

FlyTED: the *Drosophila* Testis Gene Expression Database

Jun Zhao¹, Graham Klyne¹, Elizabeth Benson², Elin Gudmannsdottir²,
Helen White-Cooper^{2,*} and David Shotton^{1,*}

¹Image Bioinformatics Research Group and ²*Drosophila* Spermatogenesis Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received August 24, 2009; Revised and Accepted October 16, 2009

ABSTRACT

FlyTED, the *Drosophila* Testis Gene Expression Database, is a biological research database for gene expression images from the testis of the fruit fly *Drosophila melanogaster*. It currently contains 2762 mRNA *in situ* hybridization images and ancillary metadata revealing the patterns of gene expression of 817 *Drosophila* genes in testes of wild type flies and of seven meiotic arrest mutant strains in which spermatogenesis is defective. This database has been built by adapting a widely used digital library repository software system, EPrints (<http://eprints.org/software/>), and provides both web-based search and browse interfaces, and programmatic access via an SQL dump, OAI-PMH and SPARQL. FlyTED is available at <http://www.fly-ted.org/>.

INTRODUCTION

Our activities

We have determined the mRNA expression patterns of genes involved in spermatogenesis in the *Drosophila* testis, including many that show differences in expression level between wild type flies and meiotic arrest mutant strains exhibiting abnormal spermatogenesis. Gene expression studies in the *Drosophila* testis have the advantage of a clear correlation between the position of the developing germ cell within the elongated testis and its developmental stage (see diagram at http://www.fly-ted.org/images/Spermatogenesis_diagram.png). It is thus possible to infer at what developmental stage a particular gene product is likely to act, simply by observing its mRNA expression pattern.

We have created the *Drosophila* Testis Gene Expression Database (FlyTED; <http://www.fly-ted.org>) to provide web access to these gene expression images and their metadata, including the primer sequences that ultimately define the gene product being localized (1). This research and development, ongoing since 2003 with funding from BBSRC and the JISC, has led to the publication of expression images of 817 genes, about 10% of all genes expressed in the testis and the male genital tract, and have given new understanding about post-meiotic gene expression by the discovery of two hitherto unknown classes of *Drosophila* genes named ‘comets’ and ‘cups’, whose expression show characteristic sub-cellular localization patterns that proved hitherto unknown post-meiotic gene expression (2,3).

Other *Drosophila* image databases

There are a number of other public databases of *Drosophila* gene expression images, mostly showing expression in the *Drosophila* embryo, that complement FlyTED. These include the Berkeley *Drosophila* Genome Project expression database (BDGP; <http://www.fruitfly.org/>) (4); Fly-FISH (<http://fly-fish.cabr.utoronto.ca/>), a new database of mRNA localization patterns at the subcellular level during early *Drosophila* embryogenesis determined by fluorescence *in situ* hybridization (5); FlyView (<http://flyview.uni-muenster.de/>), that contains pictures from enhancer-trap lines; FlyMove (<http://flymove.uni-muenster.de/>) providing didactic images, movies and interactive diagrams of the embryonic development of *Drosophila melanogaster* (6); FlyEx (<http://flyex.ams.sunysb.edu/FlyEx/>), showing embryonic segmentation gene expression patterns (7); FlyBrain (<http://flybrain.neurobio.arizona.edu/>), an online *Drosophila* nervous system atlas that contains some gene expression data revealed by antibody labelling;

*To whom correspondence should be addressed. Tel: +1865 271193; Fax: +1865 310447; Email: david.shotton@zoo.ox.ac.uk
Correspondence may also be addressed to Helen White-Cooper. Tel: 029 2087 5492; Fax: 029 2087 4116 ; Email: white-cooperh@cardiff.ac.uk
Present address:

Helen White-Cooper, School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK.

Elizabeth Benson, Journal of Biology, 336 Gray's Inn Road, London WC1X 8HL, UK.

Elin Gudmannsdottir, Department of Social Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.

FlyProt (<http://www.flyprot.org/>), an exon trap database containing *Drosophila* gene expression images from all stages of development and tissue types; and FlyTrap (<http://flytrap.med.yale.edu/>), a protein trap database (8).

