



Robinson, S.W., Herzyk, P., Dow, J.A.T., and Leader, D.P. (2013)  
FlyAtlas: database of gene expression in the tissues of drosophila  
melanogaster. Nucleic Acids Research, 41 (D1). D744-D750. ISSN 0305-  
1048

Copyright © 2012 The Authors

<http://eprints.gla.ac.uk/75557/>

Deposited on: 20th February 2013

# FlyAtlas: database of gene expression in the tissues of *Drosophila melanogaster*

Scott W. Robinson<sup>1</sup>, Pawel Herzyk<sup>2,3</sup>, Julian A. T. Dow<sup>2,\*</sup> and David P. Leader<sup>1,\*</sup>

<sup>1</sup>School of Life Sciences, College of Medical, Veterinary and Life Sciences, <sup>2</sup>Institute of Molecular Cell and Systems Biology, College of Medical, Veterinary and Life Sciences and <sup>3</sup>Glasgow Polyomics, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

Received September 17, 2012; Revised October 23, 2012; Accepted October 24, 2012

## ABSTRACT

The FlyAtlas resource contains data on the expression of the genes of *Drosophila melanogaster* in different tissues (currently 25—17 adult and 8 larval) obtained by hybridization of messenger RNA to Affymetrix *Drosophila* Genome 2 microarrays. The microarray probe sets cover 13250 *Drosophila* genes, detecting 12533 in an unambiguous manner. The data underlying the original web application (<http://flyatlas.org>) have been restructured into a relational database and a Java servlet written to provide a new web interface, FlyAtlas 2 (<http://flyatlas.gla.ac.uk/>), which allows several additional queries. Users can retrieve data for individual genes or for groups of genes belonging to the same or related ontological categories. Assistance in selecting valid search terms is provided by an Ajax 'autosuggest' facility that polls the database as the user types. Searches can also focus on particular tissues, and data can be retrieved for the most highly expressed genes, for genes of a particular category with above-average expression or for genes with the greatest difference in expression between the larval and adult stages. A novel facility allows the database to be queried with a specific gene to find other genes with a similar pattern of expression across the different tissues.

## INTRODUCTION

*Drosophila melanogaster* is one of the most important model eukaryotic organisms, in part because of the great detail with which its genetics have been studied. This genetic heritage has provided an invaluable context to the sequence of its genome (1), which contains many homologues of human genes, including 75% of those known to be involved in disease (2). The sequence of the

genome has enabled numerous microarray studies of gene expression. Although these have produced much valuable information, they have sometimes suffered from the limitation of studying gene expression in the whole animal, rather than in individual tissues, potentially obscuring significant changes occurring in tissues that constitute only a small proportion of the overall body mass. To provide additional genome-wide insights into both gene and tissue function, a comprehensive atlas of gene expression (using the authoritative Affymetrix platform) across multiple tissues and life stages was produced and is available online at <http://flyatlas.org/> (3). These data have been taken up with enthusiasm by both the *Drosophila* and broader biological communities, and the original article has been cited >700 times since 2007 on Google Scholar. Even this figure probably understates the usage of the resource, as the data are also published through the established genome project resource, FlyBase (4,5), as well as the *Drosophila* data portal, FlyMine (6). An example of a listing from a search using this facility is shown in Supplementary Figure S1.

Despite its great utility, the original FlyAtlas web facility is not without limitations: a small portion of the information derived from the Affymetrix annotation was outdated or incorrect, and some searches using *Drosophila* genetic symbols could either fail or produce an unmanageable number of 'hits'. This latter problem was related to the simple 'flat-file' format of the FlyAtlas dataset and the algorithm used to search it. Rather than addressing these issues by updating the existing data records and program scripts, it was decided to replace the flat files by a relational database that would not only enable the problems to be resolved, but would also facilitate other enhancements to the web application.

One enhancement relates to an aspect of the relationship between individual probe sets used in the hybridization and *Drosophila* genes—one that is particularly important for the user of the facility to appreciate. Some genes (1367 of the 13250 for which we have data) are detected by more than one probe set (referred to as

\*To whom correspondence should be addressed. Tel: +44 1 41 33 05 905; Email: david.leader@glasgow.ac.uk  
Correspondence may also be addressed to Julian A. T. Dow. Tel: +44 1 41 33 04 616; Email: julian.dow@glasgow.ac.uk

‘duplicates’ for convenience), whereas other probe sets (1188 of the 14438) detect more than one gene (referred to as ‘degenerate’). This degeneracy means that for certain genes (717 of the 13250) the data do not allow unambiguous conclusions to be drawn regarding expression in different tissues. The relational database has enabled us to deal with these situations more explicitly.

