Research

Expression profiling of *Drosophila* imaginal discs

Ansgar Klebes*, Brian Biehs*, Francisco Cifuentes*† and Thomas B Kornberg*

Addresses: *Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA. †Current address: Agilent Inc., 395 Page Mill Rd, Palo Alto, CA 94303, USA.

Correspondence: Thomas B Kornberg. E-mail: tkornberg@biochem.ucsf.edu

Published: 24 July 2002

Genome ***Biology*** 2002, **3(8)**:research0038.1–0038.16

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/8/research/0038

Received: 10 May 2002
Revised: 7 June 2002
Accepted: 7 June 2002

(Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

Background: In the *Drosophila* larva, imaginal discs are programmed to produce adult structures at metamorphosis. Although their fate is precisely determined, these organs remain largely undifferentiated in the larva. To identify genes that establish and express the different states of determination in discs and larval tissues, we used DNA microarrays to analyze mRNAs isolated from single imaginal discs.

Results: Linear amplification protocols were used to generate hybridization probes for microarray analysis from poly(A)$^+$ RNA from single imaginal discs containing between 10,000 and 60,000 cells. Probe reproducibility and degree of representation were tested using microarrays with approximately 6,000 different cDNAs. Hybridizations with probes that had been prepared separately from the same starting RNA pool had a correlation coefficient of 0.97. Expression-profile comparisons of the left and right wing imaginal discs from the same larva correlated with a coefficient of 0.99, indicating a high degree of reproducibility of independent amplifications. Using this method, we identified genes with preferential expression in the different imaginal discs using pairwise comparisons of discs and larval organs. Whereas disc-to-disc comparisons revealed only moderate differences, profiles differed substantially between imaginal discs and larval tissues, such as larval endodermal midgut and mesodermal fat body.

Conclusion: The combination of linear RNA amplification and DNA microarray hybridization allowed us to determine the expression profiles of individual imaginal discs and larval tissues and to identify genes expressed in tissue-specific patterns. These methods should be widely applicable to comparisons of expression profiles for tissues or parts of tissues that are available only in small amounts.

Background

During the development of multicellular organisms, complexity builds sequentially in discrete steps as cells proliferate and their descendants choose between alternative developmental fates. Much of our understanding of these processes in *Drosophila* has come from mutants that have developmental defects. There are many examples. For instance, mutations in the homeodomain protein genes of the Antennapedia and Bithorax complexes provide evidence for the role of transcription factors in making developmental decisions. Flies with gain-of-function mutations in the *Antennapedia* (*Antp*) gene cannot grow normal antennae,

but instead make extra legs with their antennal cells. Conversely, loss of Antp function causes transformation of the second leg to antenna-like structures [1]. Another example is the *glass* gene, which encodes a zinc-finger transcription factor that is expressed in many of the cell types in the eye imaginal disc and is required for photoreceptor development [2]. The common functions of these and many other genes in regulating gene expression suggest that developmental fates are manifested in part in the transcripts that different cell types produce.

Although the number of *Drosophila* mutants with interesting developmental phenotypes is large, we cannot assume that the genetic screens that have been carried out identified all the relevant genes. Many *Drosophila* genes have been refractory to genetic analysis, either because they are duplicated and may code for redundant functions, or because they have mutant phenotypes that are difficult to recognize. In addition, genetic and biochemical approaches have yielded few downstream genes that the key transcription factors regulate. The recent release of the *Drosophila* genome sequence [3] and the demonstration that cDNA microarrays can be used to catalog the transcriptome of both unicellular and multicellular organisms [4-20] opens up a new approach. Hybridization of mRNA pools to DNA microarrays can potentially identify the differences in gene activity that define every developmental state. Thus, transcriptional regulators and their downstream targets can be identified simultaneously.

The *Drosophila* larva contains two developmentally distinct sets of organs. One set comprises the vital organs of the larva - the epidermis and musculature, the central and peripheral nervous systems, and the organs of the digestive tract. Most of these tissues will be destroyed at metamorphosis and will be replaced by imaginal (adult) cells. The imaginal precursor cells have no functional role in the larva. For some adult structures such as the eyes, antennae, legs, wings, halteres and genitalia, the imaginal precursors are sequestered as physically distinct primordia - the imaginal discs. The different discs are determined as to disc type, but do not differentiate until the late larval and subsequent pupal stages when they form the adult structures. The distinct developmental phases of the imaginal discs, their easy accessibility, along with the numerous advantages of *Drosophila* as a model genetic and developmental organism make imaginal disc development an ideal system to explore the feasibility and usefulness of such an analysis at a genomic scale.

The principal technical difficulty such a study must surmount is that the quantity of probe needed for efficient hybridization to microarrays is large relative to the amount of RNA contained in a single *Drosophila* imaginal disc. Approximately 4 μg poly(A)$^+$ RNA is required to make a probe, but a third instar wing imaginal disc has only about 1-4 ng (Figure 1). Thus, several thousand discs would need to be dissected for each set of experiments. In previous reports of expression profiling, *Drosophila* embryos, adult flies and adult heads, and collections of animals were pooled after carefully timing or otherwise selecting individual animals [9,14,18,20]. However, any scheme to stage animals is inherently inexact. In addition, parameters such as genetic variability, nutritional state, pathogen exposure and effects of the isolation procedure cannot be easily controlled or measured. Pools of animals therefore yield only an average of their varied inputs, and depending on the extent of variability,

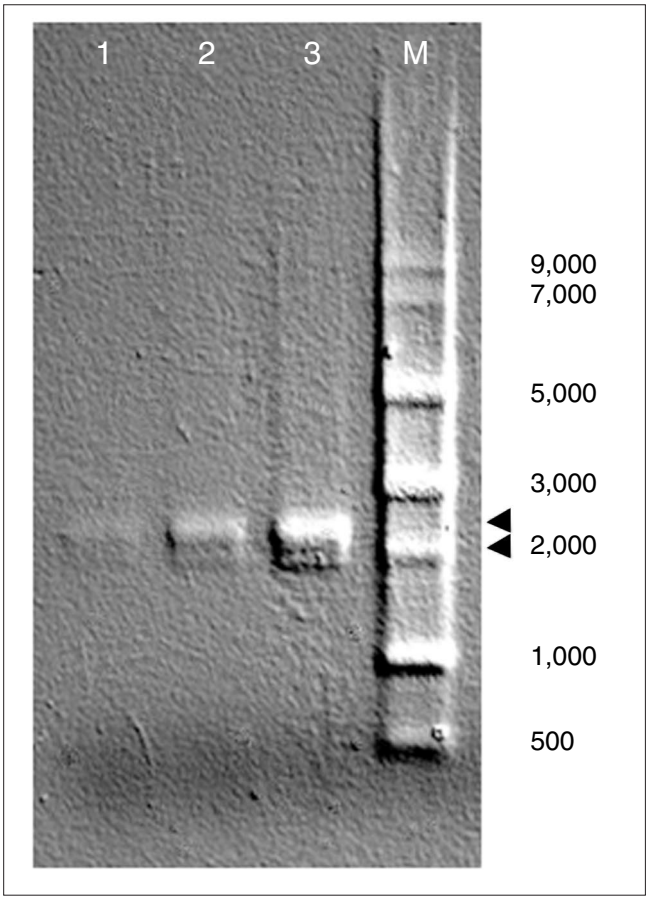


Figure 1
Quality and amount of total RNA preparations from imaginal discs. Total RNA was prepared separately from 2, 4 and 12 third-instar larval wing imaginal discs using the Mini RNA Isolation Kit (Zymo Research). The total amount of RNA, based on the absorbance at 260 nm (A_{260}), was 462, 540 and 1,530 ng, respectively. The A_{260}/A_{280} ratios of 2.08, 2.2 and 2.27 for the separate preparations is indicative of high-quality RNA preparations. The amount of total RNA per disc was calculated to be 230, 140 and 130 ng, respectively. Assuming a poly(A)$^+$ RNA content of 1-2%, this amount of total RNA roughly corresponds to 1-4 ng poly(A)$^+$ RNA per wing disc (62,000 cells). The other discs are smaller and contain fewer cells. The equivalent of 1, 2 and 6 discs was separated on a denaturing agarose gel (lanes 1, 2 and 3, respectively). The two prominent bands represent the 18S and 28S ribosomal RNA populations (arrowheads). Poly(A)$^+$ RNA is detected as a smear. No obvious small-molecular-weight products were observed. Lane M contains molecular-weight markers, with numbers indicating the approximate lengths in nucleotides.

critical differences between samples may be submerged. Furthermore, the level of resolution is relatively low when RNA preparations from whole animals are used. For many biological questions, it is necessary to detect differences in subsets of cells that might account for only a small fraction of the many thousands of cells in the animal. To circumvent these problems, we used techniques that allow us to carry out several hybridizations with probes derived from different tissues of a single animal.

We applied linear RNA amplification methods that were first introduced by Eberwine and co-workers [21,22] and have been refined by Baugh *et al.* [23]. We used two rounds of reverse transcription (RT) and *in vitro* transcription (IVT), and achieved as much as 1-5 x 10^5-fold amplification. The amplified RNA (aRNA) from different tissues of single larva was used to perform pairwise comparisons and to identify sets of genes with preferential expression. The preferential expression of many of the genes in these sets had been previously shown by genetic or molecular analysis. We also confirmed the preferential expression of ten genes that had not been previously characterized by *in situ* hybridization to imaginal discs. Several of these genes had particularly interesting patterns of expression. Our findings show that this method can be applied to determine the expression profile of individual tissues and to identify candidate genes with specific expression patterns.

Results and discussion

Principle of amplification

The amplification procedure we used is based on protocols described by Eberwine [21,22,24], Wang *et al.* [25], and Baugh *et al.* [23], and details are given in Materials and methods. Briefly, double-stranded cDNA was synthesized either from total RNA isolated from imaginal discs or from poly(A)$^+$-purified embryonic RNA. The cDNA was used as a template for *in vitro* transcription, routinely generating an RNA product that represented an amplification of approximately 1,000 fold. This RNA was subjected to a second round of RT and IVT. The yield from the second round represented an approximately 100-fold amplification, resulting in an overall amplification of 1-5 x 10^5. Probes for hybridization were synthesized by RT and labeled with either Cy5 (red) or Cy3 (green). Arrays were hybridized simultaneously with Cy5- and Cy3-labeled probes, and the intensities and ratios of bound fluor were measured for each spot. Because of inherent variability in the hybridization efficiency of individual microarrays, only the ratios of the red and green fluorescence were interpreted to indicate relative transcript abundance in the starting pool of poly(A)$^+$ RNA.

