differentially expressed in the *brat*$^{k06028}$ tumor condition (Table1).

## The core set of genes differentially expressed in *brat*$^{k06028}$ neoplasms

To strengthen the significance of our data, we considered those genes for further analysis that passed the filter criteria in both experiments. This was the case for 321 transcripts, representing a core transcriptional signature that characterizes the neoplastic tissue of adult homozygous *brat*$^{k06028}$ mutants in a highly reproducible manner (Table 1). In both experiments, the majority of the 321 differentially regulated genes (279) showed an increased expression level in neoplastic adult *brat*$^{k06028}$ mutants. 42 transcripts out of 321 differentially regulated genes showed a decreased expression level in the *brat*$^{k06028}$ condition. (5 transcripts out of the 321 genes showed an opposite differential expression change in experiment A and B). As shown in Figure 2, the majority of the 321 genes differentially regulated in *brat*$^{k06028}$ neoplasms displayed comparable expression level changes in both experiments. Only 88 genes showed significantly higher expression level changes in experiment B as compared to experiment A. This may be due to the fact that in experiment B cRNA hybridization to the array involved a signal amplification step (see materials and methods).

The 321 genes differentially expressed in both experiments were grouped into functional classes according to their annotation in *Flybase*. The function of most of the differentially expressed transcripts were unknown (n=67). A list of these genes is shown in Figure 8. The remaining genes of known function were grouped into the functional classes *metabolism* (n=56), *cytoskeleton/structural proteins* (n=31), *RNA binding/processing* (n=27), *transcription/replication/repair* (n=20), *translation* (n=19), *cell cycle* (n=18), *transcriptional regulation* (n=17) *chromatin structure* (n=13), *signal transduction* (n=12), *chaperones* (n=11), *transport* (n=10), *proteolytic systems* (n=8), *stress response* (n=7), *cell surface receptors/CAMs/ion channels* (n=4), and *apoptosis* (n=1). You can find a list of these 321 genes along with the fold change between experiments in Supplementary Table 3 [see Additional file 3].

## Genes involved in metabolism, cell cycle and apoptosis

Neoplastic overgrowth involves essential alterations in cell physiology, loss of cell proliferation control and resistance to cell death. Accordingly, we expected to observe the differential expression of genes linked to metabolism, cell cycle regulation and apoptosis in our gene expression profile studies. Indeed, transcript profiling of *brat*$^{k06028}$ mutants identified transcripts of this type of which misregulation appears to be associated with malignant growth.

Among the genes involved in *metabolism*, we identified 41 transcripts with elevated expression levels (Figure 3). These genes included *eyes absent* (*eya)***,** whose activity is required for the survival of progenitor cells at a critical stage in eye morphogenesis [35] as well as *uninitiated* (*und),* the *Drosophila* methionine aminopeptidase 2, which was shown to be a putative regulator of translation initiation required for cell growth and patterning [36]. We also detected *selenide, water dikinase* (*SelD)* as differentially upregulated in $brat^{k06028}$ neoplasms. *SelD* encodes a product involved in selenocysteine biosynthesis, and mutational inactivation in *SelD* results in a reduction in cell proliferation in the imaginal discs and the larval brain [37]. Within the functional class *metabolism*, we also found 15 transcripts as significantly down-regulated in the $brat^{k06028}$ tumor condition (Figure 3). Among these were *bubblegum* (*bgm)*, encoding a product with long-chain-fatty-acid-CoA-ligase activity. Mutant analysis of *bgm* revealed that in young homozygous mutant flies the optic lobes appear normal however as the flies grow older the optic lobes show signs of regional degeneration [38].

As expected, all of the transcripts positively regulating cell cycle showed increased expression levels in the tumor condition as compared to the wildtype situation. Among these we identified *cyclin A*, *cyclin B*, *cyclin E*, *string (stg), cdc2*, *Cdk7*, *Cks*, *Set*, *abnormal spindle (asp), fizzy (fzy), polo,* and *Myb oncogene-like (Myb).* These genes are involved in the regulation of highly conserved aspects of cell proliferation such as progression through G1, S and G2/M phases, spindle orientation and the maintenance of genomic integrity throughout mitosis.

Of particular interest is the highly significant upregulation of the genes *asp*, *fzy*, *rough deal (rod),* and *Myb*. *asp* encodes a microtubule-associated protein that associates with the polar regions of the mitotic spindle. Mutants of *asp* are larval lethal, with a high frequency of aberrant (e.g. polyploid) cells arrested in metaphase in the larval brain, suggesting that *asp* may play a role in spindle pole organization during mitosis [39]. *fzy* encodes a product involved in cyclin catabolism and *fzy* mutants also show metaphase arrest with compact condensed chromosomes like *asp* mutants [40]. *rod* encodes a product involved in mitotic chromosome segregation which is localised to the kinetochore, and mutations in *rod* result in mitotic segregational failure due to delayed or incomplete release of sister chromatids [41]. Finally *Myb* encodes a proto-oncogene with transcriptional activator activity involved in centrosome cycle which is required to sustain the appropriate rate of proliferation, to suppress formation of supernumerary centrosomes, and to maintain genomic integrity [42].

It is noteworthy that the activity of all of these genes involved in cell cycle regulation has been linked to several aspects of neurogenesis [43, 44]. Thus, all of these genes are known to be expressed during embryonic and larval development in mitotically active cells of the nervous system, the neuroblasts and ganglion mother cells (GMCs), suggesting that their elevated levels of transcriptional activity correlate with an elevated and aberrant activity of the cell cycle machinery in neoplasms of homozygous $brat^{k06028}$ mutants.

We did not detect any significant alteration of gene expression concerning transcripts belonging to the functional class *apoptosis*, except elevated expression levels for *thioredoxin peroxidase 2* (*Jafrac2)*, which is involved in the induction of apoptosis [45]. The fact that we only identified one transcript involved in programmed cell death suggests that the apoptotic program is either impaired below detectability or is unaltered in neoplastic tissue of adult $brat^{k06028}$ mutant brain tumors.

**Genes involved in translation and RNA binding/processing**

Recent studies have shown that *brat* acts as a translational repressor [29]. Moreover, mutant analyses demonstrated that *brat*-mutant cells are larger than wildtype cells and have enlarged nucleoli, which are associated with an increase in total rRNA production. Also, the C. elegans homolog of brat, ncl-1, is known for being a repressor of RNA polymerase I and III transcription and for being an inhibitor of cell growth. Loss of function mutations in ncl-1, result in enlarged nucleoli. The rates of rRNA and 5S RNA transcription are increased and cells are enlarged. [46]. These data suggest that excess ribosomal synthesis and cell growth may be important aspects of the tumorous phenotype of $brat^{k06028}$ mutants [30]. Indeed, in addition to an elevated and aberrant activity of the cell cycle machinery, our gene expression profile of $brat^{k06028}$ adult brain tumors detected genes involved in all aspects of ribosome synthesis and translation control. Thus, transcript profiling of $brat^{k06028}$ mutants identified 19 transcripts as differentially expressed belonging to the functional class *translation* and 27 transcripts as differentially expressed in the functional class *RNA binding/processing*. All of these transcripts showed increased expression level in the tumor condition as compared to the wildtype situation (Figure 4).

Among the transcripts that belong to the functional class *translation* we identified *Eukaryotic initiation factor 1A* (*eIF-1A), Eukaryotic initiation factor 3 p40 subunit* (*eIF-3p40*), and *Int6 homologue* (*Int6)* that encode proteins with translation initiation factor activity involved in protein biosynthesis [47]. Moreover, genes involved in translation elongation factor activity like *Elongation factor 1α48D* (*Ef1α48), Ef1γ*, and *eEF1δ* were detected as significantly upregulated in the tumor condition. In addition, genes involved in 35S primary transcript processing (*Fibrillarin*, *Fib)* and in rRNA processing (*Nop5*) were identified as upregulated. Interestingly, we also detected upregulation of *Ribosomal protein L1* (*RpL)* and *Ribosomal protein L3* (*RpL3),* both of which encode structural constituent of the ribosome involved in protein biosynthesis. *RpL3* is of particular interest as it has been mapped to the chromosomal region harbouring the *M(3) 86D* Minute mutation [48].

Among the transcripts that belong to the functional class *RNA binding/processing* we identified *Rrp4* and *Csl4,* both encoding gene products with 3'-5' exoribonuclease activity involved in mRNA processing [49]. Also *Developmental embryonic B* (*DebB)* and *SC35* with pre-mRNA splicing factor activity involved in mRNA splicing were detected as significantly upregulated [50]. In addition, *Rm62* encoding a product with ATP dependent RNA helicase activity involved in RNA interference was upregulated in the tumor condition, as were *U2*

*small nuclear riboprotein auxiliary factor 50* (*U2af50)* and *U2 small nuclear riboprotein auxiliary factor 38* (*U2af38),* both encoding products with pre-mRNA splicing factor activity involved in mRNA splicing (reviewed in [47]). Within this functional class we also identified upregulated transcripts which are involved in various aspects of cell growth and proliferation, such as *polyA-binding protein* (*pAbp)* which encodes a product with poly(A) binding involved in positive regulation of translation. *pAbp* mutants display cytokinesis defects, and mutant analyses demonstrated that *pAbp* is required to connect the centrosome to the minus-ends of free microtubules (reviewed in [47]). Intriguingly, we also detected *Small ribonucleoprotein Sm D3* (*SmD3),* which encodes a protein involved in mRNA splicing, as upregulated in the brain tumor condition. This is of particular interest since homozygous 12-day old *SmD3* mutant larvae show overgrowth of the brain, hematopoietic organs and imaginal discs and die during third instar larval or pupal stages [51]. Finally, we also identified *staufen* (*stau)*, which has been shown to be involved in the asymmetric localization of cell fate determinants during neuroblast divisions [52], as upregulated in neoplastic tissue.

## Genes involved in chromatin structure, transcription/replication/repair, and transcriptional regulation

In addition to the expected upregulation of genes involved in cell cycle regulation as well as ribosome synthesis and RNA processing, we also observed that the basal transcriptional machinery was significantly upregulated in the $brat^{k06028}$ tumor condition (Figure 5). Within the functional class *chromatin structure* we found 13 genes with elevated expression levels. Among those are genes involved in various aspects of chromatin modeling such as *Nucleoplasmin* (*Nlp)*, involved in nucleosome spacing, and *Nucleosome assembly protein 1* (*Nap1)* as well as *Chromatin assembly factor 1 subunit* (*Caf1)***,** both encoding proteins with histone-specific chaperone activity [53]. We also detected two suppressors of variegation, namely *Suppressor of variegation 3-9* (*Su(var)3-9)* with histone methyltransferase activity [54] and *Suppressor of variegation 205* (*Su(var)205)* involved in the establishment of chromatin silencing. Interestingly, mutant analysis of *Su(var)205* revealed that larval brain neuroblasts show cytokinesis defects during mitosis [54]; this is also the case for *gluon* (*glu*), which we detected as significantly upregulated in the tumor condition. In *glu* mutants, larval brain neuroblasts show distinct abnormalities during chromosome segregation and lethality occurs during the late larval stage [55].

In the functional class *transcription/replication/repair* we identified 20 genes as differentially regulated in the $brat^{k06028}$ tumor condition, all of them having elevated expression levels as compared to wildtype. Among them are the *minichromosome maintenance (MCM)* genes, which encode an evolutionary conserved family of molecules that form an important part of the pre-replicative complex, required for DNA replication [56]. We found two members of this complex (*mcm2* and *mcm3*) as upregulated in $brat^{k06028}$ neoplasms. Interestingly, *Mcm2* mutants persist as third instar larvae for several days, and dissection of wandering larvae demonstrates that their CNS is smaller than that of wildtype and that they lack identifiable

imaginal discs [57]. We also detected *mutagen-sensitive 209* (*mus209)***,** involved in mitotic spindle assembly as upregulated. Mutants of *mus209* show abnormal chromosome condensation [58], whereas mutations of *rad50,* a gene involved in DNA repair, result in elevated levels of anaphase bridges in dividing cells of third instar larval brain and imaginal discs. Moreover, *RNA polymerase II 15kD subunit* (*RpII15)*, involved in transcription from Pol II promoter [59] and *absent, small or homeotic disc 2* (*ash2*), a member of the trithorax group were detected as upregulated. Interestingly, mutant alleles for ash2 are larval pupal lethals and display imaginal disc and brain abnormalities [60].

Among the transcripts belonging to the functional class *transcriptional regulation,* we identified two members of the *snail* family of zinc-finger transcription factors, namely *worniu (wor)* and *snail (sna)* as significantly upregulated in $brat^{k06028}$ neoplasms. During wildtype embryogenesis, both genes are expressed during neuroblast delamination, and mutant analyses suggest that Snail and Worniu function in neuroblasts, around the time of division to give rise to ganglion mother cells (GMCs) [61]. Moreover, we also detected upregulation of several other transcription factors involved in the control of proliferative activity of embryonic and postembryonic neuroblasts. Thus, we identified *tailless (tll), deadpan (dpn*), and *castor* (*cas*) as upregulated. *tll* is expressed in procephalic neuroblasts and required for their specification as *tll* mutants lack the anterior brain [62]. *dpn* encodes a pan-neural gene shown to be expressed in all neuronal lineages during embryogenesis as well as in neuroblasts of the larval CNS and in precursors of sensory neurons in imaginal discs. Loss- and gain-of-function experiments suggest that *dpn* activity is critical for the proper regulation of cell proliferation in the larval brain [63]. *cas* is involved in a cascade of sequentially expressed transcription factors, which translates information on timing of GMC formation [64]. In addition, *cas* is known to be involved in post-embryonic brain development as small homozygous somatic mutant *cas* clones in the adult brain lead to ellipsoid body and mushroom body defects [65]. Finally, we also detected *diminutive* (*dm)*, the *Drosophila* homolog of the proto-oncogene c-myc, as significantly upregulated in the $brat^{k06028}$ tumor condition. *Myc* genes link patterning signals to cell division by regulating primary targets involved in cellular growth and metabolism [66].

**Genes involved in cytoskeleton/structural and signal transduction**

In previous studies, histological analysis of $brat^{k06028}$ mutants indicated that the brain tumor phenotype is primarily due to uncontrolled divisions of optic lobe neuroblasts and GMCs [33]. Our microarray expression analysis supports the assumption that neuroblasts and/or GMCs are at the cellular origin of these brain tumors. This is not only evident by the significant upregulation of genes involved in neuroblast delamination and proliferation such as *wor*, *sna, tll, dpn,* and *cas,* but also by the elevated transcription levels of specific genes belonging to the functional classes *cytoskeleton/structural* and *signal transduction* (Figure 6).

In the functional class *cytoskeleton/structural,* we found high levels of transcription in $brat^{k06028}$ tumorous tissue for the genes *miranda* (*mira)* and *inscuteable (insc)*. This is of

particular interest, since both proteins play essential roles in the asymmetric localization of cell fate determinants during neuroblast divisions. Miranda creates intrinsic differences between sibling cells by mediating the asymmetric segregation of transcription factor Prospero into only one daughter cell during neural stem-cell division. In *insc* mutants, mitotic spindles in neuroblasts fail to rotate into proper position and neuroblasts divide in random orientation resulting in general disorganization of the neuroblast array and defective neuroblast morphology (reviewed in [67]). Moreover, we also identified as significantly upregulated in the tumor condition *ciboulot (cib),* an actin binding protein, and the *Drosophila* profilin protein *chickadee (chic),* both of which cooperate in central brain metamorphosis [68]. Intriguingly, expression of *big brain* (*bib)* appears to be downregulated in the $brat^{k06028}$ tumor condition. *bib* encodes a product with connexon channel activity, which is required autonomously in epidermal precursors to prevent neural development (reviewed in [69]).

Among those transcripts belonging to the functional class *signal transduction*, we identified several members of GTPase signal transduction pathways as significantly upregulated in $brat^{k06028}$ neoplasms. Thus, we identified *RacGAP50C,* which has GTPase activator activity and is involved in Rho protein signal transduction, and also three members of the Ran GTPase signal transduction pathway, namely *Segregation distorter* (*Sd), ran,* and *moleskin* (*msk)*. Moleskin, for example, appears to be involved in EGF receptor signalling pathway initiation of gene expression in response to *Drosophila* receptor tyrosine kinase signalling [70]. We also detected *MAP kinase activated protein-kinase-2* (*MAPk-Ak2)*, which has protein serine/threonine kinase activity, as significantly upregulated. In addition, we identified *Notch* (*N*) as significantly upregulated in the $brat^{k06028}$ tumor condition. *Notch* encodes a transmembrane receptor mediating cell-cell communication, and Notch signalling has been implicated in a wide variety of cellular processes, including the maintenance of stem cells, specification of cell fate, differentiation, and proliferation [71]. Surprisingly, however, the *Insulin-related peptide (Ilp2),* a component of the insulin signalling pathway, was downregulated in the tumor condition. Overexpression of Ilp2 alters growth control in a Insulin receptor-dependent manner, suggesting a role for Ilp2 in controlling organismal size by augmenting both cell number and cell size of different organs (reviewed in [66]).

**Genes involved in cell surface receptors/CAMs, chaperones, proteolytic systems, stress response and transport**

Previous studies have shown that tumor cells derived from homozygous *brat* larval brain tissue can grow not only *in situ*, but also after transplantation into adult host flies. The transplanted cells grow rapidly, forming metastases and secondary malignant tumors that finally kill their host [34]. This suggests that loss of *brat* leads to the acquisition of invasive and metastatic properties of the resulting tumor cells. This acquisition of invasive and metastatic abilities is associated with altered binding specificities of cell adhesion molecules (CAMs), the activation of proteases, as well as chaperones and heat shock proteins mediating stress response [1, 72]. In accordance with this, our transcriptional analysis of $brat^{k06028}$ tumor

cells identified genes as differentially regulated in the tumor condition belonging to the functional classes *cell surface receptors/CAMs, chaperones, proteolytic systems,* and *stress response* (Figure 7). Within the class *cell surface receptors/CAMs* we identified *Neurotactin (Nrt)*, a transmembrane protein that localizes to cell-cell junctions where it mediates cell adhesion and cell signalling, as upregulated [73]. Three other genes belonging to the class *cell surface receptors/CAMs* were differentially downregulated in the $brat^{k06028}$ tumor condition. Among the classes *proteolytic systems*, *chaperones* and *stress response*, we identified genes upregulated like *Ubiquitin carboxy-terminal hydrolase* (*Uch)*, involved in protein deubiquitination, and *granny smith* with leucyl aminopeptidase activity. Also the chaperones *T-complex Chaperonin 5 (Cct5)* and *Cctgamma* with chaperonin ATPase activity involved in protein folding and the heat shock proteins *Hsp27*, *Hsp26*, and *Hsc70-4* were identified as significantly upregulated in the $brat^{k06028}$ tumor condition. Finally, we detected 10 transcripts belonging to the functional class *transport* as differentially regulated in neoplastic $brat^{k06028}$ tissue. Among these is *Pendulin* (*Pen)*, an adaptor molecule which is expressed in embryonic neuroblasts and the proliferating regions of the larval brain lobes and may be required for the normal transmission and function of proliferative signals in the cells [74].

