Resources for Functional Genomics Studies in *Drosophila melanogaster*

Stephanie E. Mohr,*,†,1 Yanhui Hu,*,† Kevin Kim,† Benjamin E. Housden,† and Norbert Perrimon*,†,‡
**Drosophila* RNAi Screening Center and †Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, and
‡Howard Hughes Medical Institute, Boston, Massachusetts 02115

ABSTRACT *Drosophila melanogaster* has become a system of choice for functional genomic studies. Many resources, including online databases and software tools, are now available to support design or identification of relevant fly stocks and reagents or analysis and mining of existing functional genomic, transcriptomic, proteomic, *etc*. datasets. These include large community collections of fly stocks and plasmid clones, "meta" information sites like FlyBase and FlyMine, and an increasing number of more specialized reagents, databases, and online tools. Here, we introduce key resources useful to plan large-scale functional genomics studies in *Drosophila* and to analyze, integrate, and mine the results of those studies in ways that facilitate identification of highest-confidence results and generation of new hypotheses. We also discuss ways in which existing resources can be used and might be improved and suggest a few areas of future development that would further support large- and small-scale studies in *Drosophila* and facilitate use of *Drosophila* information by the research community more generally.

THE availability of powerful genetic approaches and a well-annotated genome sequence have made *Drosophila melanogaster* a model system of choice for a wide range of functional genomic studies. Researchers are increasingly dependent upon online resources for navigation and management of the rapidly growing datasets associated with large-scale "omics" and other *Drosophila* studies. As a veritable encyclopedia of fly gene information, the online FlyBase resource is an essential online tool, providing access to curated information and resources for individual genes in an accessible format (McQuilton *et al.* 2012; St Pierre *et al.* 2014). However, throughout the course of many studies, such as large-scale screens or proteomics analyses, researchers need to turn to many other online resources for search, analysis, integration, and visualization of new or existing data. As a result, the use of a variety of software tools and online resources has become integral to *Drosophila* functional genomics studies, contributing to everything from experimental design to interpretation of results. Here, we present an overview of freely available noncommercial resources for functional genomics approaches in *Drosophila*; provide advice on how the associated online tools can be used by biologists and bioinformaticians; explore some of the roadblocks that prevent easy navigation; and suggest what changes or new resources might be of highest priority in the future. Although our focus will be on tools relevant to *D. melanogaster*, a number of the resources we discuss are relevant to other *Drosophila* and noninsect species.

Most online informatics tasks relevant to functional genomics can be divided into three basic categories. First, online resources can be used to identify or design reagents useful for a particular experimental approach, including fly stocks (Table 1), cell and molecular reagents (Table 2), and genome engineering resources (Table 3). Second, online resources that support search and view of existing information (Table 4) and enrichment approaches (Table 5) can be used to narrow down a list of candidate genes, such as to compare experimental *vs*. predicted results or to home in on the highest confidence results following a large-scale study. Third, a similar comparison of gene or protein lists to existing data in the curated literature or from 'omics datasets can be used to expand a list of candidate genes, such as to include all genes or proteins for which existing data suggest coexpression, binary interaction, cocomplex or colocalization, all genes encoding proteins that share a common domain or biochemical function, or all genes that share a common phenotype or predicted function (Table 4 and Table 5). Additional

doi: 10.1534/genetics.113.154344
Manuscript received January 9, 2014; accepted for publication February 9, 2014
[1]Corresponding author: 77 Avenue Louis Pasteur, Department of Genetics NRB 336, Boston, MA 02115. E-mail: stephanie_mohr@hms.harvard.edu

more specialized tasks include identification of related genes in other species, including fly genes related to human genes that have been implicated in disease (Table 5) and identification of predicted gene targets of miRNAs (Table 6).