A known drawback of linear amplification is that sequences at the 5´ end of transcripts can be preferentially lost. This reduction in sequence complexity can probably be attributed to incomplete RT, to priming from subterminal locations, and to incomplete IVT. To overcome this problem, we supplemented the random hexanucleotides that were used to prime the second RT reaction with a template-switch (TS) primer [25]. This primer includes a guanosine triplet that is designed to pair with a 3´ cytidine overhang created by the reverse transcriptase (Clontech [25]). Control amplifications with the TS primer alone or random hexanucleotides alone yielded similar quantities of aRNA. When these aRNA preparations were used for microarray hybridizations, many spots had significantly different signal intensities (data not shown). The 25 cDNAs with the most intense signals for the probe amplified with the TS primer alone had an average length of 1.6 kilobases (kb). In contrast, after amplification with hexanucleotides alone, the 25 cDNAs with the most intense signals had an average length of 1.1 kb. We interpret the bias towards long transcripts in TS-primer amplified probes as indicating better preservation of 5´-complexity.

Representation of transcripts in amplified probes

The most important feature of RNA pools that amplification must preserve is proportionality, the maintenance of the relative concentration of different RNAs. We established the quality of our amplification method in several ways. First, we isolated poly(A)$^+$ RNA from embryos and subjected 100 pg, 1 ng and 10 ng aliquots to two rounds of linear amplification. The size distribution of the aRNA ranged from about 100 nucleotides to high-molecular-weight products, with the majority between 240 and 2,000 nucleotides (Figure 2a). aRNA from independent amplifications was then used to generate probes and to compare hybridization efficiencies of pairs of probes. Such separate amplifications of aliquots of 1 ng of the same embryonic poly(A)$^+$ RNA produced signals with similar intensities. A scatterplot of the Cy3 and Cy5 signal intensities showed a tight correlation coefficient (CC) of 0.97, with a low standard deviations (SD) of the red/green (R/G) ratios for all spots (SD 0.25, $\log_2$-transformed) (Figure 2b). Comparisons of probes made from 100 pg and 1 ng and from 1 ng and 10 ng poly(A)$^+$ RNA yielded essentially the same level of reproducibility (CC 0.95, SD 0.29; data not shown). We conclude that the amplifications were reproducible and that good proportionality of RNA was maintained during amplification even if the concentration of poly(A)$^+$ RNA differed by tenfold.

Comparisons between amplified and unamplified RNA indicated that essentially no transcripts were lost during amplificaton. Of the spots that hybridized to probes generated from amplified and unamplified embryonic poly(A)$^+$ RNA 99% (5,514 out of 5,574) did so with both probes. Differences in intensities were, nevertheless, significant (CC 0.53 and SD 1.13, data not shown). We do not know whether these differences in intensities were a result of changes in relative abundance or in transcript length, both of which will affect hybridization signals. Nevertheless, the good correspondence of the hybridizing spots shows that the aRNA is an essentially complete representation of the population of

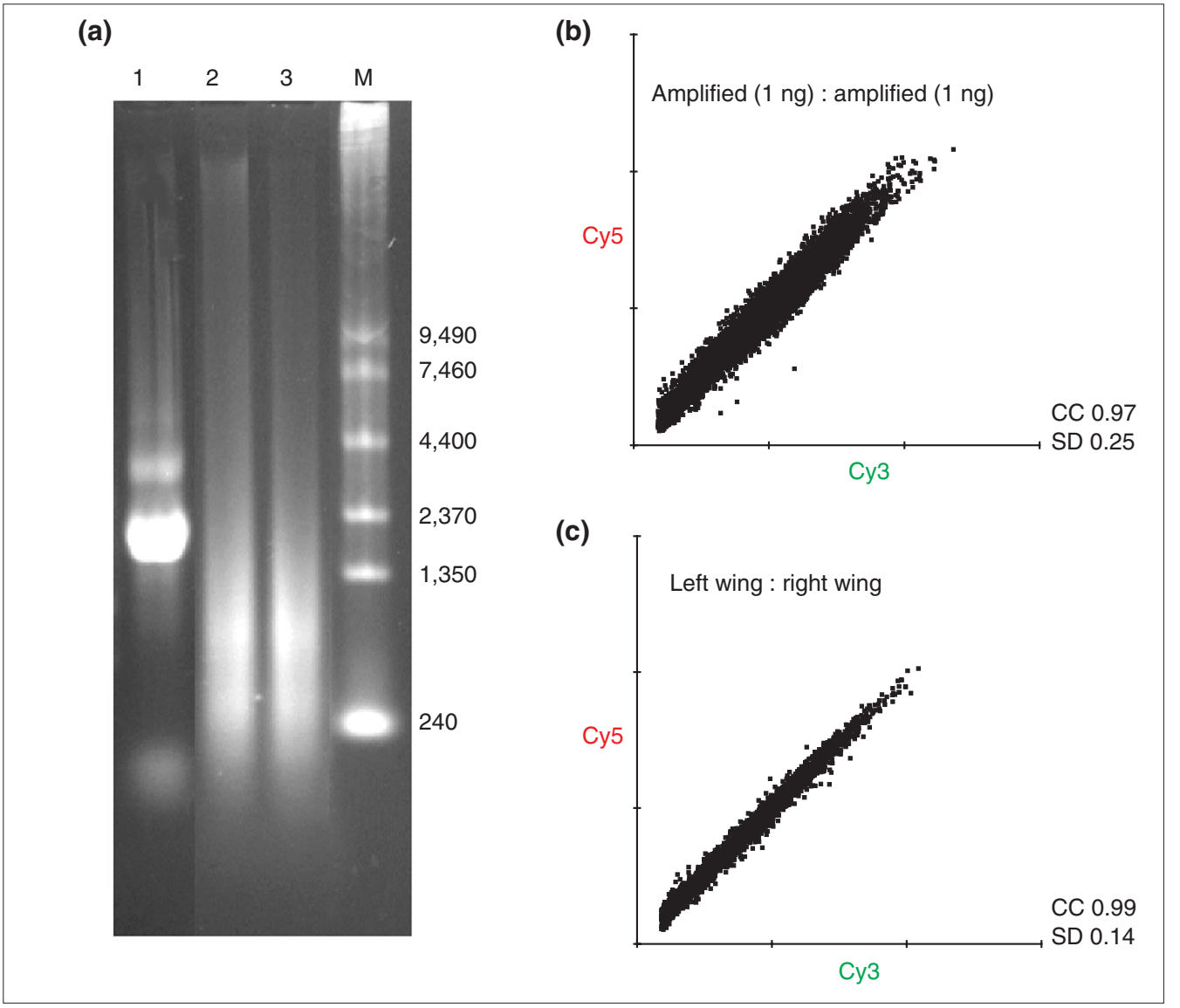


Figure 2
Linear amplification is highly reproducible. **(a)** Denaturing gel electrophoresis shows the size distribution of aRNA obtained after two rounds of linear amplification of 1 ng poly(A)$^+$ RNA. Lane 1, total RNA; lanes 2 and 3, products of independent amplifications; M, molecular-weight markers, with numbers indicating the approximate lengths in nucleotides. **(b,c)** Scatterplots of the Cy3 and Cy5 signal intensities from hybridizations of two probes derived from independent amplifications of embryonic or wing-disc RNA. Starting materials were (b) 1 ng embryonic poly(A)$^+$ RNA and (c) left and right wing imaginal discs of one wandering third-instar larva. CC, correlation coefficient; SD, standard deviation. SDs were calculated on the normalized $\log_2$-transformed R/G ratios.

transcripts present in the initial pool. We conclude that comparisons of two different probes on microarrays are valid if the RNA pools have been amplified in the same way.

Pairwise comparisons of imaginal discs and larval organs

We carried out 43 separate hybridizations, comparing 13 different pairwise combinations of tissues (Table 1). This analysis identified many genes that are preferentially expressed in specific tissues; it also provided an independent and quantitative measure of the distinct nature of larval tissue and imaginal disc cells.

Hybridizations that were carried out were of two types: with probes generated from different imaginal discs; and with probes from imaginal discs and larval organs. Comparisons between imaginal disc probes revealed a high degree of similarity, but comparisons between imaginal disc and larval organ probes identified many differentially expressed genes. We now give a general description of these results, starting

Table 1

Overview of the comparisons

Comparison	Number of experiments
Wing : wing	1
Wing : eye-antenna	11
Wing : leg1	2
Leg1 : leg2	3
Leg1 : leg3	2
Leg2 : leg3	2
Haltere : genital	5
Wing : fat body	**5**
Leg1 : fat body	**2**
Haltere : fat body	**1**
Wing : brain/optic lobe	**3**
Wing : salivary gland	**3**
Wing : midgut	**3**

Forty-three pairwise comparisons were carried out. Of these, 26 were between imaginal discs; 17 were between an imaginal disc and a larval organ (indicated in bold). The numbers indicate the times each type of comparison was repeated.

with those tissues that were most alike. It should be noted that these results were obtained with microarrays that represent a pool of cDNAs obtained from the Berkeley *Drosophila* Genome Project (BDGP) containing approximately 44% (6,000/13,600) of the transcription units predicted to be in the *Drosophila* genome. As this pool may not be a truly random subset, extrapolations to the whole genome may be of limited value. Indeed, a preliminary comparison of our cDNA arrays with microarrays that contain short cDNA sequences for all annotated genes suggests that the total number of genes that produce a signal with these probes is similar. It is possible that the set of 6,000 cDNAs distributed by the BDGP is biased toward genes expressed in the tissues we analyzed.

Expression profiles of imaginal discs

Left and right wing discs

We first asked whether microarray hybridization could detect differences in expression profiles between left and right wing discs. Although few differences might be expected, previous studies have revealed a left/right difference in wing size, suggesting that the respective developmental programs are not identical [26]. When probes were prepared independently from the poly(A)$^+$ RNA isolated from left and right wing discs of a third instar larva, only minor differences in the intensities were observed (CC 0.99 and SD 0.14) (Figure 2c). This tight correlation again shows the reproducibility of the amplification procedure.