## Vertebrate homologs of genes differentially regulated in $brat^{k06028}$ neoplasms are involved in mammalian cancer formation

Our analysis of the 321 genes differentially regulated in $brat^{k06028}$ neoplasms revealed genes involved in general aspects of tumorigenesis such as elevated and aberrant metabolic activity, elevated and aberrant activity of the cell cycle and basal transcriptional machinery, as well as elevated and aberrant activity of ribosome synthesis and translation control. Considering the level of gene and pathway conservation, the similarity of cellular processes, and the emerging evidence of functional conservation of *Drosophila* genes that are homologous to mammalian oncogenes or tumor suppressor genes [16, 17], we wondered whether our data might reveal any parallels to the genetic alterations associated with mammalian cancer formation. This is of particular interest since one of the two human homologs of *brat*, TRIM3, has been mapped to chromosome 11p15 [75, 76], a region that has been termed the "multiple tumor-associated chromosomal region 1" due to the presence of numerous cancer-related genes in this region [77, 78].

We therefore searched available databases in order to identify vertebrate homologs of the 321 genes differentially regulated in $brat^{k06028}$ neoplasms and to determine their possible involvement in mammalian cancer formation. We identified 62 out of 321 *Drosophila* genes as having mammalian homologs, of which misregulation has been associated with various types of tumors (see Table 2). Mammalian homologs of genes involved in the regulation of conserved aspects of cell proliferation were found to be implicated in mammalian cancer formation. Among these are mammalian homologs of cyclins, Cdks, and genes involved in various other aspects of cytokinesis and the maintenance of genomic integrity. Thus, the

*string* homolog *Cdc25B*, encoding a dual-specific phosphatase that mediates cell cycle progression by activating cyclin-dependent kinases, has been shown to possess oncogenic potential and is frequently overexpressed in human prostate cancer tissues [79]. *Drosophila CG9993* encodes for a long-chain fatty acid transporter and its rat homolog (FATP) shows elevated transcript levels in hepatoma cell lines [80]. Homologs of *Drosophila* MCM family of proteins are known as proliferation markers in vertebrates, and evidence exists that HsMCM2 is upregulated in primary human tumor tissue [81]. Furthermore, it has been shown that the rate and level of Mcm3 and MCM4 expression appears to be higher in cancer cells than in normal proliferating cells of the uterine cervix and dysplastic cells.

Prominent examples of genes involved in mammalian cancer formation are the proto-oncogenes *Myb* and *Myc*. The *Myc* gene is a central regulator of proliferation, differentiation, cell survival, and neoplastic transformation. It is found mutated or overexpressed in up to 30% of human cancers and the Myc network appears to regulate a large number of genes, approaching 10% of human genes [82]. Similarly, the human homolog of *Drosophila CG9772,* called $SCF^{Skp2}$, is an ubiquitin ligase encoding onco-protein that is one of the components of the machinery used to control Myc levels through the ubiquitin pathway. Interestingly, this ubiquitin ligase is also required for induction of Myc-responsive genes, suggesting that ubiquitination not only promotes Myc turnover but also "licenses" its transcriptional activity [83]. Examples of homologous genes involved in human brain tumor formation are the *tailless* ortholog Nr2f1 involved in retinoblastoma, and the homologs of *Myb (MYB), dm (MYC)*, and *Notch (NOTCH2)*, whose misregulation has been associated with human gliomas [71, 84].

## Discussion

### A core transcriptional signature of neoplastic brain tissue in *Drosophila*

In order to gain insights into the transcriptional alterations associated with brain tumors *in vivo*, we have determined the transcriptional signature of an adult brain tumor in *Drosophila* caused by homozygous mutation in the tumor suppressor gene *brat*. To this end, we have conducted two independent genome-wide gene expression studies using two different oligonucleotide microarray platforms to compare adult wildtype flies with flies displaying the adult $brat^{k06028}$ mutant brain tumor. Cross-validation of our comparative analysis of adult wildtype heads compared to $brat^{k06028}$ heads versus dissected $brat^{k06028}$ brains compared to dissected adult brains of flies generated by transposon excision of the $brat^{k06028}$ P-element identified a core transcriptional signature of $brat^{k06028}$ neoplastic tissue. The obtained transcriptional signature revealed genetic alterations associated with malignant growth in a highly reproducible manner, despite the fact that both the biological material and the array platforms used differed considerably between the two experiments. This involved the use of stringent filter criteria such as a significance value of $p \leq 0.0001$ (t-test). Due to these very stringent filter criteria, the 321 transcripts detected as differentially regulated in the $brat^{k06028}$ tumor condition probably represent only a subset of genes that could be involved in brain tumorigenesis. For example, it has been shown that PTEN plays a crucial role in the highly conserved insulin/PI3-kinase (PI3K) signaling pathway involved in cell growth and proliferation (for review see [66] and the human homolog of PTEN is frequently lost in gliomas (reviewed in [84]). Although we did detect PTEN as differentially down-regulated in the brat tumor condition in both experiments, it was not incorporated into the core set of genes differentially regulated in $brat^{k06028}$ neoplasms due to p-values below our cutoff value (see supplementary material).

Nevertheless, our genome-wide microarray analysis in *Drosophila* identified characteristic features of malignant growth such as those found in human cancer. Thus, we detected gene expression changes indicative of elevated and aberrant metabolic activity, elevated and aberrant activity of the cell cycle and basal transcriptional machinery, and elevated and aberrant activity of ribosome synthesis and translation control associated with the adult $brat^{k06028}$ mutant brain tumor. Our data are in accordance with assumptions that the vast catalog of cancer cell genotypes is a manifestation of essential alterations in cell physiology that collectively dictate malignant growth: self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed cell death (apoptosis), limitless replicative potential, and tissue invasion and metastasis [1]. This is further exemplified by the fact that for 20% of the genes differentially regulated in the *Drosophila* $brat^{k06028}$ tumor condition it was possible to identify mammalian homologs associated with various types of tumors. These include homologs of *Myb*, *Myc*, cyclins as well as other genes involved in cell growth and proliferation control whose misregulation is frequently associated with mammalian cancer formation and malignant growth [1].

**Brat acts as a tumor suppressor during *Drosophila* brain development**

Brat was originally identified as a growth suppressor of the larval brain of *Drosophila*, with mutant brains growing up to eightfold larger than normal [28, 32]. Loss of *brat* also conferred metastatic potential upon tissues transplanted into wildtype host flies [34]. Moreover, our analysis of the $brat^{k06028}$ allele shows that a substantial amount (15%) of homozygous mutant flies eclose and show a strong neoplastic adult brain phenotype with 100% penetrance. These data demonstrate that Brat functions as a tumor suppressor during brain development of *Drosophila*.

In general, tumor suppressors act as molecular antagonists to the formation of malignancy, a physiological state that is characterized by indefinite cellular growth and division [1]. This in turn requires elevated levels of transcriptional activity of genes involved in all aspects of aberrant cell cycle activity. Indeed, our genome-wide microarray expression analysis of $brat^{k06028}$ neoplastic tissue revealed genetic alterations that can be attributed to malignant growth. We identified essential components of the cell cycle machinery like *cyclin A, B, E, string* and *Myb* as significantly upregulated in the tumor condition. This elevated cell cycle activity is accompanied by significant upregulation of genes involved in DNA replication, chromatin structure and cytokinesis, such as *gluon*, *Mcm2* and *Mcm3*, and the *Drosophila Myc* homolog *diminutive*. Gain of function of these genes is frequently associated with malignant overgrowth and cancer in *Drosophila* [16]. Moreover, de-regulation of the proto-oncogenes *Myb* and *Myc* characterizes one of the hallmarks of cancer, as the oncogenic activity of these genes in turn impairs differentiation and promotes growth [66]. In addition, we also observed significant upregulation of the Ran GTPase pathway since three members of it, namely *Sd, ran,* and *msk*, show elevated expression levels in the $brat^{k06028}$ tumor condition. Recent data suggest that Ran GTPase signalling serves as a positional marker for the eukaryotic genome throughout the cell cycle by regulating microtubule nucleation and nuclear envelope formation around chromatin (reviewed in [85]). This together with the elevated and aberrant activity of several other structural components required for proliferation and cytokinesis emphasises the self-sufficiency of the cellular machinery during unrestrained growth.

**Proliferation control via ribosome synthesis regulation by *brat*?**

Recent results demonstrate that *brat* functions to repress ribosomal RNA synthesis and cell growth. *brat* mutant cells are larger than control cells, have enlarged nucleoli and contain excess rRNA. Furthermore, *brat* overexpression inhibits clone and organ growth, and leads to a decreased level of rRNA per cell [30]. Based on these observations it has been suggested, that in contrast to the cell growth regulation pathway of the insulin receptor and its effectors, *brat* affects cell growth not through the activity of ribosomes, but rather through the regulation of their synthesis [30]. Indeed, disruption of one or more of the steps that control

protein synthesis has been associated with alterations in the cell cycle and regulation of cell growth. For example mutations that inactivate the tumor-suppressor activity of RB or p53, or both, result in aberrant upregulation of essential components in the protein synthesis machinery and increase ribosome biogenesis, leading to enhanced mRNA translation rates (reviewed in [86]). Among the 321 genes differentially regulated due to loss of the tumor suppressor activity of *brat*, we identified several genes involved in ribosome biogenesis and RNA processing as significantly upregulated in neoplastic tissue. One of the most prominent examples is ribosomal protein RpL3 which maps to the chromosomal region harbouring the M(3) 86D Minute mutation [48]. However, although we observe several genes involved in protein synthesis and ribosome biogenesis as differentially upregulated in the $brat^{k06028}$ tumor condition, it remains to be determined whether these phenomena represent the cause or consequence of tumor formation.

### Neuroblast polarity and asymmetric division

Interestingly, we find that *inscuteable* (*insc*), *staufen* (*stau*), *snail* (*sna*), *miranda* (*mira*), and *worniu* (*wor*) have elevated transcript levels in $brat^{k06028}$ brain tissue. Both *insc* and *stau* are involved in asymmetric segregation of cell fate determinants in dividing neuroblasts and *insc* expression is thought to be regulated by *sna* [52, 61]. Miranda is an adaptor protein involved in the distrinution of the Prospero gene product. Prospero is required to distinguish the proliferative capabilities of NBs and GMCs [67]. As it is known that the *Discs large* (*Dlg*), *Scribble* (*Scrib*) and *Lethal giant larvae* (*Lgl*) tumor suppressor proteins regulate multiple aspects of neuroblast asymmetric cell division these findings strengthen the link between tumorigenesis and asymmetry in cell division [87].

## Conclusions

Taken together, our genome-wide microarray expression analysis of an adult brain tumor in *Drosophila* caused by mutation in the tumor suppressor gene *brat* identifies a core set of genes whose deregulation can be attributed to the genetic alterations associated with tumor formation and malignant growth. We detect aberrant and elevated activity in the transcription of genes involved in metabolic activity, cell cycle regulation, cytokinesis, as well as in basal transcriptional regulation. 20% of the genes we identified as differentially regulated in the $brat^{k06028}$ tumor condition have mammalian homologs whose misregulation has been associated with various types of tumors. Moreover, we observed significant alterations in transcriptional activity of genes involved in ribosome biogenesis, translation and RNA processing due to loss of the translational repressor encoded by *brat*. This may indicate that loss of *brat* leads to overexpression of ribosomal proteins and increased ribosome function, which in turn may cause malignant growth. Considering the evolutionary conservation of gene structure and function, we propose that our results obtained by the first genome-wide

expression analysis of an adult brain tumor in *Drosophila* will also be valid for tumor formation in mammals. It will now be important to elucidate at the cellular and molecular level the mechanisms involved in the genetic alterations caused by mutation of the tumor suppressor *brat*, which ultimately lead to tumor formation, metastasis and invasiveness.

## Methods

### Fly strains and genetics

The wildtype was Oregon R. The *brat* allele used in this study is $brat^{k06028}$, balanced over $CyO,P\{w^{+mc}=ActGFP\}JMR1$. Flies used for histology and for transcript profiling were 1 to 2 days old homozygous $P\{w^{+mc}=lacW\},brat^{k06028}$ males and females, identified by the absence of GFP. For the control condition of the oligonucleotide array experiment, either wildtype Oregon R were used or flies generated by transposon excision of the k $brat^{k06028}$ P-element. Excision of the $brat^{k06028}$ transposon resulted in several viable strains, which were further characterized by genomic PCR. Using the primer pair 5'AACAACCAAAACAACGGCAACC 3' and 5' AAACGGAGATAAGCCGACTTAC 3', flanking the insertion site of the $brat^{k06028}$ transposon in the *brat* genomic DNA, a fragment of 211 bp was PCR amplified and sequenced from both strands. For all characterized strains, this sequence was indistinguishable from the genomic *brat* sequence of Oregon R control flies; strain "$brat^{k06028}$ *jumpout*" was chosen for further work. For assessing the percentage of adult escapers in $brat^{k06028}$, homozygous third instar larvae were identified by the absence of GFP, counted (n = 979) and collected into bottles containing standard food. Bottles were then subsequently checked on a daily basis for hatched flies during the following 4 weeks. All flies were kept on standard cornmeal/ yeast/ agar medium at 25°C.

### Histology

For sectioning, adult flies were fixed, dehydrated, embedded in paraffin, and cut into 7µm sections as described by Heisenberg and Böhl [88]. Sections were mounted on coated glass slides with DePeX (Fluka), and neural structures were visualized with a Zeiss Axioskop microscope by autofluorescence (wavelength 488nm) and recorded using a Prog/Res/3008 digital camera (Kontron, Zürich).

### Experimental paradigms and oligonucleotide arrays

Two oligonucleotide array experiments were carried out. In the first experiment, a custom-designed *Drosophila* oligonucleotide array (roDROMEGAa, Affymetrix) was used [89]. This array contains 14,090 sequences representing 13,369 single transcripts encoding *Drosophila* proteins deposited in SWISS-PROT/TrEMBL databases (Celera Genome/BDGP Release

no.1; [10] as well as prokaryotic and custom chosen control sequences. This array was used to compare the transcriptome from Oregon R control heads with heads of homozygous *brat*$^{k06028}$ flies; 5 replicates were done for the wildtype condition and 6 were done for homozygous *brat*$^{k06028}$ (experiment A). In the second experiment, a commercially available Affymetrix *Drosophila* full genome array was used (*Drosophila* Genome Array, Affymetrix) which contains probe sets interrogating more than 13,500 genes from *Drosophila melanogaster* (experiment B). Sequences used to design this array can be found in Flybase (version 1). Over 8,000 of the genes represented have at least one EST/cDNA match. Additionally, approximately 5,500 genes were identified using prediction algorithms. This array was used to compare the transcriptome of dissected *Drosophila* brains derived from *brat*$^{k06028}$ *jumpout* with that of homozygous *brat*$^{k06028}$ brains. 6 replicates were done for each of the two conditions.

**Preparation of biotinylated cRNA.**

Total RNA was isolated from 100 fly heads or 150 fly brains, respectively, using guanidinium isothiocyanate in combination with acidic phenol (pH 4.3) (fast RNA tube green kit from BIO101) in a fast prep homogenizer FP120 (Bio 101). After precipitation, 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'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-(T)24VN-3') was PAGE-gel purified. For first strand cDNA synthesis, a typical 40µl reaction contained 10 –15µg RNA, 200 pmoles T7-(T)24 primer, 500µM 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 340µl, using 20 units *E. coli* DNA ligase, 80 units *E. coli* DNA polymerase I and 4 units RNase H in the presence of 250µM of each dNTP. After 2nd strand cDNA synthesis, 0.5µl RNase A (100mg/ml) (Qiagen) was added and the samples were incubated at 37˚C for half an hour. Thereafter 7.5µl 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 20µl reaction volume contained between 0.3-1.5µg 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 2µl 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

analyzed by agarose gel electrophoresis. Samples were precipitated overnight at -20˚C, taken up in 20µl DEPC treated water and spectrophotometrically quantified. Susequently, the biotinylated antisense cRNA was fragmented by heating the sample to 95˚C for 35 min in a volume of 25µl, containing 40 mM tris-acetate (pH 8.1), 100 mM KOAc, 30 mM MgOAc. After the fragmentation, the samples were placed on ice.

**Hybridization and Scanning**

Gene chips were equilibrated at RT and prehybridized with 280µl hybridization buffer (1 x MES (pH 6.7)/ 1M NaCl/ 20 mM EDTA/ 0.01 % Tween 20) for 15 min at 45 $^0$C with rotation at 60 rpm. Hybridization was done for 16 h at 45 $^0$C (60 rpm) in a final volume of 220µl hybridization buffer, 0.1 mg/ml herring sperm DNA, 0.5 mg/ml acetylated BSA containing 50 pM control oligo B2, spiked bacterial control RNAs (BioB, BioC, BioD, cre) and 15µg fragmented biotinylated cRNA. After hybridization, the arrays were briefly rinsed with 6 x SSPE-T buffer (0.9 M NaCl/ 0.06 M $NaH_2PO_4$/ 6mM EDTA/ 0.01 %Tween 20). Washing and staining were carried out on a Fluidics station 400 (Affymetrix, MAS 5.0) using 100 mM MES, 0.1M NaCl, 0,01 % Tween 20 as stringent wash buffer. Two different staining protocols were used. When comparing the transcriptome of Oregon R fly heads and homozygous *brat*$^{k06028}$ fly heads, staining was carried out in 220µl 1 x MES buffer, 2.0 mg/ml acetylated BSA, 10 ng/ ml StreptavidinR-phycoerythrin conjugate (Molecular Probes) at 40 $^0$C for 15 min. In the second experiment comparing dissected brains of *brat*$^{k06028}$ *jumpout* flies with homozygous *brat*$^{k06028}$ brains, an additional antibody amplification step was included. Prior to staining, 220µl 1 x MES buffer containing 2 mg/ml acetylated BSA and 2µg/ ml biotinylated anti-streptavidin antibody (Vector) was applied onto the array for 30 min at 40 $^0$C. Following a washing step, staining was then performed using streptavidinR-phycoerythrin as described above. Following a final washing step, arrays were scanned with a commercial confocal laser scanner (Agilent).

**Data analysis**

The single arrays were analysed using Microarray Suite 5.0 (Affymetrix). Detailed analysis was carried out using Race-A (Roche), Excel2000 and Access2000 software (Microsoft), and GeneSpring (Silicon Genetics). In this process, the mean of all signal values of each array were set to 1, and all other signal values were expressed relative to it (normalization). A Nalimov outlier test was performed [90]. The data was grouped into control- and experimental conditions. For every gene in all 4 conditions a single mean signal value over the replicates was calculated. Also an unpaired t-test was performed for every gene per experiment to assess significance of change between conditions. The annotation of the two different arrays was linked (see below) using information from NetAffx [91] and Flybase [92]. Genes were judged as present when their presence attribute in the RACE-A software was above 0.75. Both datasets were filtered using a cutoff on the mean signal (≥ 10 in the condition the gene had to be present in), the changefold (≥ 2) and the t-test (≤ 0.0001). The

various control sequences present on the arrays were excluded from analysis. Subsequently the overlap of the two datasets was calculated using the database mentioned above. When more than one probe set was representing a gene the probe set with the highest statistical significance as judged by the t-test was chosen. In those cases were the p-value was identical, the probe set with the higher expression levels was chosen.
The raw data of the two experiments can be found in Supplementary Table 1 and Supplementary Table 2 [see Additional files 1 and 2].