FlyBase (<http://flybase.org/>), the definitive global database of information concerning the genes and genomes of several *Drosophila* species including *D. melanogaster*, while containing no gene expression image data, is of relevance to all of the aforementioned image databases. Sites that integrate these distributed heterogeneous resources include FlyMine (<http://www.flymine.org/>) and 4DXpress (<http://4dx.embl.de/4DXpress/>) (9,10). Additionally, we have created a small demonstration system, OpenFlyData (<http://www.openflydata.org/>), to show how Semantic Web technologies can be used to integrate information from FlyTED, FlyBase, FlyAtlas (<http://www.flyatlas.org/>) (11) and BDGP into a single user interface 'on the fly' (12).

A number of global testis-specific microarray studies have been published (13–15). We have conducted our own microarray analysis to compare gene expression in wild type testes with that in several meiotic arrest mutants (White-Cooper, H., unpublished data), and have used these data, in conjunction with the published array and EST data, to identify testis-expressed genes. A subset of these testis-expressed genes were selected for analysis by mRNA *in situ* hybridization. Most of the genes selected for analysis were dependent on the meiotic arrest genes for full expression in testes, while others were expressed independently of the meiotic arrest genes.

DATABASE METHODS

About the dataset

The primary spermatocyte stage of *Drosophila* spermatogenesis lasts ~3.5 days, and is characterized by extensive cell growth, associated with activation of expression of a large repertoire of testis-specific genes. Typically, primary spermatocytes transcribe genes required in the primary spermatocytes themselves and in the spermatids that develop from them. The transcripts required after meiosis are stabilized and stored in the cytoplasm in a translationally repressed state for up to 4 days (16). In *Drosophila* spermatogenesis, meiotic cell cycle progression is linked to spermatid differentiation by the function of the meiotic arrest genes. Mature primary spermatocytes in testes from a meiotic arrest mutant male arrest during differentiation, and show no signs of entering either the meiotic divisions or spermatid differentiation (17,18). These meiotic arrest genes fall into two phenotypic classes—*aly*-class (*aly*, *comr*, *topi*, *tomb* and *achi/vis*) (19–24), and *can*-class (*can*, *mia*, *sa*, *nht*, *rye*) (25,26). The failure of meiotic arrest mutant germ cells to progress past the mature primary spermatocyte stage is due to failure to activate expression of genes required for meiotic cell cycle progression (e.g. *twine*) and for spermatid differentiation (e.g. *fzo*) (27). One of our aims has been to determine the expression patterns for genes that require the meiotic arrest genes for their expression, in comparison with those whose expression in testes is

independent of the meiotic arrest genes. To achieve this, since the mutant testes are morphologically easily distinguished from wild type testes, the two genotypes are mixed and stained in the same hybridization well. Thus, although mRNA *in situ* hybridization is not quantitative, qualitative judgements of gene expression level in mutant versus wild type can be made on the basis of side-by-side comparisons.

Data acquisition

The experimental methodology used to obtain the gene expression images within FlyTED is summarized at <http://www.fly-ted.org/meth.html>, and is more fully documented by White-Cooper (28). In brief, testes from young male *Drosophila* (0–1 day old) were dissected, hybridized to probes specific for the gene under study, stained and then examined using DIC microscopy, typically using a 10× objective magnification. Images were captured with a digital colour camera and were not subjected to post-capture digital manipulations. For each gene, pictures were taken of at least one wild type and one mutant strain testis, with additional pictures, including higher magnification views, being taken if the staining pattern looked interesting. Images were also acquired if staining occurred in the somatic cells of the testis.

Metadata structure

Every FlyTED image was annotated manually at the time of capture by the biologist concerned, with metadata that is compliant with the emerging MISFISHIE standard (29). The gene expression pattern revealed in each image is described using controlled vocabulary terms from the *Drosophila* Anatomy Ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=fly_anatomy). In addition to the gene name, each image is also annotated with the FlyBase identification number (gene id) that uniquely identifies each gene, which is linked to the corresponding FlyBase gene report page, and with the CG (Computed Gene) number, by which biologists can search FlyTED if they are not familiar with the gene name used.

Database content

FlyTED, the *Drosophila* Testis Gene Expression Database, was constructed by customizing an instance of the EPrints open source repository software system (<http://eprints.org/software/>), as detailed on the 'About the Database' page of the FlyTED website. Currently, the database contains 2762 mRNA *in situ* hybridization images and ancillary data revealing the patterns of expression of 817 individual genes involved in spermatogenesis in the testis of the fruit fly, *D. melanogaster*, both in flies with normal spermatogenesis (wild type; typically, we used the strain *red e*), and in seven meiotic arrest mutant strains of flies exhibiting abnormal spermatogenesis: *aly* (always early), *achi/vis* (*achintya* and *vismay*), *can* (*cannonball*), *comr* (*cookie monster*), *nht* (*no hitter*), *tomb* (*tombola*) and *topi* (*matotopetli*). Full details of the alleles used are at <http://www.fly-ted.org/meth.html#strain> (19,21,22, 24,30). The database also contains a small number of

Browse by Expression Location

Please select a location to browse from the list below. (The numbers in parenthesis show the number of images for each expression location and its subtypes.)