Leg discs

We then asked if any of the genes represented on our arrays were differentially expressed in first, second or third leg discs. Among the several genes that have been shown to be differentially required or expressed in the different leg discs (for example, *Ultrabithorax* [27], *Antp* [28] and *Sex combs reduced* [29,30]), only *Antp* was among the set of cDNAs on our microarrays. Comparisons of probes from first, second and third leg discs yielded CC values in the 0.93 ranges. To identify differentially expressed genes in the leg discs, cluster analysis was used to group genes with similar expression profiles [31]. For this and all subsequent analyses a threshold setting of > 1.74 (= 0.8 $\log_2$-transformed ratios) was applied to identify subclusters. We identified 2, 12 or 17 differentially expressed genes in the leg1:leg3, leg2:leg3 and leg1:leg2 disc arrays, respectively (Figure 3). The induced expression of *Antp* in the second leg discs was noteworthy. This finding is consistent with previous studies showing that although *Antp* is expressed in all three leg discs, it is expressed at a much greater level in second leg discs [28].

Leg and wing discs

Comparisons were also made between the first leg and wing imaginal discs (two experiments). Cluster analysis identified 23 genes that were preferentially expressed in the leg discs and 8 in the wing discs (Figure 4). Notable in the wing cluster are *broad*, which is expressed in imaginal cells and encodes a family of transcription factors [32], and *apterous*, which is expressed in the dorsal wing cells and is required for wing formation [33].

Wing and eye-antennal imaginal discs

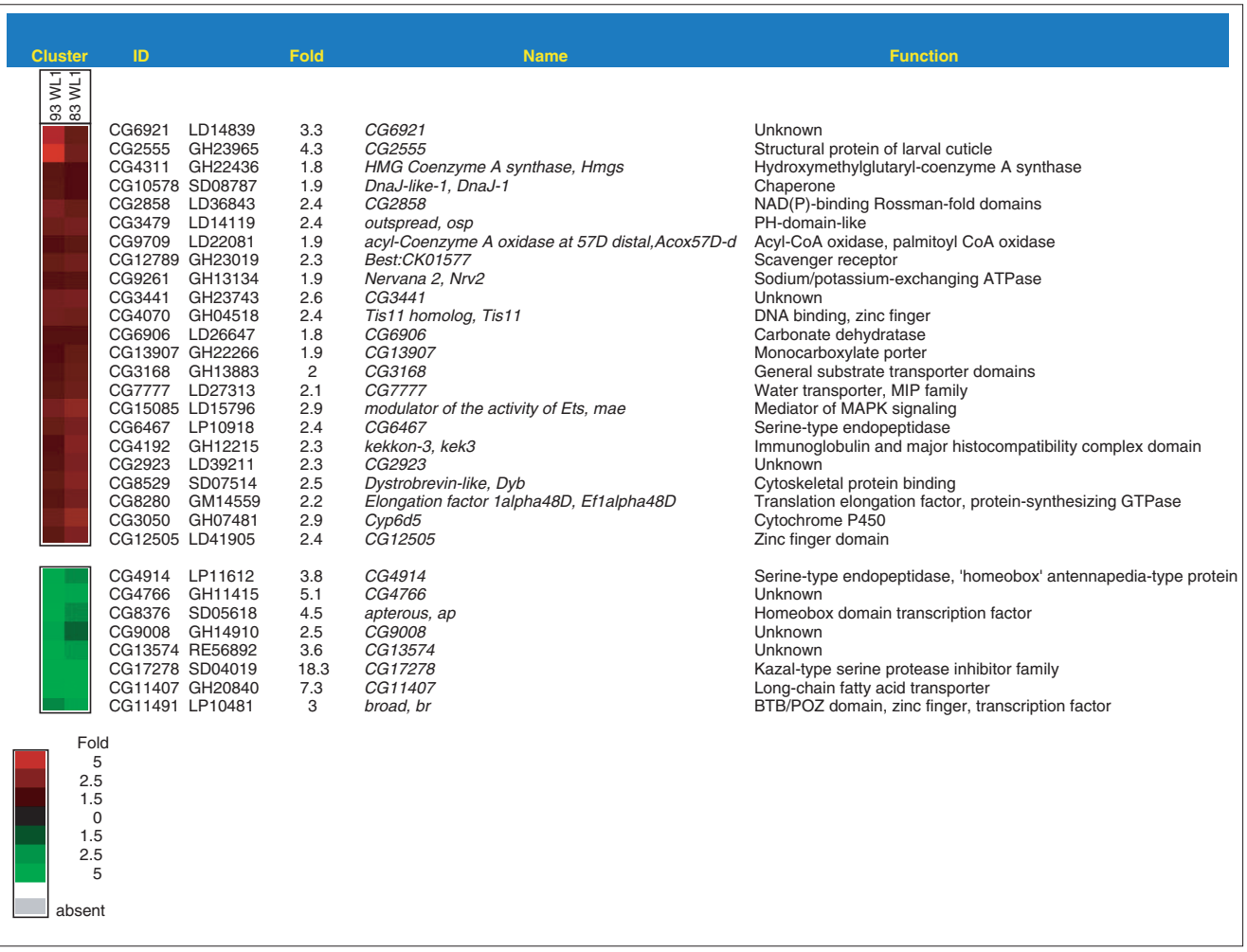
Comparing probes from the wing and eye-antennal discs isolated from the same larva revealed a higher degree of divergence. Eleven independent wing:eye-antennal comparisons were made. Twenty-four genes were identified that had elevated expression in the wing discs, and 73 genes were identified with elevated expression in eye-antennal discs (Figure 5). Many of these genes are known from previous studies to be expressed specifically in these tissues. Among the genes with elevated expression in wing discs were *apterous* and *engrailed*. *apterous* is expressed in the dorsal compartment of the wing disc, but not in the eye-antennal disc; *engrailed* is expressed in the posterior compartment cells of both the wing and eye-antennal discs. However, inclusion of *engrailed* in the wing group is reasonable as the posterior compartment represents approximately half of the wing disc, but is less than a quarter of the eye-antennal disc. Among the genes with increased expression in eye-antennal discs were *white*, which encodes an eye pigment precursor transporter and the eye-specific transcription factor-encoding gene, *glass*. Consistent with the neuronal fate of many cells in the eye-antennal disc, ten genes known to be expressed and to function in neuronal or glial cells were grouped into the eye-antennal cluster: *prospero*, *atonal*, *fasciclin I, longitudinal lacking*, *locomotion defects*, *Rapgap1*, *unc-13*, *sanpodo*, *Caps* and *beta-amyloid protein precursor-like*.

In addition to the genes whose tissue-specific expression was confirmed by this array analysis, many have not been

Figure 3

Pairwise comparisons of prothoracic, mesothoracic and metathoracic leg discs (leg1, leg2 and leg3, respectively). Cluster analysis was carried out on the dataset with the requirement to show induction > 1.74 (0.8 of the $\log_2$-transformed R/G ratio). In the color representation of the cluster results in this and the subsequent figures, the columns represent the different experiments and the rows indicate the genes. Seventeen genes were found to be differentially expressed between the first (ten genes, green cluster) and second leg discs (seven genes, red cluster) in three independently repeated experiments (numbers 91, 77 and 142). The comparison of first and third leg discs (two experiments, numbers 92 and 94) produced two genes in the leg1 cluster (green). Two experiments (numbers 106 and 81) revealed 12 genes that are differentially expressed between leg2 disc (three genes, green cluster) and leg3 disc (nine genes, red cluster). Note the expression of *Antp* in the second leg disc in the leg1-to-leg2 and leg2-to-leg3 comparisons, with more than threefold induction in both cases (see text). The columns indicate the subclusters with consistent induction in one channel, the gene identification numbers (ID), the average fold induction of the two or three comparisons (Fold), the gene name and function as published on Flybase [44]. The color code is indicated below with the numbers representing the fold induction.

characterized previously. One of these novel genes, CG9335, was notable for its 16-fold relative induction in eye-antennal discs (Figure 5). We also single out *Gliolectin (glec)* in the eye-antennal cluster, as this is not consistent with previous studies. *glec* is expressed in midline glia cells, but previous *in situ* hybridization and immunohistochemistry failed to



Cluster	ID		Fold	Name	Function
	CG6921	LD14839	3.3	*CG6921*	Unknown
	CG2555	GH23965	4.3	*CG2555*	Structural protein of larval cuticle
	CG4311	GH22436	1.8	*HMG Coenzyme A synthase, Hmgs*	Hydroxymethylglutaryl-coenzyme A synthase
	CG10578	SD08787	1.9	*DnaJ-like-1, DnaJ-1*	Chaperone
	CG2858	LD36843	2.4	*CG2858*	NAD(P)-binding Rossman-fold domains
	CG3479	LD14119	2.4	*outspread, osp*	PH-domain-like
	CG9709	LD22081	1.9	*acyl-Coenzyme A oxidase at 57D distal,Acox57D-d*	Acyl-CoA oxidase, palmitoyl CoA oxidase
	CG12789	GH23019	2.3	*Best:CK01577*	Scavenger receptor
	CG9261	GH13134	1.9	*Nervana 2, Nrv2*	Sodium/potassium-exchanging ATPase
	CG3441	GH23743	2.6	*CG3441*	Unknown
	CG4070	GH04518	2.4	*Tis11 homolog, Tis11*	DNA binding, zinc finger
	CG6906	LD26647	1.8	*CG6906*	Carbonate dehydratase
	CG13907	GH22266	1.9	*CG13907*	Monocarboxylate porter
	CG3168	GH13883	2	*CG3168*	General substrate transporter domains
	CG7777	LD27313	2.1	*CG7777*	Water transporter, MIP family
	CG15085	LD15796	2.9	*modulator of the activity of Ets, mae*	Mediator of MAPK signaling
	CG6467	LP10918	2.4	*CG6467*	Serine-type endopeptidase
	CG4192	GH12215	2.3	*kekkon-3, kek3*	Immunoglobulin and major histocompatibility complex domain
	CG2923	LD39211	2.3	*CG2923*	Unknown
	CG8529	SD07514	2.5	*Dystrobrevin-like, Dyb*	Cytoskeletal protein binding
	CG8280	GM14559	2.2	*Elongation factor 1alpha48D, Ef1alpha48D*	Translation elongation factor, protein-synthesizing GTPase
	CG3050	GH07481	2.9	*Cyp6d5*	Cytochrome P450
	CG12505	LD41905	2.4	*CG12505*	Zinc finger domain
	CG4914	LP11612	3.8	*CG4914*	Serine-type endopeptidase, 'homeobox' antennapedia-type protein
	CG4766	GH11415	5.1	*CG4766*	Unknown
	CG8376	SD05618	4.5	*apterous, ap*	Homeobox domain transcription factor
	CG9008	GH14910	2.5	*CG9008*	Unknown
	CG13574	RE56892	3.6	*CG13574*	Unknown
	CG17278	SD04019	18.3	*CG17278*	Kazal-type serine protease inhibitor family
	CG11407	GH20840	7.3	*CG11407*	Long-chain fatty acid transporter
	CG11491	LP10481	3	*broad, br*	BTB/POZ domain, zinc finger, transcription factor

Figure 4
A small number of genes are preferentially induced in the prothoracic leg disc. Two comparisons of the prothoracic leg disc (L1) to the wing disc (W) (numbers 83 and 93) are shown. The cluster analysis for these comparisons revealed 23 and 8 induced genes in the L1 (red) and W (green) subclusters, respectively. Note the expression of *apterous* in the wing disc cluster (see text).

detect it in discs [34]. We did not characterize *glec* further and cannot therefore distinguish whether this discrepancy should be attributed to greater sensitivity of the array analysis or to possible contamination by glial cell RNA in the eye-antennal RNA.