**Linking the two different oligonucleotide arrays**

Using the probe set annotation information provided for the roDROMEGa array, we determined which probe sets on the Affymetrix "DrosGenome1" GeneChip were analogous to those on the Roche custom chip. To this end, a database was built with the following information: 1) Affymetrix "DrosGenome1" probe set annotations, 2) roDROMEGa annotations, and 3) FlyBase gene annotations.

The Affymetrix "DrosGenome1" annotations were extracted from the June 2003 DrosGenome1_annot_csv.zip file obtained from the NetAffx [91] Affymetrix web site (http://www.affymetrix.com/analysis/download_center.affx). The database tables contained probe set ID, gene symbol, FlyBase FBgns, and synonyms as annotated by Affymetrix for each of the 14,010 probe sets on DrosGenome1. The roDROMEGa array annotations were extracted from text dump files obtained from Roche. The database table contained probe set ID and gene symbol information for each of the 14,090 probe sets on the Roche custom chip. FlyBase gene annotations were extracted from the 29-08-2003 FlyBase genes dump file located at ftp://flybase.org/flybase/genes/genes.txt. The resulting database tables contained gene symbol, FBgns, and synonyms for all *43,177* annotations in FlyBase. During extraction of Affymetrix DrosGenome1 and FlyBase annotations from their source files, special cleanup was required to correct from HTML escape characters commonly used in *Drosophila* melanogaster annotation dump files. This was necessary because the annotations for the roDROMEGa array contained only fully spelled out words for gene symbols (i.e. 'alphaCop' instead of '&agr;Cop').

The 14,090 probe sets on the Roche custom *Drosophila* microarray represented 13,343 distinct genes which were mapped to Affymetrix's DrosGenome1. Utilizing the database described above, we first mapped gene symbols annotated by Roche on their custom chip to gene symbols and synonyms annotated by Affymetrix on their DrosGenome1 chip. This search yielded 13,377 probe sets (12,784 genes) on the Roche *Drosophila* custom chip. The remaining 713 probe sets on the Roche custom chip which had no direct gene symbol or synonym match to DrosGenome1 were used in a three-way search using FlyBase genes data. These 713 Roche gene symbols were searched against all FlyBase gene symbols and synonyms and in case of a match all the corresponding FlyBase FBgns for these genes were used to search the Affymetrix FBgn annotations on DrosGenome1. This search resulted in

additional 449 probe sets (401 genes) on the Roche *Drosophila* custom chip which mapped to analogous probe sets on DrosGenome1. Of the remaining 172 unmapped Roche probe sets, 92 were found to be either Affymetrix AFFX controls, Roche internal controls, or rRNA controls. Thus, 13,826 Roche probe sets (13,185 genes) have at least one analogous Affymetrix DrosGenome1 probe set. 172 probe sets (157 genes) on the Roche *Drosophila* custom chip dolfed our parameters and remained orphaned with no Affymetrix DrosGenome1 match. A list with the orphaned probe sets can be found in Supplementay Table 4 [see Additional file 4]

**Functional classification and links to mammalian homologs involved in cancer**

The 321 genes differentially regulated in both experiments were grouped into functional classes by reviewing their annotation in *Flybase*. When there were vertebrate homologs annotated in *Flybase* we searched the *PubMed* bibliographic database (http://www.ncbi.nlm.nih.gov/PubMed/medline.html) for literature that linked the gene in question to cancer in vertebrates.

## Authors' contributions

TL carried out the the array study of experiment B, performed the histology of flyheads, did the bioinformatics and statistics for both chip experiments and drafted the manuscript. RL carried out the array study of experiment A and performed the jumpout of the $brat^{k06028}$ P-element. US dissected the flies that were used as a tissue source for the RNA extractions. LH carried out the bioinformatics for linking the annotations of the two different array platforms used in our study. BE balanced the fly strains. FX was initially involved in setting up experimental conditions and carried out pilot experiments. MP, UC, KFF and HR participated in the design and coordination of the study. FH conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

## Acknowledgements

We thank Martin Müller and Bruno Bello for comments on the manuscript. We also thank Clemens Broger, Martin Strahm, and Martin Neeb (F. Hoffman-La Roche) for support with the RACE-A software. We are grateful to the students of the Neurogenomics course 2002 for actively participating in this project, to Philippe Demougin for help with the array hybridizations and to Margit Böhler for assistance with the histology. This research was supported by grants from the SNSF and Neurex/Eltem

## References


1. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**(1):57-70.
2. Evan GI, Vousden KH: **Proliferation, cell cycle and apoptosis in cancer**. *Nature* 2001, **411**(6835):342-348.
3. Ponder BA: **Cancer genetics**. *Nature* 2001, **411**(6835):336-341.
4. Wechsler-Reya R, Scott MP: **The developmental biology of brain tumors**. *Annu Rev Neurosci* 2001, **24**:385-428.
5. Kozmik Z, Sure U, Ruedi D, Busslinger M, Aguzzi A: **Deregulated expression of PAX5 in medulloblastoma**. *Proc Natl Acad Sci U S A* 1995, **92**(12):5709-5713.
6. Rasheed BK, Bigner SH: **Genetic alterations in glioma and medulloblastoma**. *Cancer Metastasis Rev* 1991, **10**(4):289-299.
7. Yokota N, Aruga J, Takai S, Yamada K, Hamazaki M, Iwase T, Sugimura H, Mikoshiba K: **Predominant expression of human zic in cerebellar granule cell lineage and medulloblastoma**. *Cancer Res* 1996, **56**(2):377-383.
8. Holland EC: **Brain tumor animal models: importance and progress**. *Curr Opin Oncol* 2001, **13**(3):143-147.
9. Bate M, Martinez-Arias A: **The Development of *Drosophila melanogaster*.** *Cold Spring Harbor, Cold Spring Harbor Laboratory Press* 1993.
10. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF *et al*: **The Genome Sequence of Drosophila melanogaster**. *Science* 2000, **287**:2185-2218.
11. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W *et al*: **Comparative Genomics of the Eukaryotes**. *Science* 2000, **287**(5461):2204-2218.
12. Fortini ME, Skupski MP, Boguski MS, Hariharan IK: **A survey of human disease gene counterparts in the Drosophila genome**. *J Cell Biol* 2000, **150**(2):F23-30.
13. Gateff E: **Malignant neoplasms of genetic origin in Drosophila melanogaster**. *Science* 1978, **200**(4349):1448-1459.
14. St John MA, Xu T: **Understanding human cancer in a fly?** *Am J Hum Genet* 1997, **61**(5):1006-1010.
15. Pagliarini RA, Xu T: **A genetic screen in Drosophila for metastatic behavior**. *Science* 2003, **302**(5648):1227-1231.
16. Potter CJ, Turenchalk GS, Xu T: **Drosophila in cancer research. An expanding role**. *Trends Genet* 2000, **16**(1):33-39.
17. Reiter LT, Potocki L, Chien S, Gribskov M, Bier E: **A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster**. *Genome Res* 2001, **11**(6):1114-1125.
18. De Lorenzo C, Mechler BM, Bryant PJ: **What is Drosophila telling us about cancer?** *Cancer Metastasis Rev* 1999, **18**(2):295-311.
19. Karim FD, Rubin GM: **Ectopic expression of activated Ras1 induces hyperplastic growth and increased cell death in Drosophila imaginal tissues**. *Development* 1998, **125**(1):1-9.

20. Lowy DR, Willumsen BM: **Function and regulation of ras**. *Annu Rev Biochem* 1993, **62**:851-891.
21. Xu T, Wang W, Zhang S, Stewart RA, Yu W: **Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase**. *Development* 1995, **121**(4):1053-1063.
22. St John MA, Tao W, Fei X, Fukumoto R, Carcangiu ML, Brownstein DG, Parlow AF, McGrath J, Xu T: **Mice deficient of Lats1 develop soft-tissue sarcomas, ovarian tumors and pituitary dysfunction**. *Nat Genet* 1999, **21**(2):182-186.
23. Xia H, Qi H, Li Y, Pei J, Barton J, Blackstad M, Xu T, Tao W: **LATS1 tumor suppressor regulates G2/M transition and apoptosis**. *Oncogene* 2002, **21**(8):1233-1241.
24. Turenchalk GS, St John MA, Tao W, Xu T: **The role of lats in cell cycle regulation and tumorigenesis**. *Biochim Biophys Acta* 1999, **1424**(2-3):M9-M16.
25. Tapon N, Harvey KF, Bell DW, Wahrer DC, Schiripo TA, Haber DA, Hariharan IK: **salvador Promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines**. *Cell* 2002, **110**(4):467-478.
26. Torok M, Etkin LD: **Two B or not two B? Overview of the rapidly expanding B-box family of proteins**. *Differentiation* 2001, **67**(3):63-71.
27. Jensen K, Shiels C, Freemont PS: **PML protein isoforms and the RBCC/TRIM motif**. *Oncogene* 2001, **20**(49):7223-7233.
28. Arama E, Dickman D, Kimchie Z, Shearn A, Lev Z: **Mutations in the beta-propeller domain of the Drosophila brain tumor (brat) protein induce neoplasm in the larval brain**. *Oncogene* 2000, **19**(33):3706-3716.
29. Sonoda J, Wharton RP: **Drosophila Brain Tumor is a translational repressor**. *Genes Dev* 2001, **15**(6):762-773.
30. Frank DJ, Edgar BA, Roth MB: **The Drosophila melanogaster gene brain tumor negatively regulates cell growth and ribosomal RNA synthesis**. *Development* 2002, **129**(2):399-407.
31. Edwards TA, Wilkinson BD, Wharton RP, Aggarwal AK: **Model of the brain tumor-Pumilio translation repressor complex**. *Genes Dev* 2003, **17**(20):2508-2513.
32. Hankins G: **Analysis of a Drosophila neuroblastoma gene.** *PhD thesis* 1991, **Department of Biology, University of Virginia, Charlottesville**:107.
33. Kurzik-Dumke U, Phannavong B, Gundacker D, Gateff E: **Genetic, cytogenetic and developmental analysis of the Drosophila melanogaster tumor suppressor gene lethal(2)tumorous imaginal discs (1(2)tid)**. *Differentiation* 1992, **51**(2):91-104.
34. Woodhouse E, Hersperger E, Shearn A: **Growth, metastasis, and invasiveness of Drosophila tumors caused by mutations in specific tumor suppressor genes**. *Dev Genes Evol* 1998, **207**(8):542-550.
35. Bonini NM, Leiserson WM, Benzer S: **The eyes absent gene: genetic control of cell survival and differentiation in the developing Drosophila eye**. *Cell* 1993, **72**(3):379-395.
36. Cutforth T, Gaul U: **A methionine aminopeptidase and putative regulator of translation initiation is required for cell growth and patterning in Drosophila**. *Mech Dev* 1999, **82**(1-2):23-28.

37. Alsina B, Corominas M, Berry MJ, Baguna J, Serras F: **Disruption of selenoprotein biosynthesis affects cell proliferation in the imaginal discs and brain of Drosophila melanogaster**. *J Cell Sci* 1999, **112 ( Pt 17)**:2875-2884.
38. Driscoll M, Gerstbrein B: **Dying for a cause: invertebrate genetics takes on human neurodegeneration**. *Nat Rev Genet* 2003, **4**(3):181-194.
39. Saunders RD, Avides MC, Howard T, Gonzalez C, Glover DM: **The Drosophila gene abnormal spindle encodes a novel microtubule-associated protein that associates with the polar regions of the mitotic spindle**. *J Cell Biol* 1997, **137**(4):881-890.
40. Dawson IA, Roth S, Akam M, Artavanis-Tsakonas S: **Mutations of the fizzy locus cause metaphase arrest in Drosophila melanogaster embryos**. *Development* 1993, **117**(1):359-376.
41. Scaerou F, Aguilera I, Saunders R, Kane N, Blottiere L, Karess R: **The rough deal protein is a new kinetochore component required for accurate chromosome segregation in Drosophila**. *J Cell Sci* 1999, **112 ( Pt 21)**:3757-3768.
42. Manak JR, Mitiku N, Lipsick JS: **Mutation of the Drosophila homologue of the Myb protooncogene causes genomic instability**. *Proc Natl Acad Sci U S A* 2002, **99**(11):7438-7443.
43. Ohnuma S, Philpott A, Harris WA: **Cell cycle and cell fate in the nervous system**. *Curr Opin Neurobiol* 2001, **11**(1):66-73.
44. Ohnuma S, Harris WA: **Neurogenesis and the cell cycle**. *Neuron* 2003, **40**(2):199-208.
45. Tenev T, Zachariou A, Wilson R, Paul A, Meier P: **Jafrac2 is an IAP antagonist that promotes cell death by liberating Dronc from DIAP1**. *Embo J* 2002, **21**(19):5118-5129.
46. Frank DJ, Roth MB: **ncl-1 is required for the regulation of cell size and ribosomal RNA synthesis in Caenorhabditis elegans**. *J Cell Biol* 1998, **140**(6):1321-1329.
47. Lasko P: **The drosophila melanogaster genome: translation factors and RNA binding proteins**. *J Cell Biol* 2000, **150**(2):F51-56.
48. Lambertsson A: **The minute genes in Drosophila and their molecular functions**. *Adv Genet* 1998, **38**:69-134.
49. Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT: **The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila**. *Nature* 2002, **420**(6917):837-841.
50. Mount SM, Salz HK: **Pre-messenger RNA processing factors in the Drosophila genome**. *J Cell Biol* 2000, **150**(2):F37-44.
51. Schenkel H, Hanke S, De Lorenzo C, Schmitt R, Mechler BM: **P elements inserted in the vicinity of or within the Drosophila snRNP SmD3 gene nested in the first intron of the Ornithine Decarboxylase Antizyme gene affect only the expression of SmD3**. *Genetics* 2002, **161**(2):763-772.
52. Broadus J, Fuerstenberg S, Doe CQ: **Staufen-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate**. *Nature* 1998, **391**(6669):792-795.
53. Ito T, Tyler JK, Kadonaga JT: **Chromatin assembly factors: a dual function in nucleosome formation and mobilization?** *Genes Cells* 1997, **2**(10):593-600.

54. Grewal SI, Elgin SC: **Heterochromatin: new possibilities for the inheritance of structure**. *Curr Opin Genet Dev* 2002, **12**(2):178-187.
55. Steffensen S, Coelho PA, Cobbe N, Vass S, Costa M, Hassan B, Prokopenko SN, Bellen H, Heck MM, Sunkel CE: **A role for Drosophila SMC4 in the resolution of sister chromatids in mitosis**. *Curr Biol* 2001, **11**(5):295-307.
56. Tye BK: **MCM proteins in DNA replication**. *Annu Rev Biochem* 1999, **68**:649-686.
57. Treisman JE, Follette PJ, O'Farrell PH, Rubin GM: **Cell proliferation and DNA replication defects in a Drosophila MCM2 mutant**. *Genes Dev* 1995, **9**(14):1709-1715.
58. Gerbi SA, Bielinsky AK: **DNA replication and chromatin**. *Curr Opin Genet Dev* 2002, **12**(2):243-248.
59. Leatherman JL, Jongens TA: **Transcriptional silencing and translational control: key features of early germline development**. *Bioessays* 2003, **25**(4):326-335.
60. Beltran S, Blanco E, Serras F, Perez-Villamil B, Guigo R, Artavanis-Tsakonas S, Corominas M: **Transcriptional network controlled by the trithorax-group gene ash2 in Drosophila melanogaster**. *Proc Natl Acad Sci U S A* 2003, **100**(6):3293-3298.
61. Hemavathy K, Ashraf SI, Ip YT: **Snail/slug family of repressors: slowly going into the fast lane of development and cancer**. *Gene* 2000, **257**(1):1-12.
62. Rudolph KM, Liaw GJ, Daniel A, Green P, Courey AJ, Hartenstein V, Lengyel JA: **Complex regulatory region mediating tailless expression in early embryonic patterning and brain development**. *Development* 1997, **124**(21):4297-4308.
63. Wallace K, Liu TH, Vaessin H: **The pan-neural bHLH proteins DEADPAN and ASENSE regulate mitotic activity and cdk inhibitor dacapo expression in the Drosophila larval optic lobes**. *Genesis* 2000, **26**(1):77-85.
64. Edenfeld G, Pielage J, Klambt C: **Cell lineage specification in the nervous system**. *Curr Opin Genet Dev* 2002, **12**(4):473-477.
65. Hitier R, Chaminade M, Preat T: **The Drosophila castor gene is involved in postembryonic brain development**. *Mech Dev* 2001, **103**(1-2):3-11.
66. Saucedo LJ, Edgar BA: **Why size matters: altering cell size**. *Curr Opin Genet Dev* 2002, **12**(5):565-571.
67. Skeath JB, Thor S: **Genetic control of Drosophila nerve cord development**. *Curr Opin Neurobiol* 2003, **13**(1):8-15.
68. Boquet I, Boujemaa R, Carlier MF, Preat T: **Ciboulot regulates actin assembly during Drosophila brain metamorphosis**. *Cell* 2000, **102**(6):797-808.
69. Hassan B, Vaessin H: **Regulatory interactions during early neurogenesis in Drosophila**. *Dev Genet* 1996, **18**(1):18-27.
70. Baker SE, Lorenzen JA, Miller SW, Bunch TA, Jannuzi AL, Ginsberg MH, Perkins LA, Brower DL: **Genetic interaction between integrins and moleskin, a gene encoding a Drosophila homolog of importin-7**. *Genetics* 2002, **162**(1):285-296.
71. Radtke F, Raj K: **The role of Notch in tumorigenesis: oncogene or tumor suppressor?** *Nat Rev Cancer* 2003, **3**(10):756-767.
72. Cavallaro U, Christofori G: **Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough**. *Biochim Biophys Acta* 2001, **1552**(1):39-45.