In this report, we describe the structure of the relational database and the new web interface that has been provided for FlyAtlas, including features that provide significant improvements in usability. We also describe a new facility that allows one to find genes with a similar pattern of expression across different tissues to that of a query gene.

## DATA COLLECTION

The number of tissues in the original version of FlyAtlas (3) was 11, but has since been increased to 25. The tissues from adult flies are currently head, salivary gland, heart, ovary, virgin spermatheca, mated spermatheca, testis, accessory glands, carcass, fat body, tubule, midgut, eye, brain, hindgut, thoracoabdominal ganglion and crop, whereas those from larvae are central nervous system, tubule, trachea, midgut, salivary gland, fat body, hindgut and carcass. Details of the dissections are provided as Supplementary Table S1.

Other experimental details are available as Supplementary Material in the previous article (3) but are briefly repeated here. The flies were wild-type *Drosophila melanogaster* of the Canton S strain. The adults were reared at 23°C on a 12h:12h light:dark regime, on standard *Drosophila* diet, and killed 7 days after adult emergence. The larvae were third instar feeding larvae, raised under the same conditions and sampled before the wandering stage. The tissues were pooled from equal numbers of males and females, except in the case of the gonads.

At least 1500 ng messenger RNA was obtained from each tissue and from whole flies. It was then amplified and hybridized to Affymetrix *Drosophila* Genome 2 expression arrays (representing 18500 transcripts) using the Affymetrix standard protocol. For each tissue, four independent biological replicates were obtained, i.e. each array corresponds to one biological replicate.

The arrays were read using standard procedures with an Affymetrix GeneChip Scanner 3000 7G. Data were analysed using Affymetrix proprietary GCOS software (v 1.4), and data from tissue samples were compared with those of the adult fly dataset using Affymetrix Data mining software (v 3.1).

## DATA REFINEMENT AND DATABASE CONSTRUCTION

Three sets of data underlie the database: a file of the results of the hybridization experiments, a file documenting the microarrays, and ontological classification data for *Drosophila* genes. The database, FlyAtlasDB, was constructed from these data in the relational database management system, MySQL. The database schema can be

found in Supplementary Figure S2 and the table attributes in Supplementary Figure S3.

The experimental data were initially in a file listing the following six items for each probe set: the tissue and stage, the number of Genechip replicates in which a signal was detected (out of the four), the mean hybridization signal (arbitrary units), the standard error of the latter, the enrichment with respect to the signal in whole flies (1.0) and whether the enrichment represented an increase, decrease or was not significantly different from the value for whole flies. These data were used to construct the ‘Experiment’ and ‘FlyAnat’ tables (Supplementary Figures S2 and S3).

Documentation of the *Drosophila* Genome 2 microarray was in the standard Affymetrix annotation file, currently *Drosophila\_2.na30.annot.csv* and downloadable from <http://www.affymetrix.com/analysis/index.affx>. It consists of a row for each probe set, together with about 40 columns of information regarding the probe set and the *Drosophila* gene or genes that it detects. Of the various gene descriptors available, we chose the FlyBase identifier as definitive. We extracted each probe set identifier and the corresponding FlyBase identifier from the file, and the latter were then checked against FlyBase (5) to determine whether they were current, using a program written for the purpose. Twenty-four were found to be outdated because what had originally been thought to be a single gene had subsequently been resolved into two or more. In these cases, each probe set was reassigned to a new identifier by visual inspection using the FlyBase genome viewer. Another 272 FlyBase identifiers were found to refer to species of *Drosophila* other than *melanogaster*, and these and their corresponding probe sets were removed from the database. Probe sets without a corresponding FlyBase identifier—notably internal controls—were also excluded. Definitive versions of gene names, symbols and CG numbers were derived from FlyBase.