To gauge the validity of these sets of genes, eight genes with uncharacterized expression in imaginal discs - four from the wing-disc cluster and four from the eye-antennal-disc cluster - were chosen for examination by *in situ* hybridization. These genes were chosen to represent a spectrum of relative intensity of expression, from just above the threshold to the maximum observed. All of the patterns we obtained correlated with the array analysis and had the preferential expression as predicted for each cluster (Figure 6). Two out of eight genes were expressed specifically in either wing (CG10962; Figure 6c) or eye-antennal (CG9335; Figure 6g) discs. CG1607 was expressed specifically in the peripodial cells of wing discs (Figure 6b). CG11798 (Figure 6f) and CG9335 (Figure 6g) were expressed specifically by cells in or near the morphogenetic furrow of eye discs. CG9335 was expressed in R8 precursor cells that also stained with the neuronal-specific 22C10 antibody in a pattern reminiscent of *atonal* (data not shown); it was also expressed in the optic lobe, in Bolwig's organ (the light-sensing organ of the larva), and in a subset of neurons in the embryonic and larval central nervous system (CNS; data not shown). Another notable but unexpected finding was that Arrestin 2, the metarhodopsin-binding protein, was expressed in all imaginal discs we tested, not only in eye discs. Indeed, *in situ* hybridization indicated that it is expressed more abundantly in the wing than in the eye-antennal disc (Figure 6d). This distribution suggests that the *Drosophila* Arrestin 2 might function outside the visual system.

Gene	Fold	Name	Function
CG3430	1.4	/	Unknown
CG6139	1.5	/	Monoamine transporter
CG4526	1.4	/	Unknown
CG1727	2.3	/	Kazal-type S protease inhibitor family
CG1140	1.8	/	Long-chain fatty acid transporter
CG9427	2	/	Thioredoxin-like
CG6166	2.6	/	Unknown
CG9015	1.9	*engrailed , en*	Transcription factor
CG1449	2.3	*Zn finger homeodomain 2 , zfh2*	Transcription factor
CG1999	1.7	/	Unknown
CG8502	3.3	/	Structural protein of cuticle
CG1843	1.7	*Papilin, Ppn*	Ecm component, trypsin inhibitor
CG1607	2.6	/	Amino acid transporter
CG8376	9.1	*apterous, ap*	Transcription factor, LIM homeodomain
CG1151	2.5	*broad , br*	Transcription factor
LP1093	1.7	/	Unknown
CG1039	3	*Cyp310a1*	Cytochrome P450
CG1096	1.8	/	Oxidoreductase
CG1057	2.7	/	Unknown
CG8664	1.5	/	Unknown
GH2350	1.5	*Arrestin 2, Arr2*	Metarhodopsin inactivation
CG7144	1.9	/	Saccharopine dehydrogenase
CG4692	1.5	/	Hydrogen translocating F-type ATPase
CG1670	1.6	/	Monophenol monooxygenase activator
CG6588	1.7	*Fasciclin1, Fas1*	Neuronal cell adhesion
CG8434	1.6	/	Ig and MHC domain
GM0292	1.5	/	Unknown
CG7788	1.4	*Ice*	Effector caspase
CG1067	1.8	/	Unknown
CG3036	1.7	/	Sodium:phosphate symporter
CG5467	2	/	Unknown
CG6003	1.8	/	Unknown
CG1057	1.8	/	Sec7 domain, PH domain-like
CG6575	2.2	*Gliolectin, glec*	Carbohydrate binding protein
CG1539	2.4	*sandopo, spdo*	Actin and tropomyosin binding
CG9808	2.1	*PFTAIRE-interaction factor 1A or 1B, Pif1A/1B*	bZIP transcription factor family
CG1616	2	*disc proliferation abnormal, dpa*	DNA replication factor, chromatin binding
CG5009	1.7	*BcDNA:GH07485*	Palmitoyl-CoA oxidase
CG4070	1.5	*Tis11 homolog, Tis11*	DNA binding, zinc finger
CG8942	1.7	*BG:DS00180.10*	Wnt-protein binding
CG5248	1.9	*locomotion defects, loco*	G-protein signaling, glia cell differentiation
CG5869	1.5	*BG:DS02740.9*	Actin binding, depolymerizing proteins
CG1202	1.8	/	Unknown
CG2086	1.5	*BcDNA:GH03529*	Cell adhesion, EGF/Laminin
CG1837	2.5	*longitudinals lacking, lola, LD28033*	Transcription factor, axon guidance
CG4716	1.8	/	Unknown
CG2999	1.9	*unc-13*	Synaptic vesicle exocytosis
CG1515	2.8	/	RNI-like
CG3831	2.2	/	Unknown
CG7777	2.7	/	Water transporter, MIP family
CG8756	2.4	/	Chitin binding domain, ligand-binding domain
CG1509	2.4	/	Actin binding
CG1722	2.3	*prospero, pros*	Transcription factor, glia cell differentiation
CG1472	2.5	/	Histamine-gated chloride channel
CG1007	1.6	*extra macrochaetae, emc*	Transcription co-repressor
CG5981	1.6	*stathmin*	Microtubule binding
CG5210	2	*Chitinase-like, Chit*	Chitinase, cuticle chitin catabolism
CG1179	1.5	/	Zinc finger domain
CG8451	1.6	/	Sodium dependent multivitamin transporter
CG1454	2	*E(spl) region transcript mbeta, HLHmbeta*	Transcription factor, myc-type
CG1383	1.6	/	Unknown
CG5185	1.8	*Twin of m4, Tom*	Signaling, Notch antagonist
CG3523	2.3	*BcDNA:GH07626*	Fatty-acid synthase
CG6682	2	*Rapgap*	RAS GTPase activator
CG5600	1.6	/	Unknown
GH0977	1.9	/	Unknown
CG1802	2	*Calcium activated protein for secretion, Caps*	Synaptic vesicle exocytosis
CG7508	3.4	*atonal, ato*	Transcription factor
CG8965	3.3	/	RA domain
CG9134	2.7	/	Ligand binding or carrier, lectin
CG1184	6.9	/	Homeodomain-like
CG9335	16	/	Unknown
CG7727	4.6	*beta amyloid protein precursor-like, Appl*	Amyloid protein
CG2040	2.3	*hikaru genki, hig*	Cell adhesion molecule, selectin
CG7672	4.1	*glass, gl*	Transcription factor, photoreceptor determining
CG7908	1.7	/	Metalloendopeptidase
CG1675	2.1	*Spinophilin, Spn; E62*	Protein phosphatase1 binding, PDZ domain
CG3050	2.3	*Cyp 6d5*	Cytochrome P450
CG2269	1.6	/	Unknown
CG2759	1.7	*white, w*	Eye pigment precursor transporter
CG1673	1.5	*sloppy paired 1, slp1*	Transcription factor, fork head domain
CG1065	1.7	/	Retinal binding
CG7906	1.7	/	Kazal-type serine protease inhibitor family
CG7870	1.9	/	Dolichyl-phosphate beta-glucosyltransferase
CG1134	1.4	/	Unknown
CG9218	1.6	*smooth, sm*	RNA binding, ribonucleoprotein
CG4468	1.3	/	Unknown
CG4778	1.5	*BcDNA:GH02976*	Structural protein of peritrophic membrane
CG3136	2.3	/	bZIP transcription factor family
CG4511	2.9	/	ATP binding, Thioredoxin-like
CG1195	1.8	*SP1029*	Aminopeptidase, NOT aminoacyclase
CG3048	2.3	*TNF-receptor-associated factor 1, Traf1*	Defense response, signal transduction
CG1500	3.4	*Ecdyson-inducible gene L2, ImpL2*	Cell adhesion
CG9691	1.6	/	Unknown
CG1922	2.2	*onecut*	Transcription factor, homeobox domain
CG1404	2.2	*SP555*	SOCS domain, C-terminus of STAT-inhibitors
CG3556	1.7	/	Terpenoid cyclase/protein prenyltransferase

Figure 5 *(see the legend on the next page)*

Haltere and genital imaginal discs

In five direct comparisons of haltere and genital imaginal discs, 130 genes were found to be preferentially expressed in the haltere disc and 54 in the genital disc (see Additional data files). As with the wing/eye-antennal analysis, known genes clustered as predicted from previous studies. For example, *caudal*, a homeobox domain transcription factor, partitioned into the genital-disc cluster. *caudal* is expressed in the genital disc but not in more anterior tissues such as the haltere disc [35]. *abdominal B*, which would be predicted to show preferential expression in the genital disc, was not represented on the arrays. These 184 differentially expressed genes correspond to 3% of the entire set of 6,000 and represent the largest cluster we obtained in all pairwise comparisons of imaginal discs.

Expression profiles of larval organs

In order to compare and contrast the expression profiles of larval organs and imaginal discs, we made probes from the poly(A)$^+$ RNA isolated from several larval organs. We directly compared probes from the wing imaginal disc with: the salivary gland, an anterior fraction of the midgut, parts of the fat body and the brain hemisphere, including the optic lobe primordium. For each experiment, probes prepared from a single larva were compared. With the exception of the comparisons between the first leg disc and fat body which was carried out once, each of the other comparisons was repeated at least three times.