73. Tepass U, Tanentzapf G, Ward R, Fehon R: **Epithelial cell polarity and cell junctions in Drosophila**. *Annu Rev Genet* 2001, **35**:747-784.
74. Kussel P, Frasch M: **Pendulin, a Drosophila protein with cell cycle-dependent nuclear localization, is required for normal cell proliferation**. *J Cell Biol* 1995, **129**(6):1491-1507.
75. El-Husseini AE, Fretier P, Vincent SR: **Cloning and characterization of a gene (RNF22) encoding a novel brain expressed ring finger protein (BERP) that maps to human chromosome 11p15.5**. *Genomics* 2001, **71**(3):363-367.
76. Reymond A, Meroni G, Fantozzi A, Merla G, Cairo S, Luzi L, Riganelli D, Zanaria E, Messali S, Cainarca S *et al*: **The tripartite motif family identifies cell compartments**. *Embo J* 2001, **20**(9):2140-2151.
77. Koi M, Johnson LA, Kalikin LM, Little PF, Nakamura Y, Feinberg AP: **Tumor cell growth arrest caused by subchromosomal transferable DNA fragments from chromosome 11**. *Science* 1993, **260**(5106):361-364.
78. Schiebe M, Ohneseit P, Hoffmann W, Meyermann R, Rodemann HP, Bamberg M: **Loss of heterozygosity at 11p15 and p53 alterations in malignant gliomas**. *J Cancer Res Clin Oncol* 2001, **127**(5):325-328.
79. Ngan ES, Hashimoto Y, Ma ZQ, Tsai MJ, Tsai SY: **Overexpression of Cdc25B, an androgen receptor coactivator, in prostate cancer**. *Oncogene* 2003, **22**(5):734-739.
80. Blask DE, Sauer LA, Dauchy RT, Holowachuk EW, Ruhoff MS, Kopff HS: **Melatonin inhibition of cancer growth in vivo involves suppression of tumor fatty acid metabolism via melatonin receptor-mediated signal transduction events**. *Cancer Res* 1999, **59**(18):4693-4701.
81. Todorov IT, Werness BA, Wang HQ, Buddharaju LN, Todorova PD, Slocum HK, Brooks JS, Huberman JA: **HsMCM2/BM28: a novel proliferation marker for human tumors and normal tissues**. *Lab Invest* 1998, **78**(1):73-78.
82. Levens DL: **Reconstructing MYC**. *Genes Dev* 2003, **17**(9):1071-1077.
83. Jin J, Harper JW: **A license to kill: transcriptional activation and enhanced turnover of Myc by the SCF(kp2) ubiquitin ligase**. *Cancer Cell* 2003, **3**(6):517-518.
84. Zhu Y, Parada LF: **The molecular and genetic basis of neurological tumors**. *Nat Rev Cancer* 2002, **2**(8):616-626.
85. Weis K: **Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle**. *Cell* 2003, **112**(4):441-451.
86. Ruggero D, Pandolfi PP: **Does the ribosome translate cancer?** *Nat Rev Cancer* 2003, **3**(3):179-192.
87. Albertson R, Doe CQ: **Dlg, Scrib and Lgl regulate neuroblast cell size and mitotic spindle asymmetry**. *Nat Cell Biol* 2003, **5**(2):166-170.
88. Heisenberg M, Böhl K: **Isolation of anatomical brain mutants of Drosophila by histological means.** *Z Naturforsch* 1979(34):143-147.
89. Montalta-He H, Leemans R, Loop T, Strahm M, Certa U, Primig M, Acampora D, Simeone A, Reichert H: **Evolutionary conservation of otd/Otx2 transcription factor action: a genome-wide microarray analysis in Drosophila**. *Genome Biol* 2002, **3**(4):RESEARCH0015.

90. Kaiser R, Gottschalk G: **Elementare Tests zur Beurteilung von Meßdaten**. *Bibliographisches Institut Mannheim/Wien/Zürich* 1972, **BI Wissenschaftsverlag**.
91. Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani-Rose MA: **NetAffx: Affymetrix probesets and annotations**. *Nucleic Acids Res* 2003, **31**(1):82-86.
92. **The FlyBase database of the Drosophila genome projects and community literature**. *Nucleic Acids Res* 2003, **31**(1):172-175.

**Table 1** - Overview of the array experiments

| | | Experiment A | Experiment B |
|---|---|---|---|
| RNA source | *brat* tumor | adult brat [k06028] heads | adult brat [k06028] brains |
| | control | Oregon RS heads | brat [k06028] jumpout brains |
| Arrays used | | roDromega full genome | Affymetrix full genome |
| Number of replicates | | 5 Oregon RS against 6 brat [k06028] | 6 brat [k06028] jumpout against 6 brat [k06028] |
| Signal amplification | | no | yes |
| Differentially expressed genes | | 725 | 1888 |
| Overlap between the experiments | | 321 | |

**Table 1**

Overview of the two independent genome-wide gene expression studies using two different oligonucleotide array platforms to compare adult control flies with flies displaying the adult *brat*$^{k06028}$ mutant brain tumor. Row 1 and Row2 indicate the source from which total RNA was extracted. Row 1 indicates source of tumor tissue, Row 2 indicates source of control tissue used in the experiments. For experiment A (left column), total RNA was isolated from homozygous *brat*$^{k06028}$ mutant heads (row 1) as compared to total RNA isolated from Oregon R wildtype heads (row 2). For experiment B (right column), total RNA was isolated from dissected homozygous *brat*$^{k06028}$ mutant brains (row 1) as compared to total RNA isolated from dissected brains derived from flies generated by transposon excision of the *brat*$^{k06028}$ P-element (termed A2, *brat*$^{k06028}$ jumpout; row 2). Row 3 denotes the two different Affymetrix oligonucelotide arrays used. Row 4 indicates the number of replicates carried out per condition for experiment A and experiment B, respectively. Row 5 displays whether or not a signal amplification step has been performed following cRNA hybridization to the arrays. Row 6 presents the number of transcripts differentially expressed between conditions for experiment A and experiment B, respectively. Row 7 indicates the number of genes that passed the filter criteria in both experiments.

**Table 2**

*Drosophila* genes differentially regulated in *brat*$^{k06028}$ neoplastic tissue and their mammalian homologs shown to be involved in cancer formation.

| Fly gene | Mammalian homolog | Cancertype |
|---|---|---|
| CG17498 | *Hs*MAD2L1 | breast cancer |
| CycE | *Hs*CCNE1 | breast cancer |
| MAPk-Ak2 | *Hs*MAPKAPK2 | breast cancer |
| Ent2 | *Hs*SLC29A1 | breast cancer |
| CG6546 | *Hs*BAF53A | breast cancer |
| FK506-bp1 | *Hs*FK506-bp1 | childhood astrocytoma |
| cib | *Hs*TMSB4X | colon cancer |
| UTPase | *Hs*DUT | colorectal |
| Ef1gamma | *Hs*EEF1G | colorectal adenoma |
| Aats-ile | *Hs*IARS | colorectal cancer |
| msk | *Hs*IPO7 | colorectal cancer |
| CG8235 | *Hs*SCYE1 | fibrosarcoma, melanoma |
| Bub3 | *Hs*BUB3 | gastric cancer |
| Myb | *Mm*MYB | glioma |
| CG5525 | *Hs*CCT4 | hepatocelluar and colonic carcinoma |
| cdc2 | *Hs*CDC2 | hepatocellular cancer |
| Lam | *Hs*LMNB1 | hepatocellular cancer |
| CG9993 | *Mm*FATP | hepatoma 7288CTC |
| CG31232 | *Hs*CCNK | Karposi sarcoma |
| Aats-his | *Hs*HARS | laryngeal epithelial carcinoma |
| ash2 | *Hs*ASH2L | leukemia |
| und | *Hs*MetAP2 | leukemia |
| GTP-bp | *Hs*SRPR | leukemia |
| N | *Hs*Notch2 | lung cancer, leukemia |
| Dpit47 | *Hs*TTC4 | melanoma |
| nop5 | *Hs*NOP5/NOP58 | melanoma |
| btd | *Hs*EGR1 | melanoma, prostate cancer |
| Jafrac2 | *Hs*PRDX4 | mesothelioma |
| Nup98 | *Hs*NUP98 | myeloid leukemia |
| Rm62 | *Hs*DDX5 | ovarian cancer |
| fzy | *Hs*CDC20 | pancreatic cancer |

| | | |
|---|---|---|
| mfas | *Hs*TGFBI | pancreatic cancer |
| CG15000 | *Hs*NAB1 | prostate cancer |
| Hsp27 | *Hs*HSPB2 | prostate cancer |
| CG8142 | *Hs*RFC4 | prostate cancer |
| ran | *Hs*RAN | prostate cancer |
| Ef1alpha48D | *Hs*EEF1A1 | prostate cancer |
| tll | *Mm*Nr2f1 | retinoblastoma |
| Ercc1 | *Hs*ERCC1 | squamous cell carcinoma of the head and neck |
| CkIIbeta | *Hs*CSNK2B | squamous cell carcinoma of the head and neck |
| GstE1 | *Hs*GSTT1 | thyroid cancer |
| Mipp1 | *Hs*MINPP1 | thyroid cancer |
| CG9344 | *Hs*SNRPF | uterine cancer |
| RnrS | *Hs*RRM2 | uterine cancer |
| SmD3 | *Hs*SNRPD3 | uterine cancer |
| Mcm3 | *Hs*MCM3 | uterine cervical carcinoma |
| polo | *Hs*PLK | various |
| Cyclin A | *Hs*CCNA1 | various |
| Mcm2 | *Hs*MCM2 | various |
| Klp61F | *Hs*KIF11 | various |
| mus209 | *Hs*PCNA | various |
| CG9772 | *Hs*SKP2 | various |
| Fib | *Hs*FBL | various |
| stg | *Hs*CDC25B | various |
| CycB | *Hs*CCNB1 | various |
| Cdk7 | *Hs*CDK7 | various |
| eEF1delta | *Hs*EEF1D | various |
| crc | *Hs*ATF4 | various |
| CG8586 | *Hs*KLKB1 | various |
| betaTub60D | *Hs*TUBB2 | various |
| BM-40/SPARC | *Hs*SPARC | various |
| dm | *Hs*MYC | various |

**Table 2**

*Drosophila* genes differentially regulated in *brat*$^{k06028}$ neoplastic tissue and their mammalian homologs shown to be involved in cancer formation. Indicated are the *Drosophila* genes, their human (Homo sapiens, Hs) or mouse (Mus musculus, Mm) homolog, and the cancer type for which misregulation of the mammalian homolog has been reported in available databases like

Flybase and PubMed. References are given in the supplementary material (References for table 2) [see Additional file 5].

## Additional files

### Supplementary table 1

Filename: SupplementaryTable1.xls
Fileformat: Microsoft Excel2000

Raw data of experiment A after analysis with the RACE-A package. Column 1: probeset ID, column 2: gene symbol, column 3: mean signal for Oregon R heads, column 4: mean signal for *brat*$^{k06028}$ heads, column 5: fold change between conditions, column 6: T-test describing significance of change between conditions.

### Supplementary table 2

Filename: SupplementaryTable2.xls
Fileformat: Microsoft Excel2000

Raw data of experiment B after analysis with the RACE-A package. Column 1: gene symbol, column 2: Flybase gene identifier, column 3: mean signal for A2 brains, column 4: mean signal for *brat*$^{k06028}$ brains, column 5: fold change between conditions, column 6: T-test describing significance of change between conditions.

### Supplementary table 3

Filename: SupplementaryTable3.xls
Fileformat: Microsoft Excel2000
All 321 genes the 2 experiments agree on as being differentially expressed between wildtype control and *brat*$^{k06028}$. Column 1: gene symbol, column 2: Flybase gene identifier, column 3: fold change between conditions of experiment A, column 4: fold change between conditions of experiment B.

### Supplementary table 4

Filename: SupplementaryTable4.xls
Fileformat: Microsoft Excel2000
The 172 probe sets of the roDROMEGa array that could not be linked to any probe set on the Affymetrix *Drosophila* genome 1 array. Column 1: roDROMEGa probe set ID, column 2: gene symbol.

### References for table 2

Filename: References for table 2.doc
Fileformat: Microsoft Word2000

List of references to the literature used for table 2 - mammalian homologs of genes differentially regulated in *brat*$^{k06028}$ neoplastic tissue.

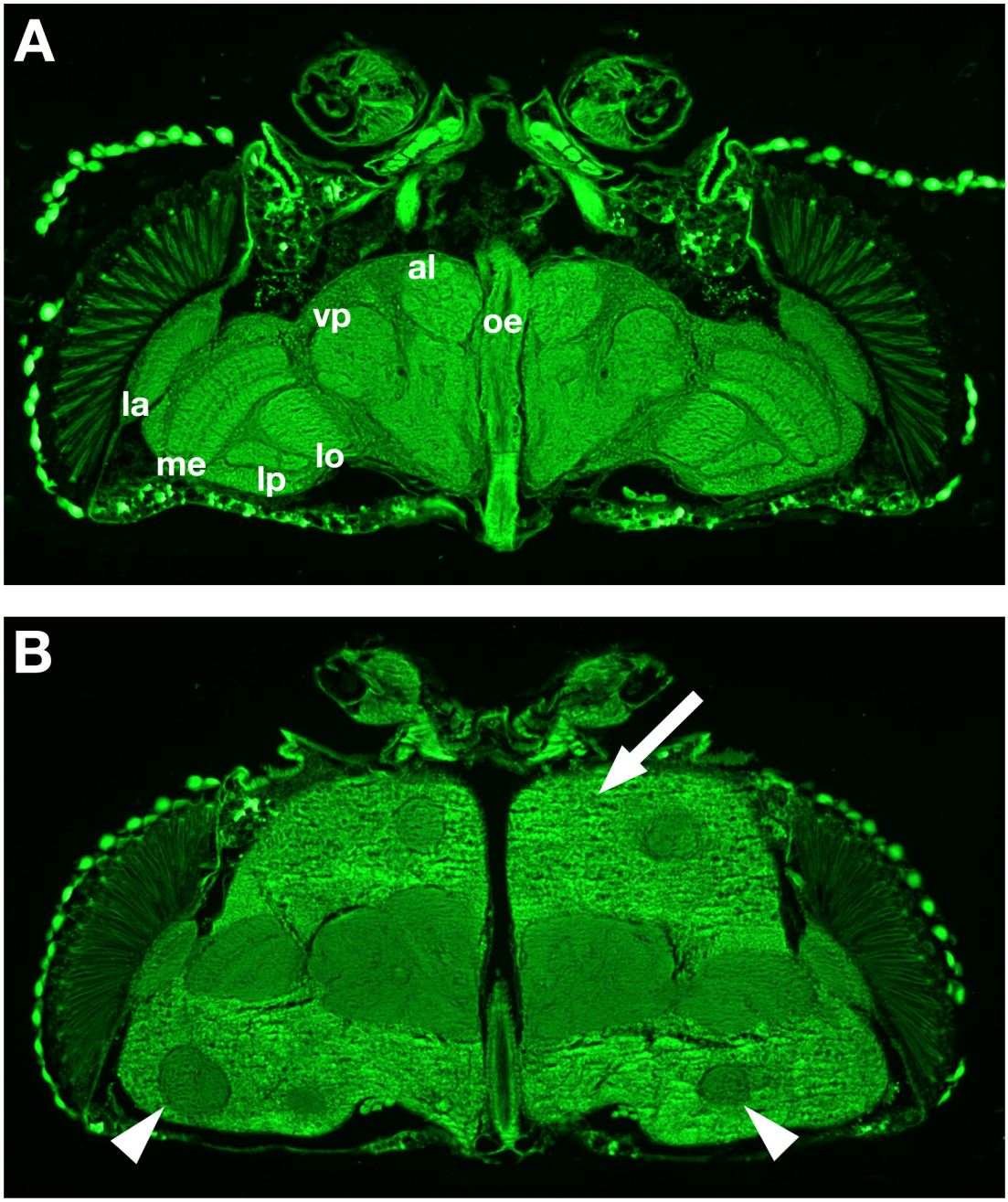



**Figure 1**
Adult flies homozygous for the *bratk06028* allele show a strong neoplastic brain phenotype. Horizontal 7 m paraffin sections at the level of the oesophagus (oe) of the heads of two days old adult Oregon R wildtype (A) and *bratk06028* (B) flies. Homozygous adult *bratk06028* mutants display a pronounced neoplastic phenotype in their heads (B, arrow) and ectopic neuropil-like structures can be observed (B, arrowheads). The size of several neuropil structures like lamina (la), medulla (me), lobula (lo), lobula plate (lp), antennal lobe (al), and ventrolateral protocerebrum (vp) appears reduced (B, compare to wildtype in A). However, their overall morphology appears to be unaffected, suggesting that the cells giving rise to these neuropil structures differentiated normally in homozygous adult *bratk06028* mutants.

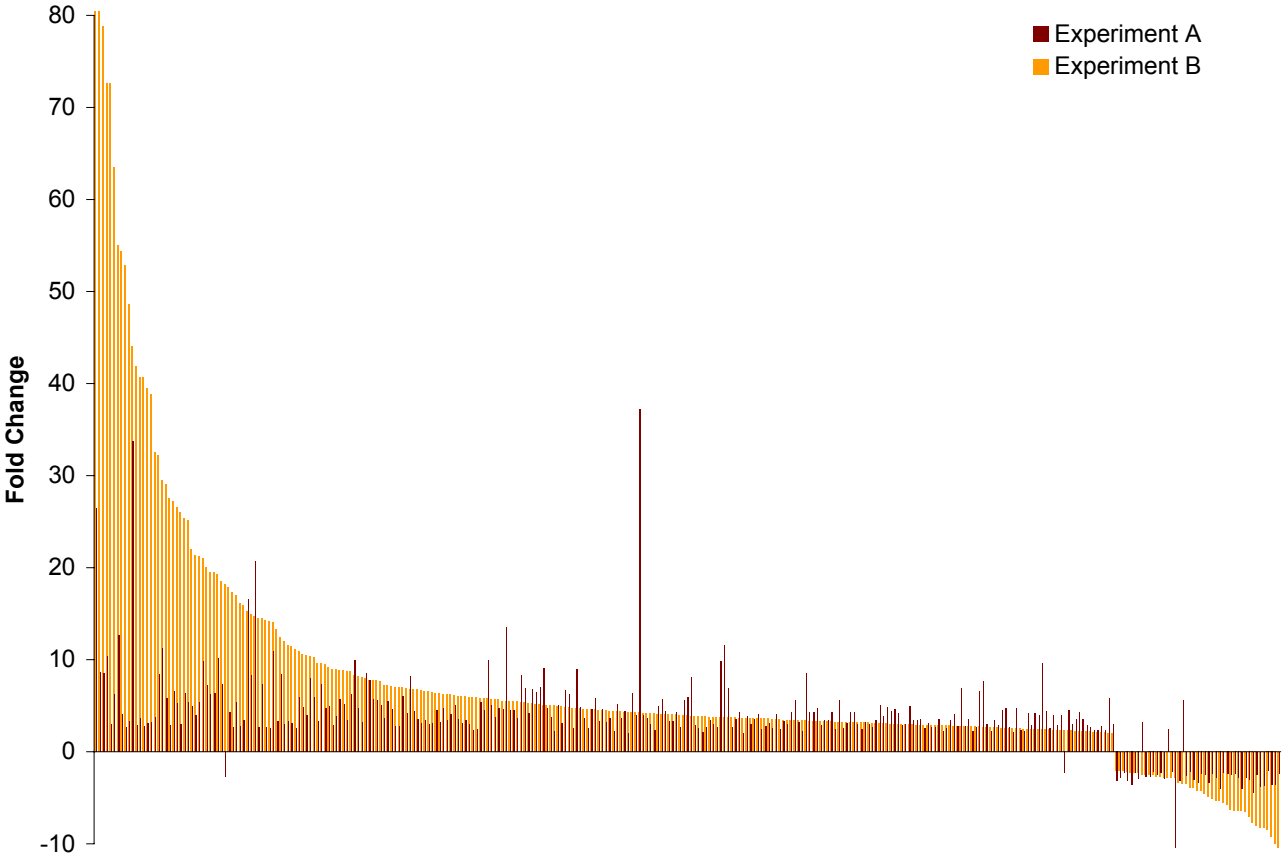



**Figure 2**
Expression level changes observed for the 321 genes identified in both oligonucleotide array experiments as differentially regulated in the brat tumor condition. X-coordinate shows relative fold-changes for each individual gene (y-coordinate) depicted in red for experiment A and depicted in orange for experiment B, respectively. Note that the two experiments A and B disagree on the differential expression level changes only for five transcripts. Please note that the highest expression levels are cut at a value of 80 due to the scaling of the y-axis.