Expression Location: Cup-like_pattern_of_distal_end_of_elongating_spermatids

- Subjects for annotation (2689)
 - Other annotations (671)
 - Cup-like_pattern_of_distal_end_of_elongating_spermatids (81)

Number of items at this level: 81.

Gene CG10113 in wt Gene CG10113 in wt Gene CG10113 in wt Gene CG10113 in wt

Gene CG10113 in wt Gene CG10113 in wt Gene CG10113 in wt Gene CG10113 in wt

- Subjects for annotation (2689)
 - Drosophila_male_reproductive_system (2215)
 - Ejaculatory_bulb
 - Ejaculatory_duct (1)
 - Male_accessory_gland (30)
 - Seminal_vesicle (4)
 - Testicular_duct (1)
 - Testis (2193)
 - Drosophila_testis_cell (2188)
 - Male_germline_cell (2148)
 - Male_germline_stem_cell
 - Primary_gonial_cell
 - Primary_spermatocyte (2130)
 - Secondary_spermatocyte
 - Secondary_spermatogonium (2)
 - Spermatid (854)
 - Elongation-stage_spermatid (854)
 - Early-mid_elongation-stage_spermatid (98)
 - Early_elongation-stage_spermatid (345)
 - Comet-stage_spermatid
 - Leaf_blade-stage_spermatid
 - Late_elongation-stage_spermatid (92)
 - Mid-late_elongation-stage_spermatid (128)
 - Mid_elongation-stage_spermatid (191)
 - Individualization-stage_spermatid
 - Pre-elongation_spermatid
 - Coalescence-stage_spermatid
 - Onion-stage_spermatid
 - Spermatozoon
 - Somatic_cell_of_testis (49)
 - Waste_bag (8)
 - Vas_deferens (3)
 - Other annotations (671)
 - Comet-like_pattern_of_distal_end_of_elongating_spermatids (116)
 - Cup-like_pattern_of_distal_end_of_elongating_spermatids (81)
 - No_staining (460)

Figure 1. FlyTED browse results, presented as an array of captioned image thumbnails, for the gene expression location Cup-like_pattern_of_dismal_end_of_elongating_spermatids.

images of testis gene expression in *Drosophila pseudo-obscura*. For most genes, the PCR primer sequences used (designed from genomic sequences) and the predicted sequence of the PCR reaction are also included in the database.

The *Drosophila* thumbnail images and accompanying metadata contained within the FlyTED Database are published as Open Access Data under the Creative Commons CC0 1.0 Universal License, in conformity with the Science Commons Protocol for Implementing Open Access Data (<http://sciencecommons.org/projects/publishing/open-access-data-protocol/>), and may be freely downloaded and reused for any purpose, or aggregated with third party metadata using tools such as OpenFlyData.org, without attribution. In contrast, the higher resolution *Drosophila* testis gene expression images contained within the FlyTED Database are published under the Creative Commons Attribution 2.0 UK: England & Wales License, which permits reuse only if each image is attributed to Dr Helen White-Cooper. Further details concerning licensing and attribution are given on the 'About the Database Licenses' page of the FlyTED website. The database is not open for public data submission.

DATABASE ACCESS METHODS

Browsing interface

Users can browse FlyTED by gene name, strain name or gene expression location. For example, the 'Browse by

Gene Name' view groups all the FlyTED images according to the gene names associated with the images, and construct a Web page listing the 817 gene names currently in our database, in alphabetical order. Each name links to one of 817 Web pages in which thumbnails of all images recorded for that gene name are displayed, ordered by strain name, each annotated with a caption containing its gene and strain name. This allows users to compare the images from different strains relating to a single gene together in a single page. Hovering the mouse over the centre of a thumbnail gives a pop-up box containing a description of the staining pattern. Once an individual image record has been selected, the user can view both the full-size image (by clicking on the thumbnail) and also a detailed information page containing an intermediate-sized image, descriptive metadata about the image and a link to the FlyBase database (by clicking on the image caption).