These data were used as the basis of the ‘Probeset’ and ‘Gene’ tables (Supplementary Figures S2 and S3). The ‘Probeset’ table also contained an additional pre-computed field ‘ProbeDegeneracy’, indicating whether a probe set was degenerate, i.e. hybridized to the messenger RNA from more than one gene. This was the case for 1188 of the 14438 probe sets. The ‘Gene’ table contains between 100 and 200 gene symbols and names that include Greek characters; so to allow search flexibility, additional fields were included with Romanized equivalents (e.g. with ‘ $\alpha$ ’ replaced by ‘alpha’).

Because of errors and ambiguities in the ontological data in the Affymetrix file, ontological data were downloaded from FlyBase and used to populate the ‘OntolOfGene’ and ‘Ontology’ tables (Supplementary Figures S2 and S3). These tables provide the functional descriptions of *Drosophila* genes that are used in ‘Category’ searches (below).

## TECHNICAL IMPLEMENTATION

The FlyAtlas web application uses a Java servlet to generate web pages and communicate with the relational database. Java packages from the Apache commons

Mathematics Library (<http://commons.apache.org/math/>) are used to calculate correlations for the 'Profile Search' (below). A separate smaller version of the servlet is used to provide programmatic access for developers.

The persistence of servlet instances depends on new HTML pages being generated from HTML form requests to the server. To avoid some of the constraints that this can place on interface design, JavaScript is used to access variables from within the page and generate and send a 'hidden' form when the user initiates a request. Thus, although Java is not required by the client web browser, JavaScript must be enabled. The JQuery JavaScript library (<http://jquery.com/>) is used for the hide/show interface features, together with the jqBarGraph plug-in for generating the bar charts for the 'Profile' search.

The 'autosuggest' uses AJAX: a JavaScript file adapted from a published source (7), a custom Java servlet and FlyAtlasDB.

## USER INTERFACE & FUNCTIONALITY OF THE WEB APPLICATION

### Presentation of tables of results

All of the search queries in the new version of the FlyAtlas web application (FlyAtlas 2) return pages that include experimental data in tabular form, formatted as shown in Figure 1. To allow comparison of corresponding tissues in adult flies and larvae, the data for these are presented side by side, an arrangement that also decreases the length of the table. Cases in which none of the arrays gave significant results are indicated by 'ND' (Not Detected), and for those in which significant signals were only obtained with between one and three arrays, the values are presented in square brackets. In the latter case, the number of arrays can be seen by holding the cursor over the value (Figure 1A). The default presentation of results shown in Figure 1A lacks standard errors, but the user may select an option to display them (Figure 1B), and once selected this choice persists between different search modes until changed.

The results table in Figure 1 uses colour to draw the user's attention to the features that are likely to be of most interest and importance. Abundance is represented on a logarithmic scale (Base 2) from white to black, whereas enrichment uses a scale running from white, through yellow, to red. Yellow is used for an enrichment value of 1 (no enrichment compared with expression in whole flies), increased enrichment runs through orange to deep red and as expression decreases the colour diminishes from yellow to white (which matches the bottom of the abundance scale). An asymmetric logarithmic scale is used to optimize perception of both relatively modest and much greater changes.

The differentiation of colours used in both tables and bar charts ('Profile' search, below) was checked using a simulator (<http://www.colblindor.com/coblis-color-blindness-simulator/>) and found to be adequate in different cases of defective colour vision.

### Search interface: concerns and approach

A major concern in designing the search interface was to address the problem mentioned in the Introduction: that of the amount and relevance of the data returned in response to valid search queries. However, we also wished to avoid, as far as possible, the situation in which a user is allowed to submit an invalid search query, wait and then receive an uninformative negative response. Our approach to both concerns has been to make extensive use of AJAX-based 'autosuggest' menus. These menus present users with a list of corresponding entries in the database, which, if used as search terms, ensure that results are returned and that these results are relevant.

Another concern was what search facilities to provide. We have assumed that users will approach the web application from one of two standpoints. In the first case, we envisage that they will be primarily interested in a specific gene or group of genes, and wish to obtain information about how these genes are expressed in different tissues. Two search facilities ('Gene' and 'Category') are designed for this purpose. In the second case, we envisage that users will primarily be concerned with a particular tissue, and wish to obtain information about genes expressed there. Three other search facilities ('Top', 'Development' and 'Tissue') address this requirement. There is a sixth—and completely new—facility, the 'Profile Search'. Although this starts from the standpoint of a particular gene, it involves expression across different tissues. Its function is to find other genes with a pattern or profile of tissue expression that is similar to the query gene.