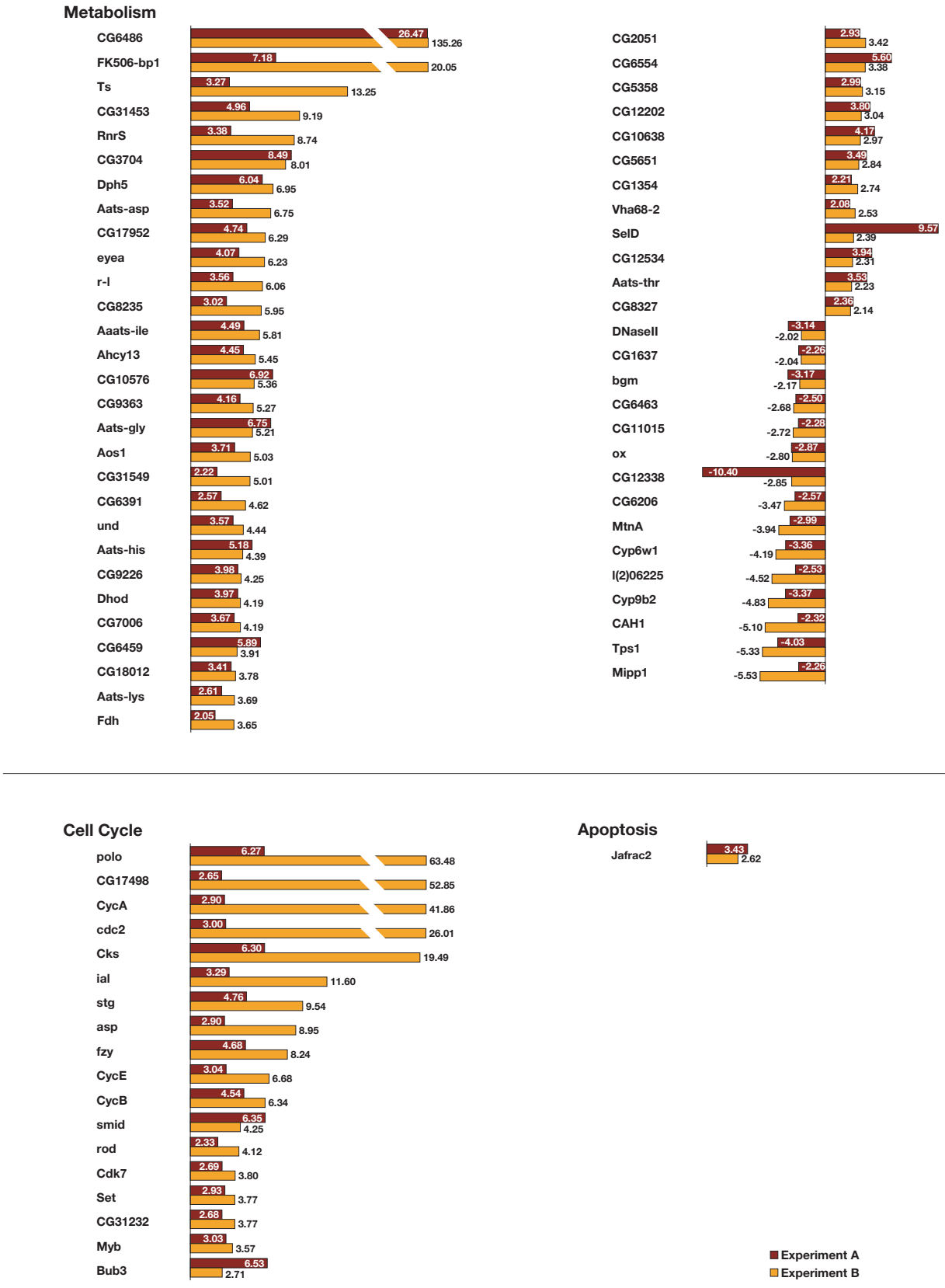



**Figure 3**
Differential expression of genes belonging to the functional classes metabolism, cell cycle and apoptosis. Fold-changes for each gene are shown in red for experiment A and in orange for experiment B, respectively.

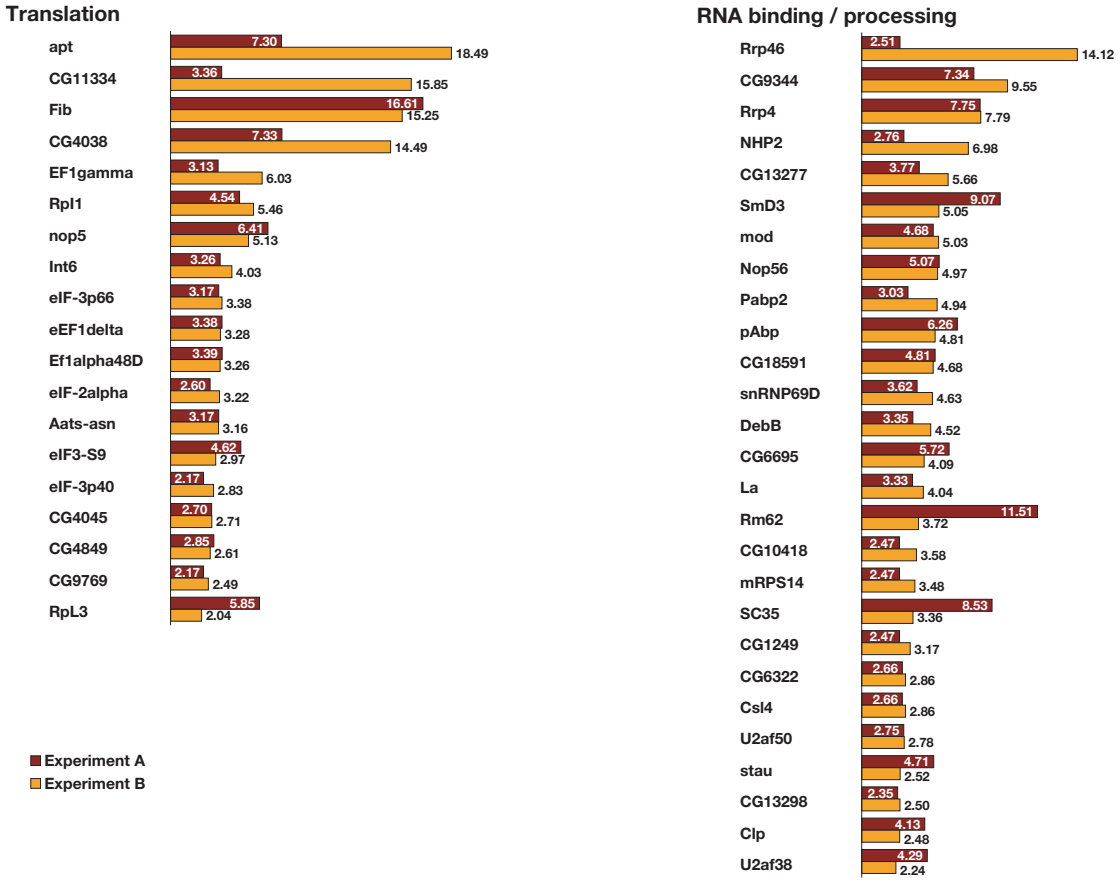



**Figure 4**
Differential expression of genes belonging to the functional classes translation and RNA binding / processing. Fold-changes for each individual gene are shown in red for experiment A and in orange for experiment B, respectively.

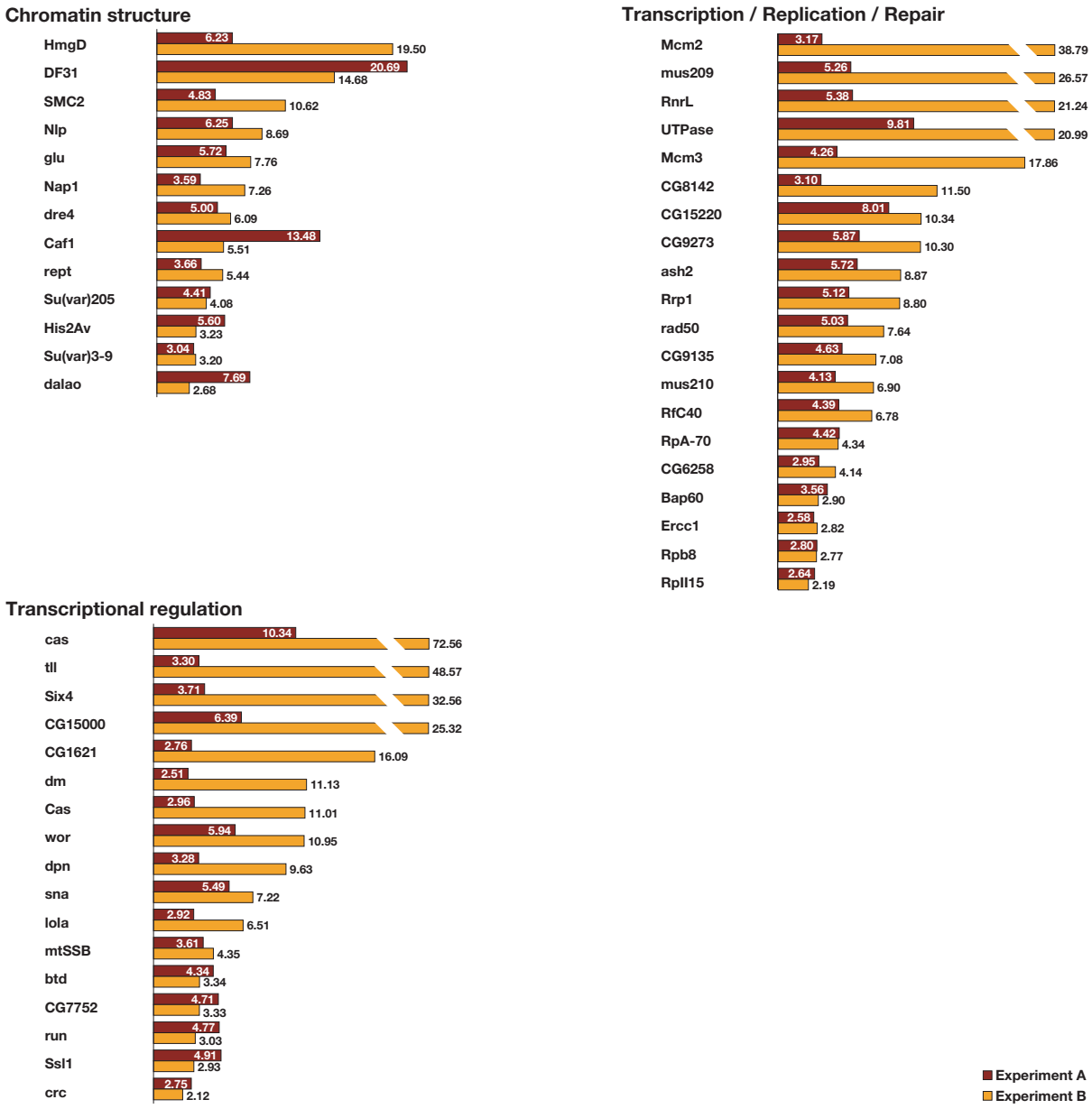



**Figure 5**
Differential expression of genes belonging to the functional classes chromatin structure, transcription / replication / repair, and transcriptional regulation. Fold-changes for each gene are shown in red for experiment A and in orange for experiment B, respectively.

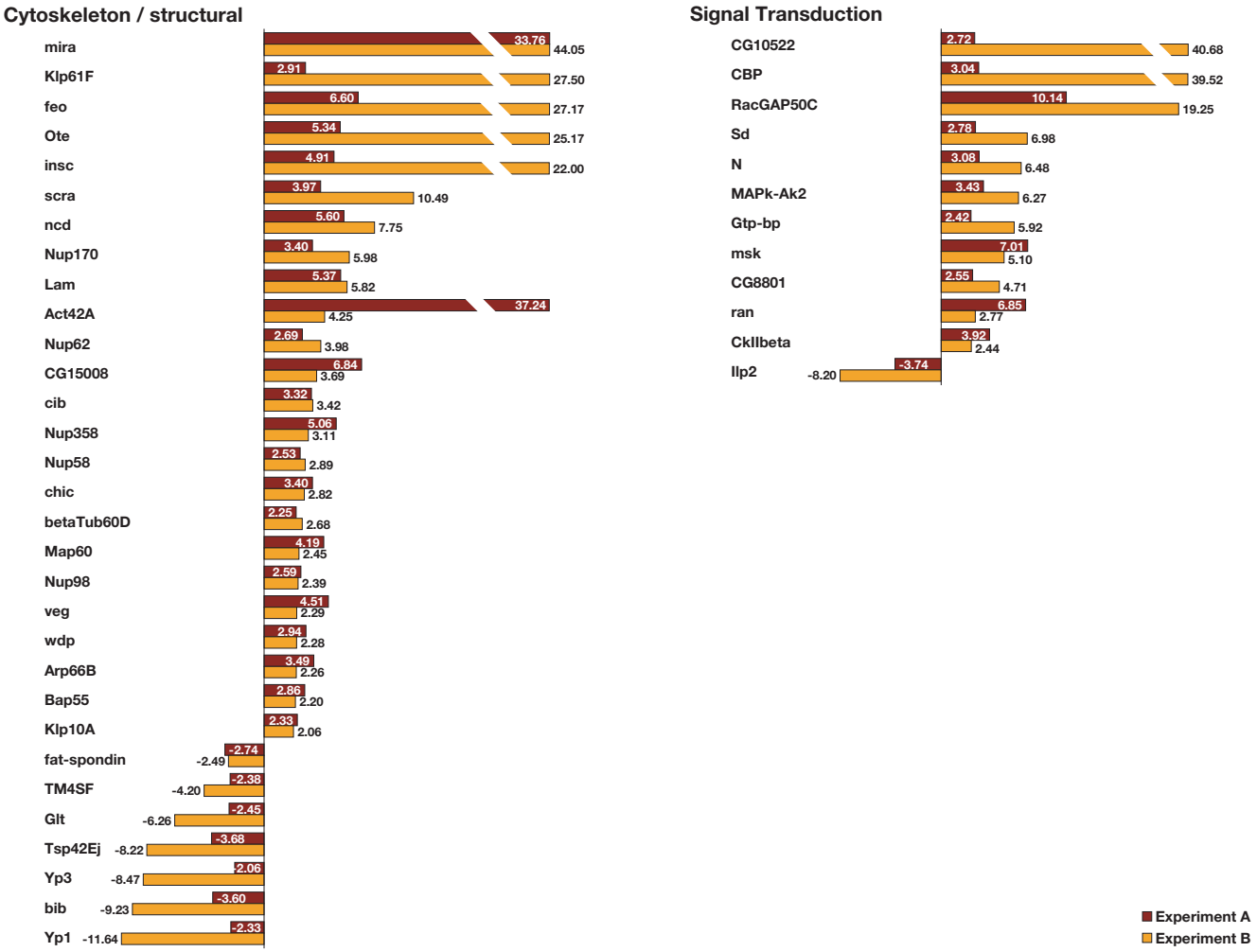



**Figure 6**
Differential expression of genes belonging to the functional classes cytoskeletal / structural and signal transduction. Fold-changes for each gene are shown in red for experiment A and in orange for experiment B, respectively.

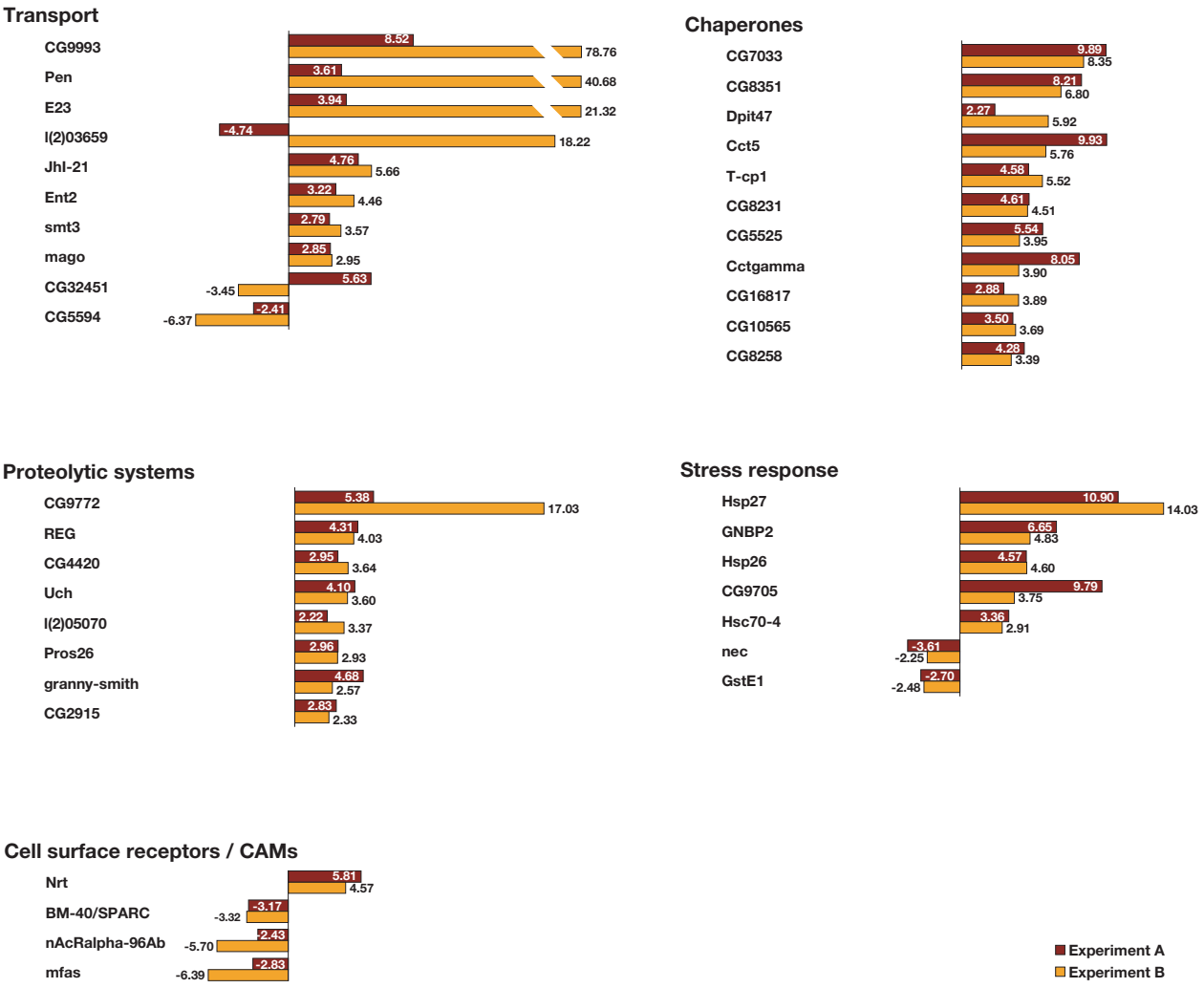



**Figure 7**
Differential expression of genes belonging to the functional classes cell surface receptors /CAMs, chaperones, proteolytic systems, stress response, and transport. Fold-changes for each gene are shown in red for experiment A and in orange for experiment B, respectively.