Visual inspection of the array scans of imaginal disc to larval tissue comparisons revealed a large number of red and green spots, indicative of a high degree of divergence. A representative block from one wing disc-to-fat body scan is shown in Figure 7b. These results contrast with disc to disc comparisons which produced mostly yellow and only few red or green spots on the arrays. For purposes of illustration, the same block of spots from a leg1 disc-to-leg3 disc scan is shown in Figure 7a. When the intensity of Cy3 and Cy5 signals for all cDNAs was plotted on a scatterplot, only a small number of data points are positioned away from the bisector on which signals of equal intensity in the two channels fall (Figure 7c). This is indicative of strong similarity in expression profiles and is reflected in a high CC (0.93). In contrast, the wing disc-to-fat body comparison had many data points located on each side of the bisector (CC 0.37) (Figure 7d). This disparity between leg disc-to-leg disc and wing disc-to-fat body comparison was also observed for all other disc-to-disc and disc-to-larval organ comparisons we carried out. Figure 8a illustrates the numbers of cDNAs that were expressed differentially in the various pairwise comparisons as a bar diagram. All disc-to-disc comparisons showed small numbers of differentially expressed cDNAs, whereas the disc-to-nondisc comparisons revealed many. The only exceptions were the comparisons of the wing imaginal disc to the brain/optic lobe (see below). To score for the differences in expression for all 6,000 genes, the standard deviations of the R/G ratios ($\log_2$-transformed) for each gene were calculated for all 42 experiments and represented in a scatterplot (Figure 8b). Standard deviations from the datasets derived from disc-to-disc comparisons (25 experiments) were all < 1, indicating the high level of similarity among the different imaginal discs. In contrast, the disc-to-nondisc comparisons (17 experiments) were, with a single exception (see below) > 1, reflecting a high degree of divergence (Figure 8b). These observations are consistent with the distinct developmental programs of imaginal and larval cells.

The only exceptions to emerge in this analysis were the cells of the optic lobe/brain hemisphere. Comparisons of these neuronal tissues with wing discs revealed only a small number of differentially expressed genes (Figure 8a) and had low standard deviations that were in the range of disc-to-disc comparisons (Figure 8b). The optic lobe primordium and brain of late third instar larvae have only limited functions in larval photoreception and contain only a small number of mature neurons. The larval brain is populated with clusters of immature neurons, neuroblasts and ganglion mother cells whose descendants will generate the neuronal and glial populations of the adult optic neuropil and the brain hemisphere. The brain hemisphere also differs from other larval tissues by not degenerating during metamorphosis. Instead, most neurons of the larva join with newly formed adult-specific neurons to build the CNS of the adult

Figure 5 *(see the figure on the previous page)*
Genes preferentially expressed in wing and eye-antennal discs. Amplified RNA from single discs of five individual larvae was used to carry out direct comparisons between a wing imaginal disc and an eye-antenna imaginal disc of the same larva (experiments 51, 88, 146, 155, 156, 157, 173 and 175). Whenever samples from one larva were used for more than one experiment, different left to right combinations of the individual discs were made. In two experiments, combinations of discs from two larvae (experiment 74) or pools of five larvae (ten wing and ten eye-antenna discs, experiment 98) were used. Cluster analysis was performed on the dataset with the requirement to show induction > 1.74 (0.8 of the $\log_2$-transformed R/G ratio) in at least five experiments. One hundred and forty genes grouped into different subclusters. After removal of double hits or genes with inconsistent induction, 97 genes remained. Twenty-four genes were induced in the wing imaginal disc (red) and 73 genes in the eye-antennal disc (green). Black indicates lack of induction. Fold induction was calculated as an average induction in all experiments (column 3). The name, description and molecular or biological function is indicated as in Flybase [44] (columns 4 and 5). Some genes with known expression in the respective tissue are included in the clusters, such as *apterous*, *engrailed* and *glass* (see text). Of this set, 22 genes are uncharacterized and do not code for known protein domains. Note the high induction of CG9335 (16 fold) and CG11849 (6.9 fold) in the eye-antennal cluster. Genes marked in red were chosen for *in situ* experiments (see Figure 6).

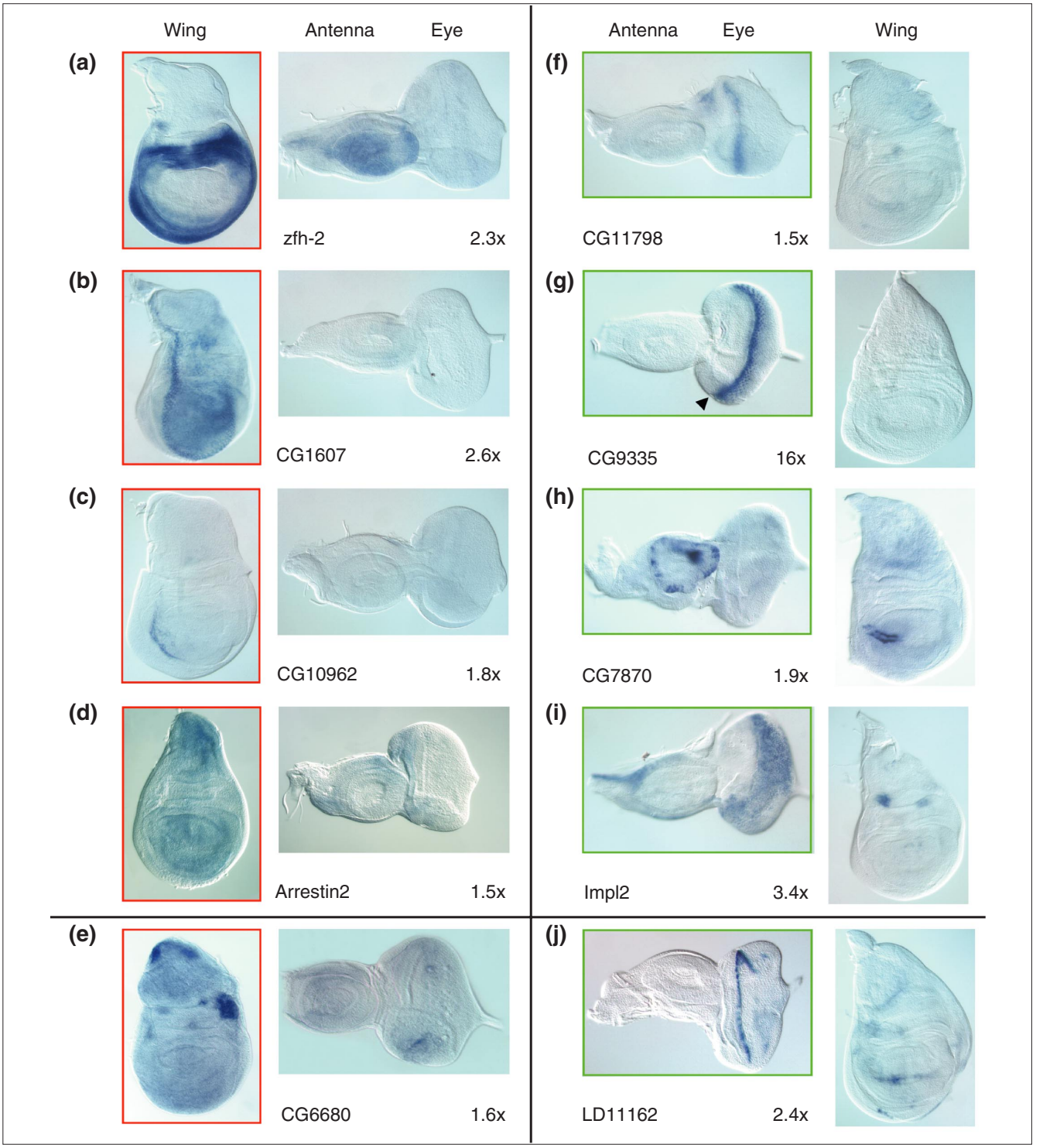


Figure 6
Expression patterns of genes in the wing and eye-antennal clusters. *In situ* hybridizations of genes from **(a-d)** the wing and **(f-i)** eye-antennal disc clusters show the disc-specific patterns of expression. Confirming the mircoarray data, the signal intensities are higher in the wing discs in (a-d) (red frame) and the eye-antennal discs in (f-i) (green frame). Note the refined expression pattern in the wing disc for CG10962 in (c). In two cases, CG10962 (c) and CG9335 (g), signal could only be detected in the predicted disc. The arrowhead in (g) indicates the morphogenetic furrow. The number indicates the average fold induction in the 11 experiments. Arrestin2 (d) and CG11798 (f) were included in the clusters because their relative induction was > 1.74 in more than five experiments. **(e)** CG6680 and **(j)** LD11162 failed the threshold criteria for the cluster depicted in Figure 4, but were part of a larger cluster with lower threshold settings. For both genes, the *in situ* patterns confirm the predicted expression. Discs are oriented anterior to the left, and dorsal uppermost.