### Gene search

The 'Gene' search allows one to determine the expression of a particular gene in the tissues for which experimental data are available. The search entry form (Figure 2A) requires that before typing a gene descriptor in the search box one specifies whether it is a symbol, name, annotation symbol (CG number) or FlyBase identifier (FBgn number). If one clicks on the 'options' box, a panel of additional options is presented (inclusion of standard errors, duplicate and ambiguous results—see below) but the default settings of these options will be appropriate for most users, so they are initially hidden to simplify the interface.

The requirement for the user to specify the type of gene descriptor to be entered relates to the specificity of the 'autosuggest' facility that operates as the descriptor is entered into the search box. This facility is tailored to specific aspects of the nomenclature of symbols for *Drosophila* genes, as documented in Supplementary Figure S4. The autosuggest menu is populated with terms from the descriptor field in the database, and only appears after an appropriate number of characters is entered. If the user chooses a search term from the autosuggest menu, the default settings (i.e. excluding duplicates and ambiguous hits) will guarantee that either a single result will be returned or a notification that the gene in question is only represented by an ambiguous probe set. If a term is not present in the autosuggest menu, it will not be present in the database, and searches with such terms

**A** Results found for 'abd-A'.

Tissue	Adult Abundance	Adult Enrichment	Larval Enrichment	Larval Abundance
Head	ND	ND		
Eye	[11]	[0.2]		
Adult Brain / Larval CNS	ND	ND	2.7	118
Thoracoabdominal ganglion	87	2.0		
Crop	[7]	[0.2]		
Midgut	21	detected in 1 out of 4 array(s)	0.72	31
Hindgut	21	0.50	0.76	33
Tubule	ND	ND	[0.2]	[7]
Fat body	22	0.49	13	552
Salivary gland	[9]	[0.2]	ND	ND
Heart	111	2.5		
Trachea			33	1434
Ovary	19	0.40		
Virgin Spermatheca	787	18		
Mated Spermatheca	948	22		
Testis	60	1.4		
Accessory glands	86	2.0		
Carcass	86	2.0	18	796

Tissue	Adult Abundance	Adult Enrichment	Larval Enrichment	Larval Abundance
Tubule	ND	ND	[0.2]	[7 ± 0.5]
Fat body	22 ± 5	0.49	13	552 ± 49

**Figure 1.** Presentation of experimental data in the FlyAtlas 2 web application. (A) Appearance without standard errors. The tooltip—obtained on holding the cursor over values in square brackets—indicates the number of replicates in which a signal was detected, if between one and four. 'ND' indicates a value of '0' for the latter. Note the green 'D' icon indicating that other probe sets besides 1637813\_at detect gene abd-A. Clicking on the arrow icon to the right of this allows one to launch a page with links for this gene in various external *Drosophila* resources. The green button on the top left with the '^' symbol allows the table to be 'collapsed' or hidden. (B) Appearance with standard errors (detail).

will naturally be futile. The single table of results is presented in an uncollapsed form, as in Figure 1, and is preceded by two lines listing the available gene descriptors and the probe set identifier. For users wishing further information about a particular gene, a link invokes a menu of external *Drosophila* resources for that gene.

The default response to genes that are only represented by ambiguous probe sets is to exclude their results, and that to genes represented by duplicate probe sets is to present the results with the probe set that gives the highest/most significant signals. This is to protect the user from more questionable data (often one duplicate has low and unreplicated signals). However the user is always alerted to the situation when duplicate or ambiguous results are available but not included, and is advised of the possibility of re-running the search with the relevant option changed. In the case of ambiguous results, as already mentioned, the alert is textual. In the case of duplicate results, the situation is indicated by a green 'D' icon (Figure 1A), with explanatory text available in a tooltip. Likewise, if a user does run a search having

selected to include ambiguous hits, the results are flagged with a red 'A' icon, and the FlyBase IDs of other genes detected by the probe set are listed. The user can then resolve the ambiguity by examining results for any unambiguous duplicate probe sets detecting these other genes. In general, this is unlikely to be productive, however.

### Category search

In the 'Category' search (Figure 2B), one does not specify a single gene, but rather chooses a term that encompasses a group of genes. The categories that are used for this search are those in the Ontology table, and the default option in the search entry form is to select from an autosuggest menu populated with descriptions that include the term that has been typed. There is also the option of specifying a particular Gene Ontology ID as the search term for a 'Category' search.