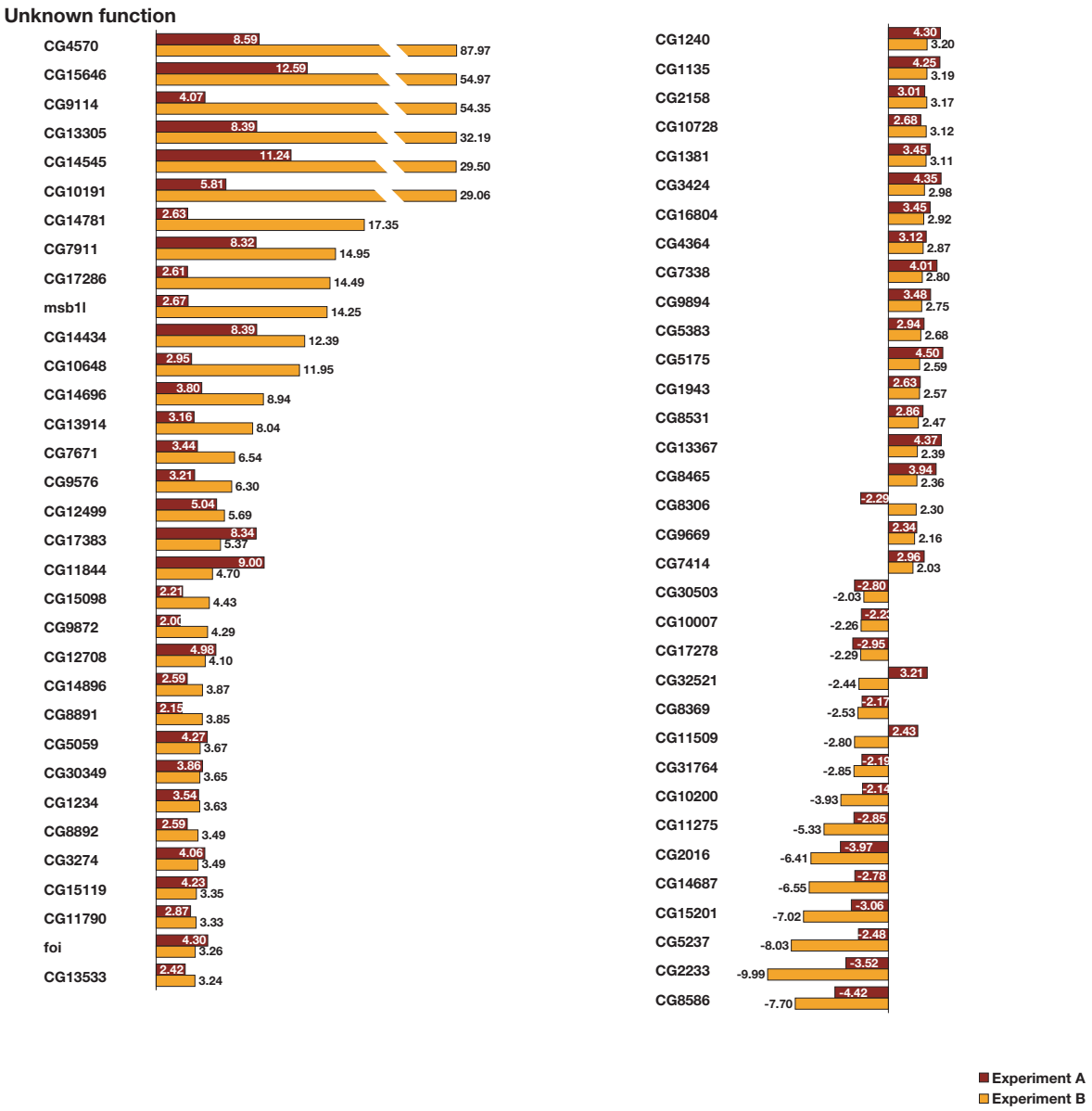



**Figure 8**
Differential expression of genes belonging to the functional class unknown function. Fold-changes for each gene are shown in red for experiment A and in orange for experiment B, respectively.

# 9. Discussion

## Functional equivalence of hometic proteins and Labial target genes

Homeotic proteins are evolutionary conserved transcription factors regulating a wide array of developmental processes. Albeit the large body of data accumulated about the DNA-binding domain of homeotic proteins, only little is known about their target genes.
We used high density oligonucleotide arrays representing approx. 1500 genes to look at the transcriptional profile of transgenic *Drosophila* embryos carrying the *lab* gene under control of a heat-shock promotor. Heat-shock treatment led to ubiquitous overexpression of *labial*. The transcriptional response in this gain-of-function situation was subject to our measurements of relative transcript levels.
Contrasting results from other labs stating that homeoproteins control the majority of zygotically active genes (Carr and Biggin, 1999; Liang and Biggin, 1998), we found a only a limited set of genes as being regulated by Labial. These nevertheless included transcripts encoding for a wide array of functions.
A drawback of our approach using ubiquitous overexpression in embryos is the averaging of gene expression through several developmental stages and throughout the complete embryo. Although array technology can give an accurate image of relative transcript abundance, this integration can lead to masking of transcript levels. Especially as homeotic genes are likely to show high tissue specificity, activation and repression of a specific gene can occur in different tissues at different time points, thereby effectively blurring the transcriptional image measured by the array experiment. This could also be a reason for the failure in trying to validate the data using *in situ* hybridization experiments in contrast to the successful validation of our results using *real-time* PCR.

As it is known that the specificity of homeotic proteins is modulated by other proteins, the lack of interaction of Labial with candidate co-factors like EXD or HTH (Nagao et al., 1998) in the source tissues may further distort our dataset.
However, our data represents a good starting point for a more in-depth functional analysis of candidate *labial* downstream genes.

Another study focusing on homeotic gene function that was conducted in the course of this thesis addressed the role of Labial in specification of the posterior tritocerebrum. Using genetic rescue experiments we replaced the *labial* gene in the developing tritocerebrum with all other homeotic genes. With the exception of *Abd-B*, all were able to replace *labial* in its function of giving segmental identity to this brain region.
*In vitro* experiments have demonstrated that homeoproteins by themselves have indeed a low DNA-binding specificity and recognize similar nucleotide sequences (Ekker et al., 1994). This specificity has to be ensured by other means, for example through interaction with other transcription factors or by additional proteins binding to the same target sequence (reviewed in (Mann and Affolter, 1998; Mann and Carroll, 2002)). The abovementioned co-factors add to the complexity of the issue, as it is known that EXD and HTH are essential for the

segmental identity of *Drosophila* brain neuromeres (Nagao et al., 1998) and that they interact with homeoproteins (Chan et al., 1996).

Hints towards why *Abd-B* is not able to replace *labial* function in the posterior tritocerebrum may be found when looking at the evolution of homeoproteins. *Labial* and *Abd-B* are most distant from each other in terms of sequence similarity because they represent the most ancestral *Hox* genes (Schubert et al., 1993).
Interaction of homeoproteins with EXD and the DNA is mediated by the so-called hexapeptide, a motif that can be found in all Hox proteins, except in *Abd-B* (Chan et al., 1996). This difference in peptide sequence could also be one of the major reasons for the functional difference of *Abd-B* when compared to the other Hox genes and represents a good entry point into further functional studies on homeoprotein target genes.

## *Glial cells missing* and regulation of neurogliogenesis

The *glial cells missing* (gcm) gene was the focus of a further array study we directed during the time of this thesis. *gcm* encodes for a transcription factor and is thought to be a key regulator on the switch between glial and neuronal cell fate. To identify the target genes that run the glial differentiation program we used full-genome oligonucleotide arrays to analyze differential expression in wild-type embryos compared to embryos in which *gcm* is misexpressed throughout the neuroectoderm. Furthermore, samples were taken at two different developmental stages, namely at the time of initial *gcm* action in neuro-glial progenitors (stage 10) and at a later stage with the glial cells already differentiated (stage 15/16). In both cases hundreds of differentially expressed candidate target genes could be found.

The list of differentially regulated genes found in our study in the *gcm* gain-of-function situation demonstrates the complexity of neuroglial fate choice. The dataset represents a complicated genetic network. Furthermore, GCM seems to be able to act on all embryonic aspects of gliogenesis in *Drosophila*. This puts the gene into a class together with the *Pax* and *Hox* genes, for which a similar high-level position in genetic pathway control has been supposed. However, no detailed knowledge has emanated yet, that could explain how a single transcription factor can achieve such a comprehensive control over various genetic networks.
Our data shows, that GCM is a direct regulator of a set of factors including REPO, PNTP1 , TTK69, and signal transducers such as the FGF-receptor HTL. These are prime candidates for the next lower level in regulating neurogliogenesis in the embryo of *Drosophila*.

## Evolutionary conservation of Otd/Otx transcription factors

Also, we have analyzed the action of the *orthodenticle* (*otd*) gene and its vertebrate homolog *Otx2* using transgenic overexpression and quantitative transcript imaging on the level of the complete genome. This resulted in hundreds of candidate target genes of *otd* and/or *Otx2*.
Comparison of the two datasets showed that the *otd* and *Otx2* transgenes control an overlapping set of genes in *Drosophila*.

Despite their strong evolutionary conservation, the two proteins display some profound differences in their molecular sequence. This can also be seen in the transcriptional profile following overexpression of the transgenes. *Otx2* does not seem to be able to replace *otd* in full, as more than half of the candidate downstream genes of *otd* do not appear in the *Otx2* dataset. This is in accordance with *in vivo* experiments focusing on the cross-phylum rescue (Leuzinger et al., 1998; Nagao et al., 1998).
The genes that do show up in both overexpession experiments are likely to embody the molecular basis for functional equivalence of the *otd*/*Otx* genes.

In our experiments we discovered approximately 300 genes that respond to the overexpession of *otd* in the embryo. This supports the notion that homeoproteins regulate a subset and not all of the zygotically active genes in the genome (Leemans et al., 2001). The fact that among these genes many transcription factors can be found is consistent with the idea that homeodomain proteins are on a high position in the hierarchy of transcriptional regulators (Kablar et al., 1996). However, with our array study we are unable to differentiate between direct and indirect target genes.

A very interesting finding of our experiments was the greater number of candidate target genes yielded by overexpression of the human *Otx2* gene when compared to the *Drosophila* transgene. After quality control 700 genes remained to be significantly influenced by overexpression of the human *Otx2* transgene. The transcript abundances of *otd* and *Otx2* are not likely to be the reason for this discrepancy, as the *otd* transgene produced higher levels of transcript. Reasons for the disparity in target gene number may be a lower binding specificity of *Otx2* compared to that of the *otd* transgene, higher promiscuity in co-factor interaction or the regulation of additional genetic cascades.

## Transcriptional profile of an adult *Drosophila* tumor

In our *brat* array study we compared *Drosophila* wildtype adult brains with *brat* mutant adult brains. The fact that mutations in the *brat* gene cause brain tumors late larval stages has already been shown (Arama et al., 2000). Looking at the adult escapers of the $brat^{k06028}$ line gave us the possibility to compare the transcriptome of a tumor in full effect with that of normal brain tissue.
We have conducted two independent genome-wide gene expression studies using two different oligonucleotide microarray platforms. Subject of the comparison were adult wildtype flies and flies displaying the adult $brat^{k06028}$ mutant brain tumor. The two experiments were cross-validated to identify a core transcriptional signature of $brat^{k06028}$ neoplastic tissue.

Albeit the considerable differences between the two experiments in terms of the biological material and the array platforms, a highly reproducible transcriptional signature of malignant growth could be generated.

Due to very stringent quality control (t-Test p≤0.0001) the 321 transcripts detected as being differentially regulated in the *brat*$^{k06028}$ tumor condition probably represent only a subset of genes that could be involved in brain tumorigenesis. An example for a gene that did not make it into our core dataset is PTEN. It is a protein with important function in the highly conserved insulin/PI3-kinase (PI3K) signaling pathway involved in cell growth and proliferation (for review see (Saucedo and Edgar, 2002)) and is frequently lost in human gliomas (Zhu and Parada, 2002). PTEN was detected as being differentially down-regulated in the brat tumor condition. Due to a p-value above our threshold it did not make it into the core dataset. However, distinctive transcriptional features of cancer cells, including those prevalent in humans, could be identified. Our dataset includes gene expression changes suggesting elevated and aberrant metabolic activity, elevated and aberrant activity of the cell cycle and basal transcriptional machinery, and elevated and aberrant activity of ribosome synthesis and translation control.

Cancer involves a multi-step process in which a cell has to acquire a variety of traits in order to become able to divide continually regardless of cellular environment (Hanahan and Weinberg, 2000). This requires a very different transcriptional program when compared to a normal neuronal cell. As expected we found a very high amount of genes being differentially expressed when compared to wildtype.
We did not detect any significant alteration of gene expression for transcripts belonging to the functional class *apoptosis*, except elevated expression levels for thioredoxin peroxidase 2 (Jafrac2), which is involved in the induction of programmed cell death (Tenev et al., 2002). The fact that we only identified one transcript in this functional class suggests that the apoptotic program is either impaired below detectability or is unaltered in neoplastic tissue of adult *brat*$^{k06028}$ mutant brain tumors.

It has been speculated that the cellular source of the *brat*$^{k06028}$ tumor are GMCs in the optic lobes of *Drosophila* (Kurzik-Dumke et al., 1992). When looking closely at the data we indeed find genes of the cascade temporally controlling GMC differentiation. In neuroblast lineage development there are five genes expressed in the neuroblasts in a temporally transitional manner. These genes are *hunchback* (*hb*), *Kruppel* (*Kr*), *pdm*, *castor* (*cas*) and *grainy head* (*gh)*. GMCs born at specific time points retain the expression of the currently active gene from this cascade and also bequeath it to their neuronal progeny (Isshiki et al., 2001; Skeath and Thor, 2003). This is data stemming from embryonic development.
Interestingly, we find *castor*, a gene from this cascade, coming up very strongly in the *brat*$^{k06028}$ tumor brains. Another gene, *grainy head* is also significantly expressed. As these transcripts can also be found in neuroblasts, the possibility remains that neuroblasts are the source of the tumor. *castor* expression could be evidence for relatively late neuroblasts continuing to divide and feeding the tumor with new cells, or for late born GMCs transformed into cancer cells. So far, no genetic marker is known to differentiate neuroblasts from GMCs in the larval stages of *Drosophila* (Ceron et al., 2001).

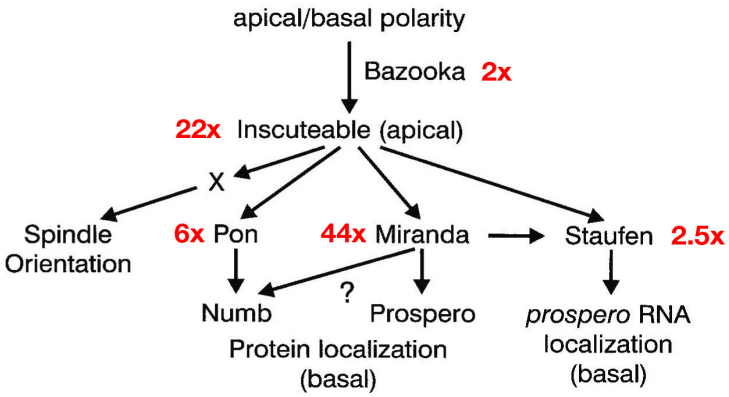



**Fig. 9.** Genes involved in asymmetric division and cellular polarity of neuroblasts. In red rounded fold changes in gene expression are depicted when comparing wt adult brains with brat mutant brains. This data is from experiment B, comparing dissected brain as RNA source.

Interestingly, we find elevated transcript levels of *insc*, *stau*, *mira, pon* and *baz* in $brat^{k06028}$ brains. This result makes it very likely that the cells responsible for this neoplasm are indeed either neuroblasts or GMCs. It also makes the pathway controlling asymmetric division of neuroblasts a prime candidate for investigating the nature of *brat* malfunction. The consistent differential expression of multiple genes contained in this pathway may even suggest that defects in asymmetry are causal for the malignant growth.

However, a decision as to which of the two cell types are responsible for the fatal proliferation can not be helped by these findings. It is still subject to research whether *insc* does segregate into the GMC after division. This question is complicated by the fact that asymmetric division of neuroblasts can happen in different modes in the larvae (Ashraf and Ip, 2001; Ceron et al., 2001).

Along these lines, the highly significant upregulation of the genes *asp*, *fzy*, *rough deal* (*rod*), and *Myb* is also of special interest. *asp* encodes a microtubule-associated protein that associates with the polar regions of the mitotic spindle. Mutants of *asp* are larval lethal, with a high frequency of aberrant (e.g. polyploid) cells arrested in metaphase in the larval brain, suggesting that asp may play a role in spindle pole organization during mitosis (Saunders et al., 1997). *fzy* encodes a product involved in cyclin catabolism. In *fzy* mutants display metaphase arrest with compact condensed chromosomes similar to *asp* mutants (Dawson et al., 1993). *rod* encodes a product involved in mitotic chromosome segregation which is localized to the kinetochore, and mutations in *rod* result in mitotic segregational failure due to delayed or incomplete release of sister chromatids (Scaerou et al., 1999). Finally *Myb* encodes a proto-oncogene with transcriptional activator activity involved in centrosome cycle which is required to sustain the appropriate rate of proliferation, to suppress formation of supernumerary centrosomes, and to maintain genomic integrity (Manak et al., 2002).