Similar presentation of images is given in three other FlyTED browse views: the 'Browse by CG Number' view, that groups images by the CG number; the 'Browse by Strain Name' view, that groups images by the strain of fly from which the images were acquired; and the 'Browse by Expression Location' view, that groups images by the pattern of gene expression revealed in the images. In the last case, because our images are annotated using controlled terms from the extended Fly Anatomy Ontology, users can browse images using the hierarchical structure of the ontology, as shown in

A

Order the results

Synonyms:

Strain: Restrict your search to the following strains: wt, aly, av, comr, nht, tomb, and topi.

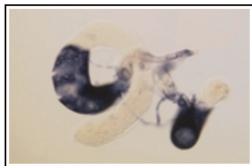
Expression Location: Select one or more values from the list (Ctrl+click), and whether you want to search for records with any one or all of those values. Default is (All).

Retrieved records must fulfill:

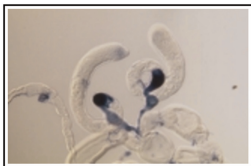
B

Synonyms matches any of "CG18628, MtnA" AND Expression Location matches "Terminal_epithelial_cell"

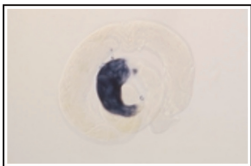
Displaying results 1 to 4 of 4.
[Refine search](#) | [New search](#)



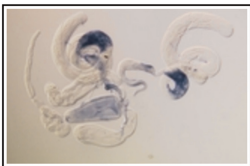
1. [Gene CG18628 in av](#)



2. [Gene MtnA in av](#)



3. [Gene CG18628 in wt](#)



4. [Gene MtnA in wt](#)

Figure 2. Example of an advanced search, a search for images of genes CG18628 and *MtnA* that are expressed in a terminal epithelial cell. (A) The interface for entering the search conditions. (B) The search results with image thumbnails.

Figure 1. The number next to each term indicated how many images in FlyTED are annotated using that term and its sub-terms.

On the FlyTED home page, in addition to a general description of the database and a few exemplar images, users can also find links to pages providing details of the dataset and the database, other *Drosophila* resources such as FlyBase, and further relevant information. The footer on all pages of the database displays the license statements given above.

Search interfaces

We provide both a simple and an advanced search interface to permit users to make specific queries across the database content. The simple search interface allows querying for images by gene name, CG number or FlyBase gene id. Queries for multiple genes can be achieved by separating the names with commas. The advanced search interface (Figure 2A) supports more complex queries, allowing users to search across multiple gene names, and/or strain names, and/or gene expression locations. The image results are presented as a tiled array

of captioned thumbnails (Figure 2B), allowing users to compare them side by side. Again, enlarged images can be viewed by clicking on the thumbnail images, and metadata can be displayed by clicking on the captions.

Programmatic access

In addition to the conventional Web interfaces permitting human access to the FlyTED Database, programmatic access is provided as detailed in the 'How to use the Database' page on the FlyTED website, involving either a database SQL dump, OAI-PMH access (<http://www.openarchives.org/OAI/openarchivesprotocol.html>), or queries against a SPARQL endpoint (<http://openflydata.org/query/flyted>) (31).

DATABASE INTEROPERABILITY

In FlyTED, we provide links to FlyBase, the central *Drosophila* genomic database. More flexible cross search of *Drosophila* information can be found in our demonstration *Drosophila* data web application OpenFlyData (<http://openflydata.org>), a web application that allows

scientists to cross search for *Drosophila* gene expression information from FlyTED, BDGP and FlyAtlas using any synonyms of a gene, either individually, by a batch of gene names or by gene expression profiles. Data integration between distributed resources containing heterogeneous data is a difficult task for which various approaches have previously been proposed (32). Our novel use of Semantic Web technologies in OpenFlyData has proven their value in promoting interoperability between the data resources, and in lowering the cost of development. For this, accurate cross-database mapping of gene names and identifiers was a key prerequisite (12). However, the maintenance of such mapping between different identifiers in a reliable way, during the ongoing churn of database revisions and updates so eloquently described by Stein (32), presents a separate problem. In recent papers (33,34), we have proposed methods employing a set of RDF patterns called Named Graphs (<http://www.w3.org/2004/03/trix/>) (35) that can be adopted to express provenance information about data identifier mappings and to record nomenclature changes. Adoption of these patterns would permit database updates to be documented in machine-processable ways, and would allow third-party annotations made using an old nomenclature to be interpreted correctly in terms of a revised or updated nomenclature.