Gene Ontology descriptions are in many cases narrow, and although in some cases this is what is required, in others wider search terms may be more appropriate. The third option, 'Free Search', provides such wider scope,

**A** For a particular *Drosophila* gene, find the pattern of expression in different tissues.

Gene Symbol (e.g. vkg) — start typing, then select from the autosuggest menu  
 Gene Name (e.g. viking)  
 Annotation symbol (e.g. CG16858)  
 FlyBase ID (e.g. FBgn0016075)

Gene:

Options

Include duplicate hits <sup>D</sup>       Include ambiguous hits <sup>A</sup>  
 Show standard errors

**B** For a particular category of *Drosophila* genes, find how each gene is expressed in different tissues.

Category ( Type a term of interest — e.g. 'wing' or 'kinase' — then select from the autosuggest menu.)  
 Gene Ontology ID (e.g. 0005201)  
 Free Search ( Type a partial or complete term — e.g. 'mitochon'. No autosuggest, all matching categories searched.)

Term:  Display:

Options

**C** For a particular tissue, find which genes have the greatest absolute expression (abundance) or relative expression (enrichment).

Select a stage and a tissue, choose abundance or enrichment, how many 'top genes' you wish listed, and press 'Search'.

Greatest:  Display:

Options

**D** For a particular tissue, find which genes show the greatest difference in expression between larva and adult.

1. Select whether expression should be greater in the adult or larval stage.

Adult > Larval  
 Larval > Adult

2. Choose the tissue, abundance or enrichment, how many genes you wish listed, and press 'Search'.

Tissue:  Greatest:  Display:

Options

**E** For a particular tissue, find which genes in a selected category have the greatest expression.

1. Select a stage and a tissue, and choose either abundance or enrichment.

Greatest:

2. Type a term of interest (e.g. wing, kinase) then select from the categories (gene ontologies) in the autosuggest menu.

Category:  Display:

Options

**F** For a *Drosophila* gene of interest, find others with a similar profile of expression across tissues.

Gene Symbol (e.g. vkg) — start typing, then select from the autosuggest menu  
 Gene Name (e.g. viking)  
 Annotation symbol (e.g. CG16858)  
 FlyBase ID (e.g. FBgn0016075)

Gene:  Display:

Options

Pearson  Spearman correlation       $r >$    
 Show standard errors

**Figure 2.** Form entry for different types of search. (A) 'Gene' search—the 'Options' panel is shown open (not the default) in this case; (B) 'Category' search; (C) 'Top' search—stage and tissue selections have already been made; (D) 'Development' search; (E) 'Tissue' search—stage, tissue and display selections have already been made; (F) 'Profile' search—the 'Options' panel has been opened.

returning genes corresponding to all ontological descriptions that the search term includes. Searches with a frequently occurring string can return results from a variety of different gene ontology classes. The output for this type of search, therefore, includes the matching Gene Ontology ID for each gene listed, with the description of the gene ontology in a tooltip (Figure 3). (The reason Gene

Ontology IDs are not included in the output of 'Gene' searches is that many genes are assigned to a large number of ontological classes).

Because 'Category' searches can produce many hits, the user has an option to limit the number of results retrieved, and the latter are presented in collapsed form initially, with only the documentation visible. Individual

hide/show buttons are provided at the left of each listing, and a master hide/show control is positioned at the top right (Figure 3).

As with the 'Gene' search, there is the option to include duplicate and ambiguous hits. The user is made aware of any excluded ambiguous hits by a listing of their FlyBase IDs at the end of the output.

### Top search

The 'Top' search (Figure 2C) is the most straightforward search facility focussed on tissues, and allows retrieval of the most highly expressed genes for a particular adult or larval tissue. The basis of the ranking can be either 'abundance' (the absolute extent of expression) or 'enrichment' (the expression relative to the average in whole flies). One can select to view from 20 to 50 'top' genes, and, as with the 'Category' search, the results are presented in collapsed form. For this type of search, it was decided to withhold the option to include duplicate or ambiguous results, as we believed that the user expected the results to be definitive. However, any genes detected by duplicate probe sets are flagged in the usual way so that the user has the option of investigating them further in the 'Gene' search facility.