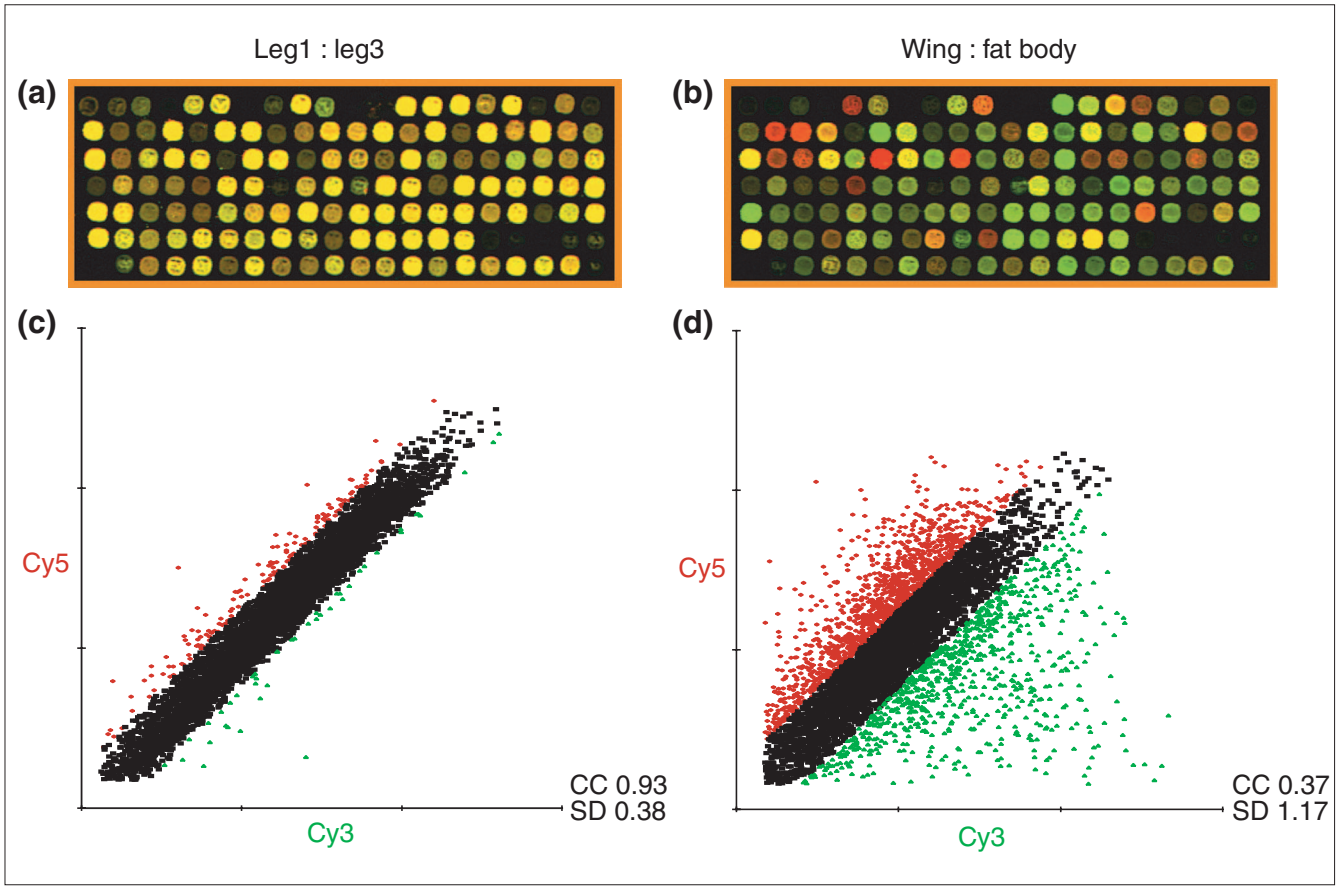


Figure 7
Imaginal discs share a similar expression profile but differ from differentiated larval tissues. **(a)** Comparison between first and third leg discs. An enlargement of a representative block out of 32 blocks on the microarray is shown. It produced mostly yellow spots on the superimposed red (Cy5 labeling) and green (Cy3 labeling) images, indicative of a high degree of similarity in the respective expression profiles. **(b)** In a comparison of wing disc and fat body, the same block contained mostly red and green spots, indicating a high degree of divergence. **(c,d)** Scatterplots of Cy3 and Cy5 intensities in comparisons (c) between leg discs and (d) between wing disc and larval fat body**.** In the leg comparison the spots are in close proximity to the bisector (CC 0.93, SD 0.38) with only a small number of genes induced in either the leg1 or leg3 disc. In contrast, spots are spread widely to both sides of the bisector for the wing-to-fat body comparison (CC 0.37, SD 1.17), indicating a large number of differentially expressed genes. The data points are color coded such that spots that are induced > 1.74 fold in the Cy5 channel are colored red and those induced > 1.74 in the Cy3 channel in green. Ratios within a threshold of 1.74 are represented in black.

([36] and references therein). We interpret the high degree of similarity in the expression profiles of imaginal discs and the brain/optic lobe preparations as a manifestation of the apparent commonalities of their developmental programs.

The number of genes that were found to be induced in the larval organs is significantly greater than the number of genes in the sets defined by the disc-to-disc comparisons. Most of the genes in these sets have not been characterized previously and therefore have no known function, but the known genes in these clusters illustrate the predictive power of the method; we mention three examples.

Many of the genes that are preferentially expressed in the optic lobe/brain hemisphere have predicted or described expression and/or function in neurons or glial cells (data not shown). One of these is *lola*, which encodes a transcription factor involved in axon guidance and is expressed in the embryonic optic lobe placodes as well as other tissues [37]. Another is *beta amyloid protein precursor-like* (*Appl*), with known expression in the larval optic lobe [38].

The expression profile of the cyclin-dependent kinase 1 (*Cdk-1/cdc2*), is also noteworthy. Cdk-1 is believed to regulate the G2 checkpoint of the cell cycle and is required to maintain diploidy. Mutations in *Cdk-1* drive cells that would normally remain diploid into endoreplication [39]. We found *Cdk-1* (CG5363) to be repressed in larval tissues when compared to imaginal discs (Figure 9). As most of the cells in the larval fat body, salivary gland and midgut are polyploid (reviewed in [40]), the array hybridizations are in good agreement with the known function of *Cdk-1*.

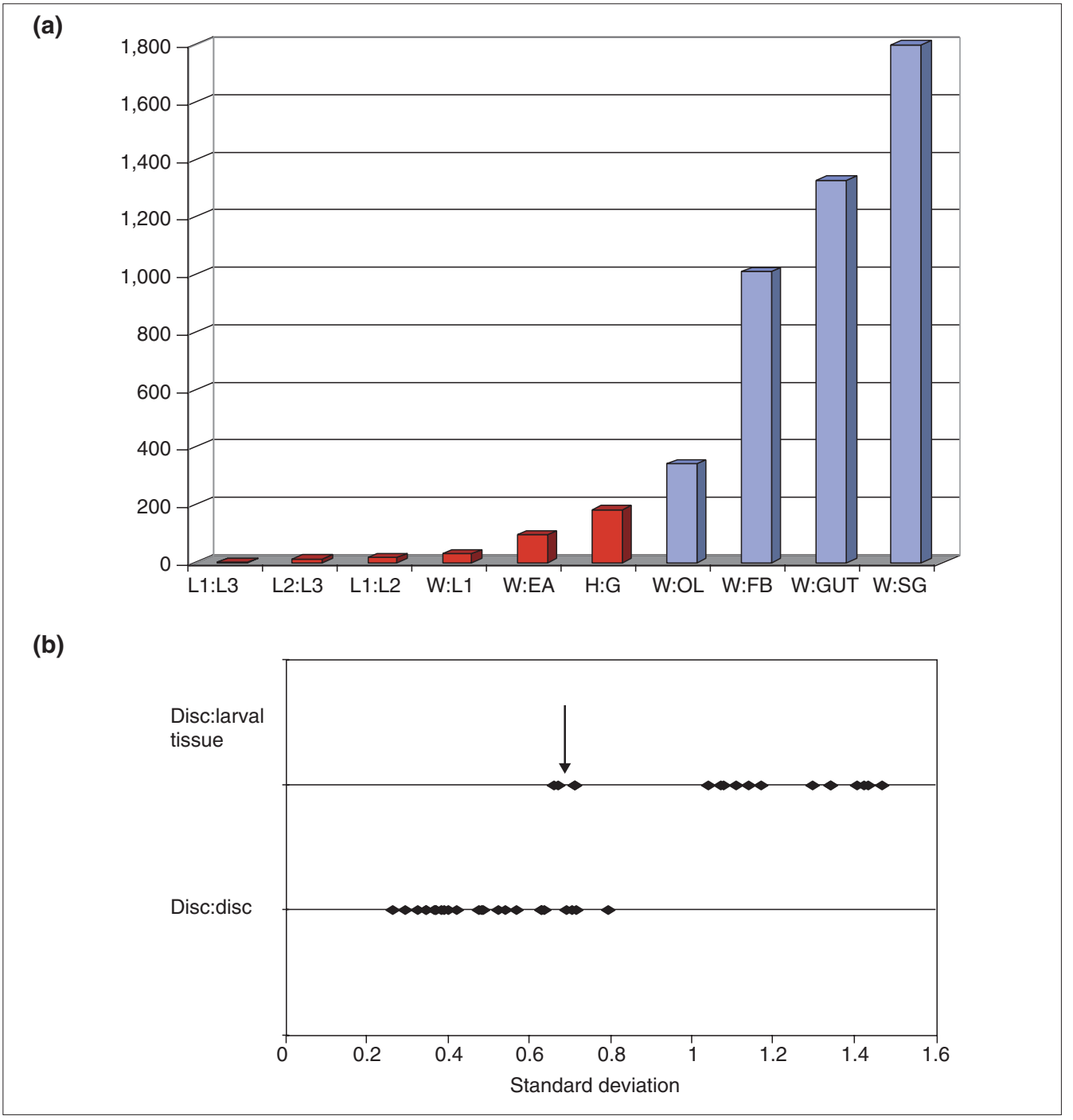


Figure 8
The mRNA populations in larval and imaginal tissues are distinct. **(a)** Graphical representation of the number of spots with intensity differences > 1.74 when imaginal and larval tissues were compared. The numbers of spots are derived from cluster analysis of the repeated experiments. All disc-to-disc comparisons are represented by red bars, whereas the disc-to-nondisc comparisons are shown in blue. EA, eye-antennal; FB, fat body; G, genital; GUT, anterior part of the midgut; H, haltere; L1, leg 1; L2, leg 2; L3, leg 3, OL, optic lobe/brain; SG, salivary gland; W, wing. **(b)** A plot of the level of divergence as measured by the SD of the $\log_2$-transformed R/G ratios for all genes in 42 experiments (leaving out the wing-to-wing experiment). Of these, 25 experiments were comparisons of one imaginal disc to another (see text and Table 1) and the remaining 17 compared imaginal discs to larval tissue, that is, salivary gland, midgut, fat body and optic lobe/brain hemisphere. The disc-to-disc comparisons were placed on the lower line, whereas disc-to-nondisc comparisons are represented on the top line. All disc-to-disc experiments grouped to the left as a result of SD < 1, indicating similarity of the expression profiles among the various imaginal discs. Disc-to-nondisc comparisons group to the right, with SD > 1. A group of three experiments of disc-to-nondisc comparisons reveals low SDs in the range of disc-to-disc comparisons. These are the wing disc-to-optic lobe/brain hemisphere comparisons (arrow).