"Meta" Sites for Fly Gene Annotation Information

A few key resources include information from many sources, making these metasites generally useful for many tasks. FlyBase (flybase.org) is a comprehensive database of fly gene information (McQuilton *et al.* 2012; St Pierre *et al.* 2014). FlyBase is the go-to source for information about specific genes, including gene annotation, mutant alleles and their phenotypes, high-throughput expression data, and publications. Curation of information by experts helps to ensure quality. Methods for querying data in FlyBase have recently been presented elsewhere (Wilson *et al.* 2008; McQuilton *et al.* 2012; St Pierre *et al.* 2014). The National Center for Biotechnology Information (NCBI) gene database (ncbi.nlm.nih.gov/gene) (2013) and Ensemble (ensembl.org/Drosophila_melanogaster/Info/Index) (Flicek *et al.* 2013) from the EMBL and Wellcome Trust Sanger Institute are also major repositories of gene annotation information, including for *Drosophila*. It should be noted, however, that gene annotations at these sites are not independent; they are imported from FlyBase. Users might see differences between NCBI or Ensemble or FlyBase when the former sites are out of synch with the current FlyBase release. The metasite FlyMine (flymine.org) provides access to overlapping sets of information as compared with these other resources (Lyne *et al.* 2007). FlyMine has the added advantage of user-friendly interfaces for handling and viewing information regarding long lists of genes, but particularly in terms of reagent information, it is less comprehensive than FlyBase. In addition, >1000 genome-scale datasets from the modENCODE project are available for download online (see data.modencode.org/?Organism=D.%20melanogaster). Gene annotations will continue to change over time as these and other datasets are further analyzed and incorporated at FlyBase.

Identification of Existing Reagents

Centralized fly stock and reagent collections

A number of facilities in the United States and overseas have collected resources from large numbers of laboratories, creating centralized distribution centers. These include fly stock collections at the Bloomington *Drosophila* Stock Center (BDSC; classical alleles, mobile element insertions, Gal4 and upstream activation sequence (UAS) stocks, RNAi fly stocks, *etc*.) (Cook *et al.* 2010a), Vienna *Drosophila* RNAi Center (VDRC; RNAi fly stocks, Vienna Tiles Gal4 stocks, *etc*.), *Drosophila* Genetic Resource Center-Kyoto Stock Center [Kyoto PiggyBac, FlyTrap, and Cambridge protein trap insertion (CPTI) fluorescent trap stocks, Gal4 stocks, *etc*.] (Yamamoto 2010) and National Institute of Genetics-Japan (NIG-Japan; Cas9 fly stocks, RNAi fly stocks, *etc*.) (for URLs see Table 1). Centralized centers also include the San Diego *Drosophila* Species Stock Center and EHIME-Fly at EHIME University, which maintain collections of non-*melanogaster Drosophila* fly stocks (for URLs see Table 1), and the *Drosophila* Genomic Resource Center (DGRC), which distributes *D. melanogaster* resources such as cultured cells, plasmid vectors, cDNAs, *etc*. (Table 2). In addition, several large-scale and consortium efforts have led to the development of large-scale collections made accessible through their own websites and/or at existing distribution centers (Table 1 and Table 2).

Major categories of in vivo fly stocks

Fly stocks useful for functional genomic studies (Table 1) can be generally grouped into (1) those in which the endogenous gene is mutated, disrupted, or tagged directly, *e.g.*, classical alleles, deletions, Gene Disruption Project (GDP) transposon insertions, and "protein trap" lines (Morin *et al.* 2001; Bellen *et al.* 2004, 2011; Kelso *et al.* 2004; Buszczak *et al.* 2007; Cook *et al.* 2010a, 2012; Venken *et al.* 2011; Singari *et al.* 2014), and (2) those for which transgenesis into a neutral or unrelated location is used to introduce an exogenous reporter construct, expression driver, reagent, *etc*. (reviewed in Cook *et al.* 2010a; Perrimon *et al.* 2010). Some fly stocks, such as GDP MiMIC element insertions, fall into both categories, as they can be useful as mutant alleles and/or used to develop reporter constructs, protein fusions, *etc*. (Venken *et al.* 2011). GDP and other efforts have been focused on complete coverage at the chromosome region or gene level, such as nearly complete overlapping coverage by deletion constructs (Cook *et al.* 2010b, 2012) and generation of at least one disruption (GDP) or RNAi construct [VDRC, Transgenic RNAi Project (TRiP), NIG-Japan] per gene (reviewed in Perrimon *et al.* 2010). It is conceivable that future development will include targeted disruption of specific gene isoforms. The FlyFos collection fly stocks (transgeneome.mpi-cbg.de) facilitate RNAi "rescue" approaches using *D. pseudoobscura* genomic fragments (see below and Langer *et al.* 2010). Overexpression of open reading frames (ORFs) using the Gal4-UAS system is supported by fly stock collections at the University of Zurich in Switzerland (FlyORF; Bischof *et al.* 2013) and the National Centre for Biological Sciences in India [*Drosophila* protein interaction map (DPiM) fly stocks; Guruharsha *et al.* 2011]. Recent areas of growth include the large number of newly generated Gal4 fly stocks, *e.g.*, the Janelia Farm (Pfeiffer *et al.* 2008) and Vienna Tiles (Jenett *et al.* 2012) collections, and the *Drosophila* Genetic Reference Panel (DGRP) for population genomics and analysis of quantitative traits (Mackay *et al.* 2012). Fly stocks for interrogation of miRNA function have also been an area of recent development; fly stocks, reagents, and online tools related to miRNAs are presented in a later section.