CONCLUSION

We report the creation by our Image Bioinformatics Research Group of FlyTED, a biological database for images of gene expression in the testis of *D. melanogaster* obtained by our *Drosophila* Spermatogenesis Research Group, the biological significance of which has been reported in the papers referenced above. FlyTED was created by adapting an existing software system, EPrints, and provides both human and programmatically accessible interfaces to the images and their metadata. In collaboration with the curators of the Fly Anatomy Ontology, we have corrected and expanded that section of the ontology dealing with the male reproductive system, to permit appropriate descriptions to be made in FlyTED of the germ cell developmental stages in which particular genes are expressed. FlyTED data acquisition is largely complete, although we are continuing work on a number of genes not yet characterized in testis that will be added to FlyTED at a later date. This work has triggered us to consider novel solutions to problems of database interoperability, including the creation of OpenFlyData, a data web to integrate *Drosophila* gene expression information.

ACKNOWLEDGEMENTS

The authors acknowledge the technical advice given by members of the University of Southampton EPrints team, particularly Christopher Gutteridge, and fruitful discussions with members of the Semantic Web research community.

FUNDING

UK Biotechnology and Biological Sciences Research Council (BBSRC BB/C503903/1 to H.W.-C. and D.S., BB/E018068/1 to D.S. and H.W.-C., BB/D009324/1 to H.W.-C.), Joint Information Systems Committee for the Defining Image Access Project and the FlyWeb Project (grant numbers not used). Funding for open access charge: Departmental funds.

Conflict of interest statement. None declared.

REFERENCES

- Benson,E., Gudmannsdottir,E., Klyne,G., Shotton,D. and White-Cooper,H. (2007) FlyTED—The *Drosophila* Testis Gene Expression Database. *Proceedings of the Twentieth European Drosophila Research Conference*, Vienna, Austria **20**: T1008. Abstract available at <http://flybase.org/reports/FBbrf0199979.html>. Abstract and presentation (pp 1–23) available at http://imageweb.zoo.ox.ac.uk/pub/2007/Benson_et_al_FlyTED_EDRC20_Vienna_2007.pdf (19 November 2009, date last accessed).
- Barreau,C., Benson,E., Gudmannsdottir,E., Newton,F. and White-Cooper,H. (2008) Post-meiotic transcription in *Drosophila* testes. *Development*, **135**, 1897–1902.
- Barreau,C., Benson,E. and White-Cooper,H. (2008) Comet and cup genes in *Drosophila* spermatogenesis: the first demonstration of post-meiotic transcription. *Biochem. Soc. Trans.*, **36**, 540–542.
- Tomancak,P., Berman,B., Beaton,A., Weiszmam,R., Kwan,E., Hartenstein,V., Celniker,S. and Rubin,G. (2007) Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.*, **8**, R145.
- Lécuyer,E., Yoshida,H., Parthasarathy,N., Alm,C., Babak,T., Cerovina,T., Hughes,T.R., Tomancak,P. and Krause,H.M. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*, **131**, 174–187.
- Weigmann,K., Klapper,R., Strasser,T., Rickert,C., Technau,G., Jäckle,H., Janning,W. and Klämbt,C. (2003) FlyMove—a new way to look at development of *Drosophila*. *Trends Genet.*, **19**, 310–311.
- Poustelnikova,E., Pisarev,A., Blagov,M., Samsonova,M. and Reinitz,J. (2004) A database for management of gene expression data *in situ*. *Bioinformatics*, **20**, 2212–2221.
- Kelso,R.J., Buszczak,M., Quinones,A.T., Castiblanco,C., Mazzalupo,S. and Cooley,L. (2004) Flytrap, a database documenting a GFP protein-trap insertion screen in *Drosophila melanogaster*. *Nucleic Acids Res.*, **32**, D418–D420.
- 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.
- Haudry,Y., Berube,H., Letunic,I., Weeber,P.D., Gagneur,J., Girardot,C., Kapushesky,M., Arendt,D., Bork,P., Brazma,A. *et al.* (2007) 4DXpress: a database for cross-species expression pattern comparisons. *Nucleic Acids Res.*, **36**, D847–D853.
- 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.
- Zhao,J., Miles,A., Klyne,G. and Shotton,D. (2009) OpenFlyData: the way to go for biological data integration. *Data Integration in the Life Sciences. Lecture Notes in Computer Science*, **5647**, 47–54.
- Parisi,M., Nuttall,R., Edwards,P., Minor,J., Naiman,D., Lü,J., Doctolero,M., Vainer,M., Chan,C., Malley,J. *et al.* (2004) A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol.*, **5**, R40.
- Andrews,J., Bouffard,G.G., Cheadle,C., Lu,J., Becker,K.G. and Oliver,B. (2000) Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res.*, **10**, 2030–2043.
- Dorus,S., Busby,S.A., Gerike,U., Shabanowitz,J., Hunt,D.F. and Karr,T.L. (2006) Genomic and functional evolution of the