The activity of these genes involved in cell cycle regulation has been linked to several aspects of neurogenesis (Ohnuma and Harris, 2003; Ohnuma et al., 2001). Thus, all of these genes are known to be expressed during embryonic and larval development in mitotically active cells of the nervous system, the neuroblasts and ganglion mother cells (GMCs), suggesting that their elevated levels of transcriptional activity correlate with an elevated and aberrant activity of the cell cycle machinery in neoplasms of homozygous $brat^{k06028}$ mutants.

A very interesting gene showing lower transcript abundance in tumor brains compared to wildtype brains was *anachronism* (*ana*). It encodes for a secreted glycoprotein that is expressed in glia cells and acts as a suppressor of neuroblast proliferation. In *ana* mutants, quiescent postembryonic central brain and optic lobe neuroblasts enter S phase prematurely (Ebens et al., 1993). This may be a hint for direct *brat* action in glia cells or for the neoplastic cells affecting gene expression in glia.

It has been shown that *hb* is expressed in embryonic neuroblasts and that it has a role in cell fate specification in the Drosophila central nervous system (Novotny et al., 2002; Urbach and Technau, 2003). The translational repression of *hb* by a complex of Nanos, Pumilio, and Brat in the very early embryo might also apply for *hb* function in later development. However, in our transcriptional screen *hb* shows no differential expression.

As the underlying neuropil structures in adult homozygous $brat^{k06028}$ flies look very similar to those in the wildtype, the cellular phenotype of the tumor most likely does not affect the complete population of a cell type. If all neuroblasts, or all glial cells would be affected it would be highly unlikely that an even remotely normal neuropil could develop.
These findings also argue the hypothesis that the tumor starts in the larval optic lobes (Kurzik-Dumke et al., 1992). Tumor growth emanating from the optic lobes would possibly disrupt neuropil structures in this region. On the other hand one could argue, that a tumor developing from the cortex of the developing optic lobes could leave the neuropil structure untouched. This will be subject to further studies.

Recent studies have shown that *brat* mutant cells are enlarged compared to control cells, have lager nucleoli and possess excess rRNA. This suggests a function of Brat in repression of cell growth and ribosomal RNA synthesis. *brat* gain-of-function experiments demonstrated inhibition of growth in clones and organs, and a decrease of rRNA per cell (Frank et al., 2002). These findings may suggest that Brat affects cell growth by regulating ribosome synthesis, and not by controlling their activity (Frank et al., 2002), representing an alternative way of cell growth regulation independent of the insulin receptor pathway.
We identified a number of genes involved in ribosome biogenesis and RNA processing as being significantly upregulated in the neoplastic tissue. Among these is the ribosomal protein RpL3 which maps to the chromosomal region harboring the M(3) 86D Minute mutation (Lambertsson, 1998). However, the question whether protein synthesis and ribosome biogenesis are cause or consequence of tumor growth in $brat^{k06028}$ neoplasms cannot be answered by interpreting our dataset.

Furthermore, we found that for 20% of the genes differentially regulated in the tumor condition it was possible to identify vertebrate homologs associated with various types of tumors. Among these were homologs of Myb, Myc, cyclins as well as other genes involved in cell growth and proliferation control that are thought to be involved in mammalian cancer formation and malignant growth (Hanahan and Weinberg, 2000). This underlines the potential

of *Drosophila* as a high-speed model system in terms of studying developmentally related diseases.

One of the two experiments of our study showed much higher sensitivity than the other (Loop et al., 2004). The magnitude of differences can be best appreciated when only looking at the second experiment that showed a very good reproducibility and a higher sensitivity than the other experiment. With reasonable criteria in terms of quality control (p-value <= 0.01, Change more than 2-fold, signal >=10 in the higher condition) we filtered over 3500 genes as being differentially expressed in the $brat^{k06028}$ adult brains. This demonstrates the fundamental difference of a cancer transcriptome compared to normal relatively homeostatic brain tissue.

## Genome-wide transcript profiling in *Drosophila* brain development

Gene expression lies at the basis of most biological processes. This holds especially true for developmental mechanisms where it is crucial to gain knowledge about spatio-temporal expression patterns of genes. For a long time in biological research transcriptional research could only be performed on a gene-per gene basis. With the advent of expression array technologies it became possible to monitor gene expression of thousands of genes with a single experiment (Duggan et al., 1999; Lipshutz et al., 1999). *Drosophila*, together with yeast and *C. elegans*, has been leading the way in application of these high-throughput technologies, as in these model organisms the complete genomic sequence was available early.

In the course of this thesis five complete GeneChip studies were conducted covering different functional aspects of *Drosophila* nervous system development and developmentally related disease of the brain. Three of these studies were performed using arrays representing the complete fly genome while the other two were based on a custom array, representing only about 1500 fly genes. As the technologies mentioned above are relatively young and were even more so at the time some of these studies were done, a lot of issues arose in terms of what can be done with these tools and how to interpret the results.

Controlling reproducibility of the transcript measurements is a prime step to a successful experiment. This can be done by doing statistics on the data. In the simplest case this means doing a Student`s t-Test to check if two measurements representing two different conditions are really different. To perform these statistical tests there have to be replicates of the measurements. During our studies we used at least 4 measurements per condition equaling 4 arrays. This is rather on the low side. In later studies we used 6 replicates raising the statistical power of the tests (Loop et al., 2004). Replicating experiments has proven to be a good way to control for between-sample variation and reduce the number of false positives in the resulting dataset.

The most crucial part of any experiment is the initial paradigm chosen to conduct it. In the case of array experiments the RNA source lies at the core of this problem. In our first study

using GeneChips, standard wildtype embryos from the developmental stages 10-17 were compared with heatshocked wildtype embryos within the same range of ages (Leemans et al., 2000). This led to the problem that the resulting transcriptional profile was averaged over time and space. It comprised expressional aspects of all embryonic tissues developing during 17 hours leading to a reduced reproducibility between samples.
The precision of the transcriptional snapshot can only be as high as the spatio-temporal precision that can be achieved in preparing the RNA source tissue. As there is still a considerable amount of RNA needed to hybridize a single GeneChip (above 5µg) this poses severe problems when very small tissue samples or even single cells are to be investigated. When there is too much RNA extracted coming from tissues that are not of interest or from time points irrelevant to the experiment, the signal-to-noise ratio in terms of transcriptional information will drop. These problems may be overcome by methods like laser microdissection (Simone et al., 1998) or cell sorting methods like FACS and MACS (Montalta-He, personal communication). Another way to increase the RNA amount is linear amplification of the target sequences by multiple rounds of cDNA synthesis and *in-vitro* transcription. This method is thought to introduce only minor artifacts into the measured expression profiles (Baugh et al., 2001; Dumur et al., 2004).

Another issue arising with array experiments on the genomic scale is the problem of independent validation of the results. As per today, no feasible method has been devised to confirm the vast amount of data springing from a GeneChip experiment. In our hands verification using real time PCR was a reliable way to confirm expressional changes. Furthermore, we used *in situ* hybridization of RNA to whole-mount embryos, an approach that proved to be less sensitive, enabling us only to verify gross changes in gene expression. Both approaches are no high-throughput methods and therefore cannot be used to validate array experiments in an exhaustive manner. This also means that stringent filtering of the data sets and raising the number of replicate experiments is important to push the number of false positives down to a reasonable level.

When using array platforms like spotted microarrays and GeneChips one has to be aware that the genomic annotation of the different model system is a work in progress. The sequences on the array represent a snapshot of the genomic annotation at the time of array-design. This sometimes poses problems when updating a dataset resulting from an array experiment to a new release of the genome, and when comparing experiments done with different platforms. In our array study comparing wt adult brains with tumor brains dissected from *brat* mutant flies, custom software had to be developed to be able to merge data from a custom *Drosophila* GeneChip with that produced with the commercially available *Drosophila* GeneChip (Loop et al., 2004).
Comparison of data from different platforms not always gives satisfactory results (Michaut et al., 2003). The reasons for this may vary, including the above-stated annotational differences, or incompatibilities in experimental design.
The use of arrays to produce transcriptional snapshots is headmost a screen for expressional change under different conditions, may they be of genotypic or environmental origin.

Depending on how far apart the experimental RNA sources are in terms of qualitative and quantitative gene expression, an array experiment will yield a smaller or bigger number of differentially expressed genes. In the case of our analysis of the *brat* mutant brains, the resulting dataset of differentially expressed genes would have been in the range of several thousands, would we have taken into account the experiment using the commercially available array only, and used generally accepted cut-off values for our stringency criteria ( a p-value lower than 0.01 is generally considered being "highly significant") (Loop et al., 2004). The idea behind our data management was to combine the two experiments using different array platforms and a slightly different experimental paradigm to strengthen the results and take into account only those genes both experiments agreed on.
The main differences between the experiments were: using fly heads as RNA source as opposed to dissected brains, the different GeneChip platforms and hybridizing to the arrays without or including signal amplification. Arguably, amplifying the signal using an antibody should be the major reason for differences, as it raises the sensitivity of the method. Transcripts of lower abundance can be detected, raising the number of detected differentially expressed genes. The noise was not increased by signal amplification, as the reproducibility as judged by the number of detected highly significant transcript changes, was still very good (Loop et al., 2004).

When compared to the other GeneChip studies done during the period of this thesis, the *brat* study shows the advantages of a defined tissue source. The extremely good statistical outcome for many genes has its reason in a lower transcriptional noise level. The wildtype adult brain is a complex tissue, nevertheless. Comparing it to a brain containing a high-magnitude tumor will clearly show the completely different transcriptional program of cancer cells.

Still, one of the major difficulties a researcher embarking on a full-genome transcription profiling experiment has to face, is the sheer amount of data to follow-up. After having done a thorough quality control hundreds or thousands of candidate genes may remain. The first steps have to include grouping these genes into functional bins. This has become a bit easier, recently, as a gene ontology (GO) is being developed (Ashburner et al., 2000). The GO consists of three structured, controlled vocabularies that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. Integrated into data-mining software like GeneSpring (Grewal et al., 2003) the GO becomes a powerful tool in getting insights into the result of high-throughput expression screens. Even more information can be gathered from an array experiment by grouping the genes into specific genetic cascades for which already functional data exists (see Fig. 9).
Having boiled down the dataset to a set of good candidates the biological follow-up has to begin. Complete biological follow-up of array experiments yielding thousands of differentially expressed genes is a task beyond the capabilities of any lab. Ways around this dilemma can be to narrow down the experimental paradigm so that the resulting dataset is small. Comparison of the transcriptional profiles of two very similar tissues would be such a

case. However, this is not always possible. Nevertheless, such expression studies add to the wealth of data the scientific community can use to accelerate research decisions.

The experiments conducted to follow-up array experiments span the complete range of methods available in biological research. In case of a classical array study that, for example, screens for target genes of a transcription factor, these would include loss- and gain-of-function experiments using the candidate downstream genes found in the dataset, if possible in mutant background of the gene of interest that was used in the array study. This tedious gene by gene approach still is the prime way to elucidate functional information on the genetic pathways controlling biological processes.

A good way to narrow down a dataset emanating from an array study focusing on transcription factors is to apply additional bioinformatics methods like *in silico* screening for enhancers containing binding sites for the protein of interest. This combinatorial approach has been exploited to great success in a screen for target genes of the morphogen *dorsal* (Alon et al., 1999).

Of exceedingly high importance will be the setup of standardized ways to annotate and publish array data. As more and more ways to analyse it will be devised, standardization is the only way to prevent old array data to become to obscure to be instructive. Leading the way was the Gene Expression Omnibus (Edgar et al., 2002). The GEO is a high-throughput gene expression / molecular abundance data repository, as well as a curated, online resource for gene expression data browsing, query and retrieval. Another initiative relevant in this respect is MIAME (minimum information about a microarray experiment), describing a minimal set of information that has to be included with array experiments to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment (Brazma et al., 2001).

The GeneChip studies conducted during this thesis have certainly proven that the method has matured to give reproducible results. Array studies have been utilized in a wide range of research fields. They do not give instant gratification but produce a huge amount of data that has to be tackled with patience.

## 10. References


**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. and Simeone, A.** (1999). The TINS Lecture. Understanding the roles of Otx1 and Otx2 in the control of brain morphogenesis. *Trends in NeuroScience* **22**, 116-22.

**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**, 2185-2218.

**Akam, M.** (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell* **57**, 347-9.

**Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. and Levine, A. J.** (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *America* **96**, 6745-50.

**Arama, E., Dickman, D., Kimchie, Z., Shearn, A. and Lev, Z.** (2000). Mutations in the beta-propeller domain of the Drosophila brain tumor (brat) protein induce neoplasm in the larval brain. *Oncogene* **19**, 3706-16.

**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.

**Asha, H., Nagy, I., Kovacs, G., Stetson, D., Ando, I. and Dearolf, C. R.** (2003). Analysis of Ras-induced overproliferation in Drosophila hemocytes. *Genetics* **163**, 203-15.

**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.

**Ashraf, S. I. and Ip, Y. T.** (2001). The Snail protein family regulates neuroblast expression of inscuteable and string, genes involved in asymmetry and cell division in Drosophila. *Development* **128**, 4757-67.

**Baugh, L. R., Hill, A. A., Brown, E. L. and Hunter, C. P.** (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* **29**, E29.

**Beltran, S., Blanco, E., Serras, F., Perez-Villamil, B., Guigo, R., Artavanis-Tsakonas, S. and Corominas, M.** (2003). Transcriptional network controlled by the trithorax-group gene ash2 in Drosophila melanogaster. *America* **100**, 3293-8.

**Bernardoni, R., Kammerer, M., Vonesch, J. L. and Giangrande, A.** (1999). Gliogenesis depends on glide/gcm through asymmetric division of neuroglioblasts. *Dev Biol* **216**, 265-75.

**Bertrand, N., Castro, D. S. and Guillemot, F.** (2002). Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* **3**, 517-30.

**Bier, E., Vaessin, H., Younger-Shepherd, S., Jan, L. Y. and Jan, Y. N.** (1992). deadpan, an essential pan-neural gene in Drosophila, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev* **6**, 2137-51.

**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. *Dev Biol* **179**, 41-64.

**Brand, M., Jarman, A. P., Jan, L. Y. and Jan, Y. N.** (1993). asense is a Drosophila neural precursor gene and is capable of initiating sense organ formation. *Development* **119**, 1-17.

**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.

**Brenman, J. E., Gao, F. B., Jan, L. Y. and Jan, Y. N.** (2001). Sequoia, a tramtrack-related zinc finger protein, functions as a pan-neural regulator for dendrite and axon morphogenesis in Drosophila. *Dev Cell* **1**, 667-77.

**Buescher, M., Yeo, S. L., Udolph, G., Zavortink, M., Yang, X., Tear, G. and Chia, W.** (1998). Binary sibling neuronal cell fate decisions in the Drosophila embryonic central nervous system are nonstochastic and require inscuteable-mediated asymmetry of ganglion mother cells. *Genes Dev* **12**, 1858-70.

**Butler, M. J., Jacobsen, T. L., Cain, D. M., Jarman, M. G., Hubank, M., Whittle, J. R., Phillips, R. and Simcox, A.** (2003). Discovery of genes with highly restricted expression patterns in the Drosophila wing disc using DNA oligonucleotide microarrays. *Development* **130**, 659-70.

**Campuzano, S. and Modolell, J.** (1992). Patterning of the Drosophila nervous system: the achaete-scute gene complex. *Trends Genet* **8**, 202-8.

**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.

**Ceron, J., Gonzalez, C. and Tejedor, F. J.** (2001). Patterns of cell division and expression of asymmetric cell fate determinants in postembryonic neuroblast lineages of Drosophila. *Dev Biol* **230**, 125-38.

**Chagnovich, D. and Lehmann, R.** (2001). Poly(A)-independent regulation of maternal hunchback translation in the Drosophila embryo. *America* **98**, 11359-64.

**Chan, S. K., Popperl, H., Krumlauf, R. and Mann, R. S.** (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *Embo J* **15**, 2476-87.

**Chan, Y. M. and Jan, Y. N.** (1999). Conservation of neurogenic genes and mechanisms. *Curr Opin Neurobiol* **9**, 582-8.

**Chia, W. and Yang, X.** (2002). Asymmetric division of Drosophila neural progenitors. *Curr Opin Genet Dev* **12**, 459-64.

**Cornell, R. A. and Ohlen, T. V.** (2000). Vnd/nkx, ind/gsh, and msh/msx: conserved regulators of dorsoventral neural patterning? *Curr Opin Neurobiol* **10**, 63-71.

**Dawson, I. A., Roth, S., Akam, M. and Artavanis-Tsakonas, S.** (1993). Mutations of the fizzy locus cause metaphase arrest in Drosophila melanogaster embryos. *Development* **117**, 359-76.

**Doe, C. Q.** (1992). Molecular markers for identified neuroblasts and ganglion mother cells in the Drosophila central nervous system. *Development* **116**, 855-63.

**Doe, C. Q., Fuerstenberg, S. and Peng, C. Y.** (1998). Neural stem cells: from fly to vertebrates. *Journal of Neurobiology* **36**, 111-27.

**Doe, C. Q., Hiromi, Y., Gehring, W. J. and Goodman, C. S.** (1988a). Expression and function of the segmentation gene fushi tarazu during Drosophila neurogenesis. *Science* **239**, 170-5.

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

**Dubnau, J., Chiang, A. S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U. et al.** (2003). The staufen/pumilio Pathway Is Involved in Drosophila Long-Term Memory. *Curr Biol* **13**, 286-96.

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

**Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J. M.** (1999). Expression profiling using cDNA microarrays. *Nat Genet* **21**, 10-4.

**Dumur, C. I., Garrett, C. T., Archer, K. J., Nasim, S., Wilkinson, D. S. and Ferreira-Gonzalez, A.** (2004). Evaluation of a linear amplification method for small samples used on high-density oligonucleotide microarray analysis. *Analytical Biochemistry* **331**, 314-21.

**Ebens, A. J., Garren, H., Cheyette, B. N. and Zipursky, S. L.** (1993). The Drosophila anachronism locus: a glycoprotein secreted by glia inhibits neuroblast proliferation. *Cell* **74**, 15-27.

**Edgar, R., Domrachev, M. and Lash, A. E.** (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-10.

**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.

**Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A.** (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *Embo J* **13**, 3551-60.

**El-Husseini, A. E., Kwasnicka, D., Yamada, T., Hirohashi, S. and Vincent, S. R.** (2000). BERP, a novel ring finger protein, binds to alpha-actinin-4. *Biochem Biophys Res Commun* **267**, 906-11.

**Ferguson, E. L.** (1996). Conservation of dorsal-ventral patterning in arthropods and chordates. *Curr Opin Genet Dev* **6**, 424-31.

**Ferguson, E. L. and Anderson, K. V.** (1992). Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the Drosophila embryo. *Cell* **71**, 451-61.

**Frank, D. J., Edgar, B. A. and Roth, M. B.** (2002). The Drosophila melanogaster gene brain tumor negatively regulates cell growth and ribosomal RNA synthesis. *Development* **129**, 399-407.