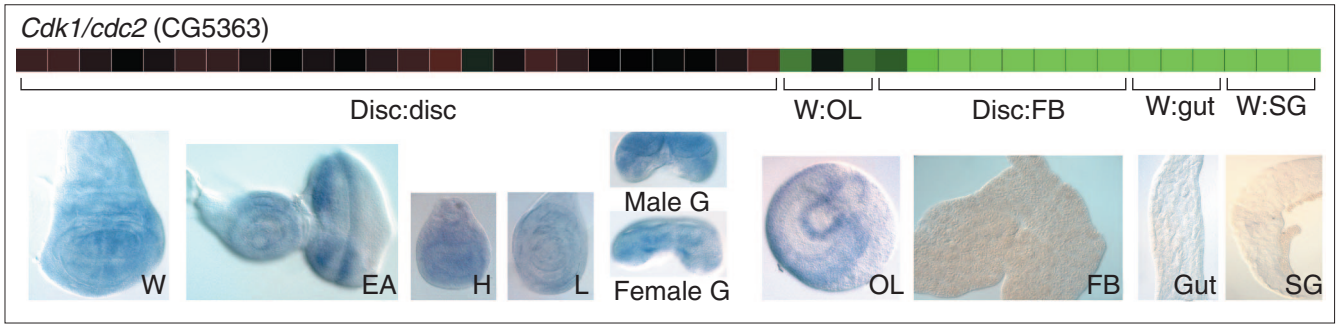


Figure 9
Cyclin-dependent kinase 1 (*Cdk1/cdc2*) is expressed in imaginal but not in larval tissues. The cluster analysis for *Cdk1* is shown for 41 pairwise comparisons with each box representing one experiment. All the disc-to-disc comparisons (see text) group to the left and display ratios of the intensities close to 1, as indicated by the dark shades of red and green (color coding as in Figure 3). The *in situ* hybridization to the various imaginal discs confirmed that *Cdk1* is expressed in all discs. Cluster analysis of the comparisons of discs (labeled in green) and larval tissues (red) showed a strong induction in imaginal discs (represented by the intense green staining). This finding is confirmed by previous descriptions that *Cdk1* is downregulated in endoreplicated tissues and by the lack of signal after *in situ* hybridization (FB, GUT and SG). The wing disc (W) to brain/optic lobe (OL) comparisons revealed less induction of *Cdk1* in the wing disc as indicated by the darker shades of green. This is in good agreement with *in situ* hybridization data that showed some expression, particularly in the proliferation zones of the optic lobe, but no signal in most parts of the brain. Abbreviations as in Figure 8.

Experimental design and data analysis

Two methods for comparative hybridizations to DNA microarrays have been used in previous studies. Most commonly, experimental samples have been compared to a common reference sample that contains an appropriate diversity of cell types and states. An alternative method, which we have used in this study, is to directly compare probes generated from two experimental samples. This second method should be the more sensitive. The amplification technique we used was such that product generated from a single imaginal disc was sufficient to make several probes and to carry out multiple array hybridizations. By comparing samples that had been isolated from a single animal, this method minimized differences due to biological variability. By repeating experiments several times and by reversing the label (dye-flip) for experimental samples in repeat hybridizations, we also attempted to minimize experimental noise.

In addition to these design aspects of the experimental protocols, we applied a number of filters to the data analysis. We selected only those genes with hybridization signals over background in 80% of all experiments and with a 1.74-fold difference between the two channels in at least 45%. Applying this method to 11 arrays that compared wing and eye-antennal discs, we found two subclusters representing genes preferentially expressed in either disc. These lists of genes are largely validated by the previously described expression pattern in either tissue and by *in situ* hybridization analysis. We understand that the use of any such threshold settings is likely to filter out biologically relevant genes. For example, when less stringent settings were used (for example, a 1.6-fold intensity difference increased the gene sets by 20%), additional genes such as *twin of eyeless* were included in the eye-antennal cluster and *Antp* and CG6680 (Figure 6e) in the wing cluster. Relaxing the requirement to produce a ratio in at least 80% of all experiments to 70% identifies yet another group of genes, and includes LD11162 (Figure 6j). Conversely, setting a more stringent threshold to 1.9-fold intensity difference decreased the gene sets by 25% and removed genes such as Arrestin2 and CG11798 (Figure 6d,f).

Setting a threshold level for the clusters is arbitrary, and as some candidate genes may be expressed by only a small subset of cells in the experimental sample, each analysis must be evaluated separately. Nevertheless, we interpret the good correlation between the array analysis and expression patterns we have described as indicating that the method is both efficient and reliable. Subsequent to the analyses described above, we carried out single-disc comparisons with microarrays that represent a nearly complete set of the approximately 14,000 annotated *Drosophila* genes. The short cDNAs spotted on these arrays were produced with specific primer pairs obtained from Incyte Genomics, Inc. Use of these arrays for disc-to-disc comparisons also identified only a relatively small number of differentially expressed genes (data not shown). The list of induced genes was apparently more complete, however, and included, for example, *vestigal* in the wing-to-leg disc comparisons (A.K., G. Schubiger and T.B.K., unpublished observations).

Although the absolute number of genes included in the lists of tissue-specific genes is not meaningful, the relatively small number of genes that differ between imaginal tissues and the significantly greater number of genes that differ between imaginal and larval tissues certainly is.

Conclusions

We have shown that linear RNA amplification is reproducible and can be used to generate probes for microarray

experiments with limited quantities of starting material. We used this technique to determine the expression profiles of imaginal and larval tissues by pairwise comparisons of tissues dissected from single larvae. Our approach was validated by the presence of characterized genes in the appropriate cluster and by confirmation of expression patterns by direct analysis. Applying stringent threshold criteria, we found that less than 3% of the genes represented on the arrays were preferentially expressed when discs were compared to each other. Many of these genes have either not been characterized previously or lack obvious similarity to known coding domains. Those that have been previously characterized or share sequence similarities represent a broad array of cellular functions. These include basic metabolic and transporting functions, structural, cell-adhesive and signaling functions, as well as transcriptional regulation. We lack sufficient understanding of the specific processes in most cell types to interpret the significance of the genes that they specifically express. Future studies will be required to identify the functional networks that the identified candidate genes describe and to establish how they contribute to the execution of specific developmental programs.

We anticipate that the methods we describe here can be applied to determine and compare the expression profiles of small cell groups in *Drosophila* tissues. We look forward to a further enhancement in resolution that can be achieved by combining cell-sorting techniques with this type of array analysis. We found that starting with as little as 10 pg poly(A)$^+$ RNA yields enough aRNA for one hybridization. This amount roughly corresponds to 300 cells of the third instar wing imaginal disc.

Materials and methods

Microarray production and labeling

Spotted cDNA microarrays were produced essentially as described at [41]. In brief, the *Drosophila* Gene Collection (DGC 1.0, kindly provided by the Berkeley *Drosophila* Genome Project) of 5,849 nonredundant cDNAs was amplified using universal primers essentially as described at [42]. The amplification products were purified with 96-well format Qiagen PCR purification columns. All PCR products were analyzed on agarose gels, and reactions with no detectable product, multiple bands or bands of unexpected size were repeated using Herculase (Amersham) or Expand polymerase (Roche). One hundred and seventy-five cDNAs from our lab collection were amplified and added to the set, resulting in a total of 6,024 cDNAs. Reactions were arrayed into 384-well plates and printed on poly-L-lysine-coated glass slides using a linear servo arrayer and ArrayMaker Version 2 control software. To minimize background caused by oxidation of the polylysine coating, slides were pretreated by a protocol suggested by Paul Ebert. Before post-processing, slides were incubated for 5 min in 3x SSC; 0.2% SDS at 65°C, washed first in water and then in 95% ethanol, followed by centrifugation to dry the slides. Slides were then treated, and the hybridization reaction was carried out as described at [42]. We indirectly labeled the hybridization probes by incorporation of amino-allyl modified nucleotides in a first-strand cDNA RT reaction. Monofunctional Cy5 or Cy3 dye (Amersham) was subsequently coupled to the reactive residues. Multiple hybridizations were carried out for most experiments, and dye labeling was reversed to avoid systematic bias.

Larval dissections, RNA isolation and *in situ* hybridization

Third instar wandering larvae were dissected and washed several times in Ringer solution (130 mM NaCl, 5 mM KCl, 2 mM Na_2HPO_4, 1.5 mM $CaCl_2$, 0.37 mM KH_2PO_4). Fat-body fragments were mainly anterior plates. Midgut was the anterior portion just posterior to the gastric caeca. Brain/optic lobe consisted of one brain hemisphere including the optic lobe primordium.

Total RNA from larval tissues was extracted using the Mini RNA Isolation Kit (Zymo Research, Orange, CA) and was eluted with RNase-free water. Total RNA from embryos was extracted with Trizol reagent (Invitrogen). Embryonic poly(A)$^+$ RNA was purified using the Oligotex mRNA Mini Kit (Qiagen). Poly(A)$^+$ RNA was not purified before amplification of disc RNA because of the small amounts (see Figure 1).

In situ hybridizations were carried out as described in O'Neill and Bier [43]. The plasmids used to generate *in situ* probes were derived by subcloning an aliquot of the PCR product that was printed onto the microarray with Topo TA cloning (Invitrogen). All cloned products were sequenced and separate hybridizations were carried out for sense and antisense probes.