### Development search

The 'Development' search (Figure 2D) is a new facility, similar to the 'Top' search, but rather than retrieving the most highly expressed genes for particular tissues, it retrieves those that show the greatest difference in expression between the adult and larval stages. Currently this facility is only available for the seven tissues in which we have data for both stages. The user chooses whether to view genes that are most highly expressed at the adult or the larval stage of the selected tissue and has the additional generic options available in the 'Top' search.

### Tissue search

Although, like the 'Top' search, the 'Tissue' search facility (Figure 2E) allows one to focus on an individual tissue, it differs from the former in that this is confined to a specific

27 genes found using the search term 'calcium channel', 25 shown

**Ca** **Ca-1D**  
 Annotation Symbol: CG4894 FlyBase ID: FBgn0001991 Probe Set ID: 1640092\_at Gene Ontology (GO): 0005245 0005891 00051925

**narrow abdomen** **na**  
 Annotation Symbol: CG1517 FlyBase ID: FBgn0002917 Probe Set ID: 1635546\_at Gene Ontology (GO): 0005245

**transient receptor potential** **trp**  
 Annotation Symbol: CG7875 FlyBase ID: FBgn0003861 Probe Set ID: 1622920\_at Gene Ontology (GO): 0005218 0005262 0008086 0008087 0015279  
 light-activated voltage-gated calcium channel activity

**trp-like** **trpl**  
 Annotation Symbol: CG18345 FlyBase ID: FBgn0005614 Probe Set ID: 1633649\_at Gene Ontology (GO): 0005262 0008086 0015279

**Figure 3.** Presentation of output from a 'Category' search. Results are listed in a 'collapsed' form, with only the first four included. The green button at the left of each entry allows individual tables to be shown, and that at the top right allows all tables to be shown. Note that the gene ontologies corresponding to the search term 'calcium channel' differ for different genes. (The description for one of them is shown in a tooltip).

group of genes, selected from gene ontologies as for the 'Category' search. Furthermore, rather than ranking within this category and returning results for e.g. the 'top 20', the search returns results for all genes that are expressed to a greater extent than the overall tissue average (abundance or enrichment, as selected), provided that this increase is statistically significant.

This facility does have the option to include duplicate and ambiguous hits, and, as with the 'Category' search, the FlyBase IDs of any excluded ambiguous hits are listed at the end of the output.

### Profile search

The 'Profile' search (Figure 2F) is a new facility that takes a query gene and compares the pattern of expression in the corresponding probe set to the others in the database, returning those with a correlation coefficient ( $r$ ) greater than a particular cut-off value (the default is 0.7). Thus, the user can identify other genes with a similar pattern of expression across tissues. The actual comparisons are of  $\log_2(\text{abundance})$ , and signals that the Affymetrix software classifies as 'not detected' are treated as 0. The Pearson correlation coefficient is used by default, but the option of using the Spearman correlation coefficient is also provided (Figure 2F). The user can also vary the cut-off value of the correlation coefficient,  $r$ ; but a fixed cut-off of 0.05 is used for the Bonferroni-corrected probability  $P_B$ .

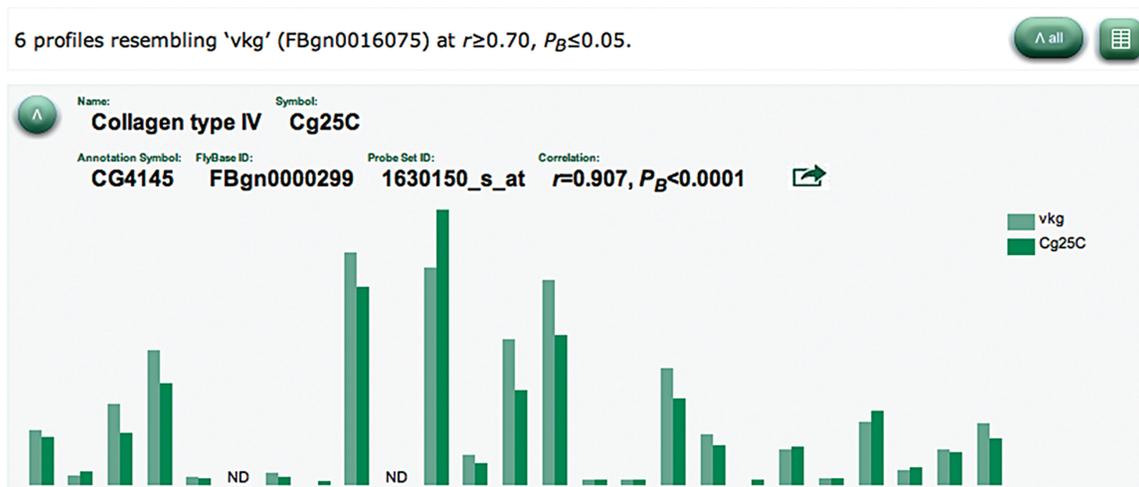
The default display is a bar chart, although a button is provided allowing the user to switch to tabular presentation of the results (Figure 4). A strong and statistically significant correlation between patterns of gene expression suggests co-ordinated regulation, and could potentially identify genes with related functions.