- Drosophila melanogaster* sperm proteome. *Nat. Genet.*, **38**, 1440–1445.
16. Schäfer, M., Nayernia, K., Engel, W. and Schäfer, U. (1995) Translational control in spermatogenesis. *Dev. Biol.*, **172**, 344–352.
 17. Lin, T., Viswanathan, S., Wood, C., Wilson, P., Wolf, N. and Fuller, M. (1996) Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development*, **122**, 1331–1341.
 18. Fuller, M.T. (1998) Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis. *Semin. Cell Dev. Biol.*, **9**, 433–444.
 19. Ayyar, S., Jiang, J., Collu, A., White-Cooper, H. and White, R.A.H. (2003) *Drosophila* TGIF is essential for developmentally regulated transcription in spermatogenesis. *Development*, **130**, 2841–2852.
 20. Jiang, J., Benson, E., Bausek, N., Doggett, K. and White-Cooper, H. (2007) Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the *Drosophila* male germline and physically interacts with Always early. *Development*, **134**, 1549–1559.
 21. Jiang, J. and White-Cooper, H. (2003) Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of *aly* and a novel meiotic arrest gene *cookie monster*. *Development*, **130**, 563–573.
 22. Perezgasga, L., Jiang, J.Q., Bolival, B., Hiller, M., Benson, E., Fuller, M.T. and White-Cooper, H. (2004) Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein matopetli. *Development*, **131**, 1691–1702.
 23. Wang, Z. and Mann, R.S. (2003) Requirement for two nearly identical TGIF-related homeobox genes in *Drosophila* spermatogenesis. *Development*, **130**, 2853–2865.
 24. White-Cooper, H., Leroy, D., MacQueen, A. and Fuller, M. (2000) Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development*, **127**, 5463–5473.
 25. Hiller, M., Chen, X., Pringle, M., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T., Marino, S. and Fuller, M. (2004) Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development*, **131**, 5297–5308.
 26. Hiller, M., Lin, T., Wood, C. and Fuller, M. (2001) Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.*, **15**, 1021–1030.
 27. White-Cooper, H., Schafer, M., Alphey, L. and Fuller, M. (1998) Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. *Development*, **125**, 125–134.
 28. White-Cooper, H. (2004) Spermatogenesis: analysis of meiosis and morphogenesis. *Methods Mol. Biol.*, **247**, 45–75.
 29. Deutsch, E.W., Ball, C.A., Bova, G.S., Brazma, A., Bumgarner, R.E., Campbell, D., Causton, H.C., Christiansen, J., Davidson, D., Eichner, L.J. *et al.* (2006) Development of the minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE). *OMICS*, **10**, 205–208.
 30. Kiger, A.A., White-Cooper, H. and Fuller, M.T. (2000) Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature*, **407**, 750–754.
 31. Zhao, J., Klyne, G. and Shotton, D. (2008) Building a Semantic Web image repository for biological research images. *Proceedings of the European Semantic Web Conference. The Semantic Web: Research and Applications. Lecture Notes in Computer Science*, **5021**, 154–169.
 32. Stein, L.D. (2003) Integrating biological databases. *Nat. Rev. Genet.*, **4**, 337–345.
 33. Zhao, J., Klyne, G. and Shotton, D. (2008) Provenance and linked data in biological data webs. *Proceedings of the Linked Open Data Workshop at the Seventeenth International World Wide Web Conference*. Beijing, China, 2008, pp. 1–5. Available at <http://events.linkedata.org/ldow2008/papers/07-zhao-klyne-provenance-linked-data.pdf> (19 November 2009, date last accessed).
 34. Zhao, J., Miles, A., Klyne, G. and Shotton, D. (2008) Linked data and provenance in biological data webs. *Brief. Bioinformatics*, **10**, 139–152.
 35. Carroll, J.J., Bizer, C., Hayes, P. and Stickler, P. (2005) Named graphs. *J. Web Semant.*, **3**, 247–267.