RNA amplification

Methods for amplification were adapted from Wang *et al.* [25] and Baugh *et al.* [23]. Total RNA isolated from dissected larval tissues or poly(A)$^+$ RNA purified from embryos was incubated with 100 ng oligo(dT)$_{24}$ -T7 primer (GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCG-GTTTTTTTTTTTTTTTTTTTTTTTT) and 125 ng TS primer (AAGCAGTGGTAACAACGCAGAGTACGCGGG) in 5 μl at 65°C for 10 min and cooled on ice. Six microliters first-strand premix were added and incubated at 42°C for 2 h (6 μl mix was made up with 2 μl 5x first-strand buffer, 1 μl 0.1 M DTT, 1 μl dNTPs (10 mM each), 0.3 μl T4gp32 (USB, 13.82 mg/ml), 0.5 μl RNasin (Promega), 1 μl Superscript II (Invitrogen)). The reaction was incubated at 65°C for 10 min and cooled on ice. Cold second-strand premix (64.5 μl prepared on ice: 45 μl RNase-free water, 15 μl 5x second-strand buffer (Invitrogen), 1.5 μl dNTPs (10 mM each), 0.5 μl *E. coli* ligase (10 U/μl), 2 μl *E. coli* polymerase (10 U/μl), 0.5 μl *E. coli* RNaseH (2 U/μl)). Enzymes for the second-strand

premix (Invitrogen) were added and incubated at 16°C for 2 h, followed by the addition of 2 units T4 DNA polymerase (Promega) and incubation at 16°C for 15 min before heat inactivation at 70°C for 10 min. Clean-up was performed with DNA clean and concentrator-5 (Zymo Research), eluting twice with 8 μl RNase-free water. The total volume was adjusted to 8 μl in a Speed Vac. The first IVT was performed with Megascript T7 (Ambion) in a 20 μl volume for 5-6 h, followed by DNA digestion. aRNA was purified using Mini RNA Isolation Kit or RNA Clean-up kit (Zymo Research), eluting in 2x 8 μl RNase-free water. Random hexanucleotides (250 ng) and TS primer (1 μg) were added and the volume adjusted to 5 μl. The mix was incubated at 65°C for 10 min before 5 μl first-strand premix was added followed by incubation at 42°C for 2 h. Before the second-strand synthesis, 200-500 ng oligo(dT)$_{24}$ -T7 primer was added and denatured at 65°C for 10 min. The second-round second-strand synthesis, clean-up and IVT were carried out as for the first round, except that no ligase was added to the second-strand premix. One-quarter to one-third of the total aRNA was applied to the clean-up columns after the second IVT to avoid overloading. All reactions were carried out in a thermocycler with heated lid or air incubator to avoid evaporation.

Data analysis

Hybridized microarrays were scanned with a GenePix 4000A Microarray Scanner (Axon Instruments, Union City, CA). Data were analyzed and displayed with Cluster and Treeview [31], AMAD [41], Genepix PRO (Axon Instruments), and Microsoft Excel. Normalization for cluster analysis and calculation of standard deviations were done with AMAD. Only genes that qualified with a combined median intensity > 300 above background in both channels in at least 80% of the repeated experiments were included in the analysis. A threshold of > 1.74 (= 0.8 of the $\log_2$-transformed ratios) was chosen for all comparisons. A requirement to show induction above threshold in 100% of the arrays was applied for experiments that were repeated three times or less. For experiments that were repeated more than three times this requirement was relaxed to 45%. Standard deviations were calculated on the $\log_2$-transformed normalized R/G ratios using the 'nonbiased' method. To calculate the correlation coefficients, the intensities of both channels were normalized to equalize the sums of the intensities.

Additional data files

A comparison of expression profiles of genital and haltere discs is available with the online version of this paper. Five comparisons between the haltere (H) and genital (G) disc were carried out. Four of the G discs were female (nos 48, 149, 150, 224) and one was male (no. 148). To control for contaminating fat body (FB) cells in the disc preparations, one direct comparison between haltere disc and FB (no. 129) was included. The columns indicate the subclusters with consistent induction in one channel (H, green; G, red), the gene identification numbers (ID), the average fold induction of the five H:G comparisons, the gene name and function as published on [44]. 130 and 54 genes were induced in the H and G subclusters, respectively, using the same threshold settings as in Figure 3. Note the 2.5-fold induction of *caudal* (CG1759) in the G subcluster.

Acknowledgements

We thank Ken Burtis, Sarah Meadows, Gerold Schubiger, Mike Eisen and Joe DeRisi for technical help and advice, and Paul Ebert for suggesting the pretreatment protocol. We thank the Merck and the Ellison Foundation for financial support.

References

1. Struhl G: **A homoeotic mutation transforming leg to antenna in *Drosophila*.** *Nature* 1981, **292:**635-638.
2. Moses K, Ellis MC, Rubin GM: **The *glass* gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells.** *Nature* 1989, **340:**531-536.
3. 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:**2185-2195.
4. 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.
5. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, *et al.*: **Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome.** *Cell* 2001, **107:**477-487.
6. Bryant Z, Subrahmanyan L, Tworoger M, LaTray L, Liu CR, Li MJ, van den Engh G, Ruohola-Baker H: **Characterization of differentially expressed genes in purified *Drosophila* follicle cells: toward a general strategy for cell type-specific developmental analysis.** *Proc Natl Acad Sci USA* 1999, **96:**5559-5564.
7. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I: **The transcriptional program of sporulation in budding yeast.** *Science* 1998, **282:**699-705.
8. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM: **Use of a cDNA microarray to analyse gene expression patterns in human cancer.** *Nat Genet* 1996, **14:**457-460.
9. Furlong EE, Andersen EC, Null B, White KP, Scott MP: **Patterns of gene expression during *Drosophila* mesoderm development.** *Science* 2001, **293:**1629-1633.
10. Hayward RE, Derisi JL, Alfadhli S, Kaslow DC, Brown PO, Rathod PK: **Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria.** *Mol Microbiol* 2000, **35:**6-14.
11. Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, Brown EL: **Genomic analysis of gene expression in *C. elegans*.** *Science* 2000, **290:**809-812.
12. Irving P, Troxler L, Heuer TS, Belvin M, Kopczynski C, Reichhart JM, Hoffmann JA, Hetru C: **A genome-wide analysis of immune responses in *Drosophila*.** *Proc Natl Acad Sci USA* 2001, **98:**15119-15124.
13. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: **Yeast microarrays for genome wide parallel genetic and gene expression analysis.** *Proc Natl Acad Sci USA* 1997, **94:**13057-13062.
14. McDonald MJ, Rosbash M: **Microarray analysis and organization of circadian gene expression in *Drosophila*.** *Cell* 2001, **107:**567-578.
15. Miki R, Kadota K, Bono H, Mizuno Y, Tomaru Y, Carninci P, Itoh M, Shibata K, Kawai J, Konno H, *et al.*: **Delineating developmental and metabolic pathways *in vivo* by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays.** *Proc Natl Acad Sci USA* 2001, **98:**2199-2204.

16. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW: **Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.** *Proc Natl Acad Sci USA* 1996, **93:**10614-10619.
17. Shalon D, Smith SJ, Brown PO: **A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization.** *Genome Res* 1996, **6:**639-645.
18. White KP, Rifkin SA, Hurban P, Hogness DS: **Microarray analysis of *Drosophila* development during metamorphosis.** *Science* 1999, **286:**2179-2184.
19. Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Brown PO, Schoolnik GK: **Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization.** *Proc Natl Acad Sci USA* 1999, **96:**12833-12838.
20. Zou S, Meadows S, Sharp L, Jan LY, Jan YN: **Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2000, **97:**13726-13731.
21. Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P: **Analysis of gene expression in single live neurons.** *Proc Natl Acad Sci USA* 1992, **89:**3010-3014.
22. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH: **Amplified RNA synthesized from limited quantities of heterogeneous cDNA.** *Proc Natl Acad Sci USA* 1990, **87:**1663-1667.
23. Baugh LR, Hill AA, Brown EL, Hunter CP: **Quantitative analysis of mRNA amplification by *in vitro* transcription.** *Nucleic Acids Res* 2001, **29:**e29.
24. Eberwine J: **Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA.** *Biotechniques* 1996, **20:**584-591.
25. Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM: **High-fidelity mRNA amplification for gene profiling.** *Nat Biotechnol* 2000, **18:**457-459.
26. Klingenberg CP, McIntyre GS, Zaklan SD: **Left-right asymmetry of fly wings and the evolution of body axes.** *Proc R Soc Lond B Biol Sci* 1998, **265:**1255-1259.
27. White RA, Wilcox M: **Protein products of the bithorax complex in *Drosophila*.** *Cell* 1984, **39:**163-171.
28. Wirz J, Fessler LI, Gehring WJ: **Localization of the *Antennapedia* protein in *Drosophila* embryos and imaginal discs.** *EMBO J* 1986, **5:**3327-3334.
29. Glicksman MA, Brower DL: **Expression of the Sex combs reduced protein in *Drosophila* larvae.** *Dev Biol* 1988, **127:**113-118.
30. Pattatucci AM, Kaufman TC: **The homeotic gene Sex combs reduced of *Drosophila melanogaster* is differentially regulated in the embryonic and imaginal stages of development.** *Genetics* 1991, **129:**443-461.
31. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95:**14863-14868.
32. Kiss I, Beaton AH, Tardiff J, Fristrom D, Fristrom JW: **Interactions and developmental effects of mutations in the Broad-Complex of *Drosophila melanogaster*.** *Genetics* 1988, **118:**247-259.
33. Williams JA, Paddock SW, Carroll SB: **Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing *Drosophila* wing disc into discrete subregions.** *Development* 1993, **117:**571-584.
34. Tiemeyer M, Goodman CS: **Gliolectin is a novel carbohydrate-binding protein expressed by a subset of glia in the embryonic *Drosophila* nervous system.** *Development* 1996, **122:**925-936.
35. Freeland DE, Kuhn DT: **Expression patterns of developmental genes reveal segment and parasegment organization of *D. melanogaster* genital discs.** *Mech Dev* 1996, **56:**61-72.
36. Ceron J, González C, Tejedor FJ: **Patterns of cell division and expression of asymmetric cell fate determinants in postembryonic neuroblast lineages of *Drosophila*.** *Dev Biol* 2001, **230:**125-138.
37. Giniger E, Tietje K, Jan LY, Jan YN: ***lola* encodes a putative transcription factor required for axon growth and guidance in *Drosophila*.** *Development* 1994, **120:**1385-1398.
38. Torroja L, Luo L, White K: **APPL, the *Drosophila* member of the APP-family, exhibits differential trafficking and processing in CNS neurons.** *J Neurosci* 1996, **16:**4638-4650.
39. Hayashi S: **A Cdc2 dependent checkpoint maintains diploidy in *Drosophila*.** *Development* 1996, **122:**1051-1058.
40. Edgar BA, Orr-Weaver TL: **Endoreplication cell cycles: more for less.** *Cell* 2001, **105:**297-306.
41. Microarrays.org - **a public source for microarray protocols and software** [http://www.microarrays.org/]
42. **Berkeley *Drosophila* Genome Project resources** [http://www.fruitfly.org/about/methods/ampUniGene.html]
43. O'Neill JW, Bier E: **Double-label *in situ* hybridization using biotin and digoxigenin-tagged RNA probes.** *Biotechniques* 1994, **17:**870, 874-875.
44. **FlyBase** [http://flybase.bio.indiana.edu/]