### PROGRAMMATIC ACCESS TO THE DATABASE

In addition to the web interface, and the free provision of the data for download (<http://130.209.54.32/atlas/20090519all.txt> and <http://flyatlas.gla.ac.uk/flyatlas/downloads/FlyAtlasDB.sql>), we have provided the means for developers to make 'Gene Search' queries directly. Programs can be written to make queries through HTTP requests and retrieve the results of such queries in either tab-separated text or extensible markup language (XML) format. Documentation (APIs) can be found at <http://flyatlas.gla.ac.uk/flydirect/docs.html>.

### Discussion

There is no doubt that FlyAtlas has proved to be a valuable scientific resource: for example, there have been >30 000 accesses to the website during the past 2 years, and Google Scholar lists 451 citations to the original article, but identifies 650 articles that mention FlyAtlas in the full text. Even this is an understatement of the true uptake of the dataset because it is now also served through the FlyBase and FlyMine portals. There is thus a demand for such data across a broad community that extends well beyond Drosophilists.



**Figure 4.** Bar chart output from 'Profile' search. The button to switch to tabular view is at the top right.

The recent restructuring of the underlying data and the redesign of the web interface should add further to the value of FlyAtlas. It is now easier for the user to select valid search terms, and the data returned is more relevant, more manageable in volume and easier to assimilate. We also believe that the recently added 'Profile Search' facility will provide new avenues for scientific investigation.

There are some aspects of the web facility that leave scope for improvement. The 'Category Search' options, aping as they do the ontology descriptions, will, no doubt, be too narrow in scope for certain searches. The structure of gene ontologies is not hierarchical (8), so that any smaller menu of broader categories will probably require manual intervention. The clarity of presentation of the bar charts in the 'Profile Search' facility might also be improved.

## FUTURE DIRECTIONS

The restructuring of the web interface to FlyAtlas is recent at the time of writing, so that changes can be anticipated both as a result of planned additions and in response to user feedback. We also envisage the addition of new data, and hope shortly to be able to expand the number of tissues from the existing 25. As the data are now in a relational database, there is the opportunity to serve them to the semantic web—for example, using the D2RQ software to generate virtual resource description framework (RDF) graphs (<http://d2rq.org/>). The facility would be even more valuable if RNAseq data were available to allow comparison of the transcripts found when a gene is expressed in different tissues. This would be an exciting prospect, despite being a major undertaking and entailing a further overhaul of the database and web interface.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–4.

## ACKNOWLEDGEMENTS

The authors thank Dr Helen Purchase for helpful suggestions regarding the visual presentation of results.

## FUNDING

This work was funded by the BBSRC. Funding for open access charge: General funds of the University of Glasgow and Glasgow Polyomics.

*Conflict of interest statement.* None declared.

## 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.
- Reiter, L.T., Potocki, L., Chien, S., Gribskov, M. and Bier, E. (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res.*, **11**, 1114–1125.
- Chintapalli, V.R., Wang, J. and Dow, J.A.T. (2007) Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.*, **39**, 715–720.
- Ashburner, M. and Drysdale, R. (1994) FlyBase — the *Drosophila* genetic database. *Development*, **120**, 2077–2079.
- McQuilton, P., St Pierre, S.E. and Thurmond, J.; FlyBase Consortium (2012) FlyBase 101 – the basics of navigating FlyBase. *Nucleic Acids Res.*, **40**, D706–D714.
- Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., McLaren, P., North, P. *et al.* (2007) FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. *Genome Biol.*, **8**, R129.
- Asleson, R. and Schutta, N.T. (2005) *Foundations of Ajax*. Apress, New York.
- 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. *Nat. Genet.*, **25**, 25–29.