**Frank, D. J. and Roth, M. B.** (1998). ncl-1 is required for the regulation of cell size and ribosomal RNA synthesis in Caenorhabditis elegans. *Journal of Cell Biology* **140**, 1321-9.

**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.

**Granderath, S. and Klambt, C.** (1999). Glia development in the embryonic CNS of Drosophila. *Curr Opin Neurobiol* **9**, 531-6.

**Grewal, A., Stockton, J. and Bolger, C.** (2003). Tools for discovery: gene expression enterprise solutions. *Curr Opin Drug Discov Devel* **6**, 333-8.

**Hahn, H., Wicking, C., Zaphiropoulous, P. G., Gailani, M. R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A. B., Gillies, S. et al.** (1996). Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* **85**, 841-51.

**Hanahan, D. and Weinberg, R. A.** (2000). The hallmarks of cancer. *Cell* **100**, 57-70.

**Hendricks, J. C., Finn, S. M., Panckeri, K. A., Chavkin, J., Williams, J. A., Sehgal, A. and Pack, A. I.** (2000). Rest in Drosophila is a sleep-like state. *Neuron* **25**, 129-38.

**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., Kaufman, T. C. and Reichert, H.** (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of Drosophila. *Development* **128**, 4781-8.

**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.

**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.

**Irving, P., Troxler, L., Heuer, T. S., Belvin, M., Kopczynski, C., Reichhart, J. M., Hoffmann, J. A. and Hetru, C.** (2001). A genome-wide analysis of immune responses in Drosophila. *America* **98**, 15119-24.

**Isshiki, T., Pearson, B., Holbrook, S. and Doe, C. Q.** (2001). Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511-21.

**Jan, Y. N. and Jan, L. Y.** (1998). Asymmetric cell division. *Nature* **392**, 775-8.

**Jan, Y. N. and Jan, L. Y.** (2000). Polarity in cell division: what frames thy fearful asymmetry? *Cell* **100**, 599-602.

**Jensen, K., Shiels, C. and Freemont, P. S.** (2001). PML protein isoforms and the RBCC/TRIM motif. *Oncogene* **20**, 7223-33.

**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.

**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.

**Kurzik-Dumke, U., Phannavong, B., Gundacker, D. and Gateff, E.** (1992). Genetic, cytogenetic and developmental analysis of the Drosophila melanogaster tumor suppressor gene lethal(2)tumorous imaginal discs (1(2)tid). *Differentiation* **51**, 91-104.

**Lambertsson, A.** (1998). The minute genes in Drosophila and their molecular functions. *Advances in Genetics* **38**, 69-134.

**Le Goff, G., Boundy, S., Daborn, P. J., Yen, J. L., Sofer, L., Lind, R., Sabourault, C., Madi-Ravazzi, L. and ffrench-Constant, R. H.** (2003). Microarray analysis of cytochrome P450 mediated insecticide resistance in Drosophila. *Insect Biochem Mol Biol* **33**, 701-8.

**Lee, J. E.** (1997). Basic helix-loop-helix genes in neural development. *Curr Opin Neurobiol* **7**, 13-20.

**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 using high-density oligonucleotide arrays. *America*.

**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**, RESEARCH0015.

**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.

**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 [published erratum appears in Development 1999 Feb;126(5):following table of contents]. *Development* **125**, 4471-82.

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

**Loop, T., Leemans, R., Stiefel, U., Hermida, L., Egger, B., Xie, F., Primig, M., Certa, U., Fischbach, K. F., Reichert, H. et al.** (2004). Transcriptional signature of an adult brain tumor in Drosophila. *BMC Genomics* **5**, 24.

**Lumsden, A. and Krumlauf, R.** (1996). Patterning the vertebrate neuraxis. *Science* **274**, 1109-15.

**Manak, J. R., Mitiku, N. and Lipsick, J. S.** (2002). Mutation of the Drosophila homologue of the Myb protooncogene causes genomic instability. *America* **99**, 7438-43.

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

**Mann, R. S. and Carroll, S. B.** (2002). Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev* **12**, 592-600.

**Matsuzaki, F., Ohshiro, T., Ikeshima-Kataoka, H. and Izumi, H.** (1998). miranda localizes staufen and prospero asymmetrically in mitotic neuroblasts and epithelial cells in early Drosophila embryogenesis. *Development* **125**, 4089-98.

**Michaut, L., Flister, S., Neeb, M., White, K. P., Certa, U. and Gehring, W. J.** (2003). Analysis of the eye developmental pathway in Drosophila using DNA microarrays. *America* **100**, 4024-9.

**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.

**Moore, M. S., DeZazzo, J., Luk, A. Y., Tully, T., Singh, C. M. and Heberlein, U.** (1998). Ethanol intoxication in Drosophila: Genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* **93**, 997-1007.

**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. *America* **95**, 3737-42.

**Nassif, C., Noveen, A. and Hartenstein, V.** (1998). Embryonic development of the Drosophila brain. I. Pattern of pioneer tracts. *J Comp Neurol* **402**, 10-31.

**Novotny, T., Eiselt, R. and Urban, J.** (2002). Hunchback is required for the specification of the early sublineage of neuroblast 7-3 in the Drosophila central nervous system. *Development* **129**, 1027-36.

**Ohnuma, S. and Harris, W. A.** (2003). Neurogenesis and the cell cycle. *Neuron* **40**, 199-208.

**Ohnuma, S., Philpott, A. and Harris, W. A.** (2001). Cell cycle and cell fate in the nervous system. *Curr Opin Neurobiol* **11**, 66-73.

**Parisi, M., Nuttall, R., Edwards, P., Minor, J., Naiman, D., Lu, J., Doctolero, M., Vainer, M., Chan, C., Malley, J. et al.** (2004). A survey of ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults. *Genome Biol* **5**, R40.

**Pedra, J. H., McIntyre, L. M., Scharf, M. E. and Pittendrigh, B. R.** (2004). Genome-wide transcription profile of field- and laboratory-selected dichlorodiphenyltrichloroethane (DDT)-resistant Drosophila. *America* **101**, 7034-9.

**Potter, C. J., Turenchalk, G. S. and Xu, T.** (2000). Drosophila in cancer research. An expanding role. *Trends Genet* **16**, 33-9.

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

**Reichert, H. and Boyan, G.** (1997). Building a brain: developmental insights in insects. *Trends in NeuroScience* **20**, 258-64.

**Riddle, R. D. and Tabin, C.** (1999). How limbs develop. *Scientific American* **280**, 74-9.

**Sarnatzki, Y., Luxenberg, C., Reem, E., Kimchie, Z. and Lev, Z.** (2003). The Brat brain tumor suppressor can block cell proliferation and cell differentiation by arresting the cell cycle. In *A. Dros. Res. Conf. 44*, (ed., pp. 459C.

**Saucedo, L. J. and Edgar, B. A.** (2002). Why size matters: altering cell size. *Curr Opin Genet Dev* **12**, 565-71.

**Saunders, R. D., Avides, M. C., Howard, T., Gonzalez, C. and Glover, D. M.** (1997). The Drosophila gene abnormal spindle encodes a novel microtubule-associated protein that associates with the polar regions of the mitotic spindle. *Journal of Cell Biology* **137**, 881-90.

**Scaerou, F., Aguilera, I., Saunders, R., Kane, N., Blottiere, L. and Karess, R.** (1999). The rough deal protein is a new kinetochore component required for accurate chromosome segregation in Drosophila. *Journal of Cell Science* **112 ( Pt 21)**, 3757-68.

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

**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.

**Schubert, F. R., Nieselt-Struwe, K. and Gruss, P.** (1993). The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *America* **90**, 143-7.

**Shen, C. P., Knoblich, J. A., Chan, Y. M., Jiang, M. M., Jan, L. Y. and Jan, Y. N.** (1998). Miranda as a multidomain adapter linking apically localized Inscuteable and basally localized Staufen and Prospero during asymmetric cell division in Drosophila. *Genes Dev* **12**, 1837-46.

**Simone, N. L., Bonner, R. F., Gillespie, J. W., Emmert-Buck, M. R. and Liotta, L. A.** (1998). Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet* **14**, 272-6.

**Skeath, J. B. and Doe, C. Q.** (1998). Sanpodo and Notch act in opposition to Numb to distinguish sibling neuron fates in the Drosophila CNS. *Development* **125**, 1857-65.

**Skeath, J. B. and Thor, S.** (2003). Genetic control of Drosophila nerve cord development. *Curr Opin Neurobiol* **13**, 8-15.

**Sonoda, J. and Wharton, R. P.** (2001). Drosophila Brain Tumor is a translational repressor. *Genes Dev* **15**, 762-73.

**Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M. and Levine, M.** (2002). Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. *Cell* **111**, 687-701.

**Steward, R.** (1987). Dorsal, an embryonic polarity gene in Drosophila, is homologous to the vertebrate proto-oncogene, c-rel. *Science* **238**, 692-4.

**Tenev, T., Zachariou, A., Wilson, R., Paul, A. and Meier, P.** (2002). Jafrac2 is an IAP antagonist that promotes cell death by liberating Dronc from DIAP1. *Embo J* **21**, 5118-29.

**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.

**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.

**Torok, M. and Etkin, L. D.** (2001). Two B or not two B? Overview of the rapidly expanding B-box family of proteins. *Differentiation* **67**, 63-71.

**Truman, J. W. and Bate, M.** (1988). Spatial and temporal patterns of neurogenesis in the central nervous system of Drosophila melanogaster. *Dev Biol* **125**, 145-57.
**Udolph, G., Urban, J., Rusing, G., Luer, K. and Technau, G. M.** (1998). Differential effects of EGF receptor signalling on neuroblast lineages along the dorsoventral axis of the Drosophila CNS. *Development* **125**, 3291-9.
**Urbach, R. and Technau, G. M.** (2003). Molecular markers for identified neuroblasts in the developing brain of Drosophila. *Development* **130**, 3621-37.
**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.
**von Ohlen, T. and Doe, C. Q.** (2000). Convergence of dorsal, dpp, and egfr signaling pathways subdivides the drosophila neuroectoderm into three dorsal-ventral columns. *Dev Biol* **224**, 362-72.
**Wechsler-Reya, R. and Scott, M. P.** (2001). The developmental biology of brain tumors. *Annu Rev Neurosci* **24**, 385-428.
**Wharton, K. A., Ray, R. P. and Gelbart, W. M.** (1993). An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the Drosophila embryo. *Development* **117**, 807-22.
**White, K. P., Rifkin, S. A., Hurban, P. and Hogness, D. S.** (1999). Microarray analysis of Drosophila development during metamorphosis. *Science* **286**, 2179-84.
**Woodhouse, E., Hersperger, E. and Shearn, A.** (1998). Growth, metastasis, and invasiveness of Drosophila tumors caused by mutations in specific tumor suppressor genes. *Dev Genes Evol* **207**, 542-50.
**Xie, J., Murone, M., Luoh, S. M., Ryan, A., Gu, Q., Zhang, C., Bonifas, J. M., Lam, C. W., Hynes, M., Goddard, A. et al.** (1998). Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90-2.
**Xiong, W. C. and Montell, C.** (1995). Defective glia induce neuronal apoptosis in the repo visual system of Drosophila. *Neuron* **14**, 581-90.
**Younossi-Hartenstein, A., Nassif, C., Green, P. and Hartenstein, V.** (1996). Early neurogenesis of the Drosophila brain. *J Comp Neurol* **370**, 313-29.
**Zhu, Y. and Parada, L. F.** (2002). The molecular and genetic basis of neurological tumours. *Nat Rev Cancer* **2**, 616-26.

## Acknowledgements

My sincere gratitude goes to my parents for their love, support, and patience over the last few years.

Also, I wish to heartily thank my supervisors Heinrich Reichert and Karl-Friedrich Fischbach for their help on many aspects of my scientific endeavors.

The effort and good will of many people have enabled completion of this dissertation. Among these I especially thank Frank Hirth, Lars Kammermeier, Ronny Leemans and Urs Stiefel for their continuing support of my work and for being enjoyable colleagues.
Of course, this applause has to be widened to all members of the Reichert- and Fischbach lab and to my collaborators on the various projects, as all those people made my Ph. D. thesis a fun business.

ERKLÄRUNG

Ich erkläre, dass ich die Dissertation

Functional genomics of brain development and developmentally related brain disease in Drosophila

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

Basel, den 7. September 2004

Thomas Loop

**Curriculum Vitae**

**Thomas Loop**

**Personal data:**

| | |
|---|---|
| Date of birth: | 11th February 1968 |
| Place of birth: | Hamburg, Germany |
| Nationality: | German |
| Marital status: | Unmarried |
| Home Address: | Gluckstrasse 9, D-79104 Freiburg, Germany |

**Education:**

| | |
|---|---|
| 1979–1988 | Elsa-Brändström-Schule Elmshorn, Germany |
| 1988 | Abitur in Elmshorn, Germany |
| 1990–1999 | Undergraduate studies in Biology at the Albert-Ludwigs-University Freiburg. Main subjects: Neurobiology, Genetics, Zoology, Informatics. |
| 1999 | Diploma in Biology at the Albert-Ludwigs-University Freiburg |
| 2004 | PhD in Zoology supervised by Prof. Dr. Heinrich Reichert, University of Basel and Prof. Dr. Karl-Friedrich Fischbach, University of Freiburg |

## Research Experience:

| | |
|---|---|
| 2000–2001 | Diploma work in Neurobiology: "*Characterisation of homeotic gene action in embryonic development of Drosophila*" supervised by Prof. Dr. Karl-Friedrich |

Fischbach, Institute of Biology III, University of Freiburg and co-supervised by Prof. Dr. Heinrich Reichert, Institute of Zoology, University of Basel.

2001-2004 PhD thesis: *"Functional genomics of brain development and developmentally related brain disease in Drosophila"*, supervised by Prof. Dr. Heinrich Reichert, Institute of Zoology, University of Basel.

## Teaching Experience:

2003 Organization and supervision of a practical student course in *Drosophila* biology, Institute of Zoology, Basel

2001-2004 Co-supervision of diploma and PhD students

2002 Supervisor in the ELTEM Summer School of Neurogenomics, Basel

## International Meetings:

### Oral presentations:

1st Neurex ELTEM Meeting on *Drosophila* Neurobiology, 2002, Engelberg, Switzerland.

### Poster presentations:

16th European *Drosophila* Research Conference, Zürich, Switzerland, 1999.

33rd USGEB meeting, Lugano, Switzerland, 2002.

9th European Symposium on *Drosophila* Neurobiology, Dijon, France, 2002.

Neurex Annual Meeting 2003 "Neurogenetics: from the molecule to man", Basel, Switzerland.

18th European *Drosophila* Research Conference, Göttingen, Germany, 2003.

Neurex Annual Meeting 2004 "Neuronal Connectivity: Development and Function", Freiburg, Germany.

## List of Publications


Hirth F., Therianos S., **Loop T.**, Gehring W.J., Reichert H., 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.

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, research 0015.1-0015.9.

Hirth F., **Loop T.**, Egger B., Miller D. F., Kaufman T. C. and Reichert H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of *Drosophila*. *Development* 128, 4781-8.

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.

Montalta-He H., Leemans R., **Loop T.**, Strahm M., Certa U., Primig M., Acampora D., Simeone A., Reichert H. (2002). Evolutionary conservation of *otd*/*Otx2* transcription factor action: a genome-wide microarray analysis in *Drosophila*. *Genome Biol.* 3, research0015.1-0015.15.

**Loop T.**, Leemans R., Stiefel U., Hermida L., Egger B., Xie F., Primig M., Certa U., Fischbach K.F., Reichert H., Hirth F. (2004). Transcriptional signature of an adult brain tumor in *Drosophila*. *BMC Genomics* 5, 24.

## Curriculum Vitae

## Thomas Loop

**Personal data:**

| | |
|---|---|
| Date of birth: | 11th February 1968 |
| Place of birth: | Hamburg, Germany |
| Nationality: | German |
| Marital status: | Unmarried |
| Home Address: | Gluckstrasse 9, D-79104 Freiburg, Germany |

**Education:**

| | |
|---|---|
| 1979–1988 | Elsa-Brändström-Schule Elmshorn, Germany |
| 1988 | Abitur in Elmshorn, Germany |
| 1990–1999 | Undergraduate studies in Biology at the Albert-Ludwigs-University Freiburg. Main subjects: Neurobiology, Genetics, Zoology, Informatics. |
| 1999 | Diploma in Biology at the Albert-Ludwigs-University Freiburg |
| 2004 | PhD in Zoology supervised by Prof. Dr. Heinrich Reichert, University of Basel and Prof. Dr. Karl-Friedrich Fischbach, University of Freiburg |

## Research Experience:

2000–2001 Diploma work in Neurobiology: "*Characterisation of homeotic gene action in embryonic development of Drosophila*" supervised by Prof. Dr. Karl-Friedrich Fischbach, Institute of Biology III, University of Freiburg and co-supervised by Prof. Dr. Heinrich Reichert, Institute of Zoology, University of Basel.

2001-2004 PhD thesis in Neurobiology: *"Functional genomics of brain development and developmentally related brain disease in Drosophila"*, supervised by Prof. Dr. Heinrich Reichert, Institute of Zoology, University of Basel.

## Teaching Experience:

2003 Organization and supervision of a practical student course in *Drosophila* biology, Institute of Zoology, Basel

2001-2004 Co-supervision of diploma and PhD students

2002 Supervisor in the ELTEM Summer School of Neurogenomics, Basel

## International Meetings:

### Oral presentations:

1st Neurex ELTEM Meeting on *Drosophila* Neurobiology, 2002, Engelberg, Switzerland.

### Poster presentations:

16th European *Drosophila* Research Conference, Zürich, Switzerland, 1999.

33rd USGEB meeting, Lugano, Switzerland, 2002.

9th European Symposium on *Drosophila* Neurobiology, Dijon, France, 2002.

Neurex Annual Meeting 2003 “Neurogenetics: from the molecule to man”, Basel, Switzerland.
18th European *Drosophila* Research Conference, Göttingen, Germany, 2003.
Neurex Annual Meeting 2004 “Neuronal Connectivity: Development and Function”, Freiburg, Germany.

## List of Publications

Hirth F., Therianos S., **Loop T.**, Gehring W.J., Reichert H., 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.

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, research 0015.1-0015.9.

Hirth F., **Loop T.**, Egger B., Miller D. F., Kaufman T. C. and Reichert H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of *Drosophila*. *Development* 128, 4781-8.

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.

Montalta-He H., Leemans R., **Loop T.**, Strahm M., Certa U., Primig M., Acampora D., Simeone A., Reichert H. (2002). Evolutionary conservation of *otd*/*Otx2* transcription factor action: a genome-wide microarray analysis in *Drosophila*. *Genome Biol.* 3, research0015.1-0015.15.

**Loop T.**, Leemans R., Stiefel U., Hermida L., Egger B., Xie F., Primig M., Certa U., Fischbach K.F., Reichert H., Hirth F. (2004). Transcriptional signature of an adult brain tumor in *Drosophila*. *BMC Genomics* 5, 24.