Other physical resources and reagents

Resources useful for functional genomics approaches include fly stocks, cultured cell lines, cell-based RNAi libraries, and cDNA or ORF clones, as well as online tools useful for the design of RNAi reagents, quantitative real-time PCR (qPCR)

Table 1 Resources for *Drosophila in vivo* functional genomics studies (fly stocks)

Reagent type	Source	Recommended online entry point(s)	References
Classical mutant alleles	Community at large	http://flybase.org/ http://flystocks.bio.indiana.edu/ http://www.dgrc.kit.ac.jp/	Cook *et al.* (2010a); McQuilton *et al.* (2012); St Pierre *et al.* (2014); Yamamoto (2010)
Transposon insertions (of various types and uses)	Community at large	http://flybase.org/ http://flystocks.bio.indiana.edu/ http://www.dgrc.kit.ac.jp/	Cook *et al.* (2010a) McQuilton *et al.* (2012); St Pierre *et al.* (2014); Yamamoto (2010)
Transposon insertions (gene disruption collection)	*Drosophila* Gene Disruption Project	http://flypush.imgen.bcm.tmc.edu/pscreen/	Bellen *et al.* (2004, 2011); Venken *et al.* (2011)
RNAi	Vienna *Drosophila* RNAi Center	http://www.flyrnai.org/up-torr/ http://stockcenter.vdrc.at/control/main	Dietzl *et al.* 2007; Hu *et al.* 2013a)
RNAi	Transgenic RNAi Project (TRiP)	http://www.flyrnai.org/up-torr/ http://www.flyrnai.org/trip http://www.flyrnai.org/rsvp http://flystocks.bio.indiana.edu/	Hu *et al.* (2013a); Ni *et al.* (2011) Cook *et al.* (2010a)
RNAi	NIG, Japan	http://www.flyrnai.org/up-torr/ www.shigen.nig.ac.jp/fly/nigfly	Hu *et al.* (2013a); Yamamoto (2010)
RNAi	Community at large	http://flybase.org/ http://flystocks.bio.indiana.edu/	Cook *et al.* (2010a); McQuilton *et al.* (2012); St Pierre *et al.* (2014)
RNAi rescue with *D. pseudoobscura* genome fragments	FlyFos project	https://transgeneome.mpi-cbg.de	Langer *et al.* (2010)
Overexpression (UAS)	FlyORF resource	http://flyorf.ch/	Bischof *et al.* (2013)
Overexpression (UAS)	DPiM fly stocks	https://interfly.med.harvard.edu/transgenic_info.php	Guruharsha *et al.* (2011)
Gal4 expression	Community at large	http://flystocks.bio.indiana.edu/Browse/gal4/gal4_main.htm http://kyotofly.kit.jp/cgi-bin/stocks/data_search.cgi#GAL4	Cook *et al.* (2010a); Yamamoto (2010)
Gal4 expression	Janelia Farms FlyLight project	http://flystocks.bio.indiana.edu/Browse/gal4/gal4_Janelia.php	Cook *et al.* (2010a)
Gal4 expression	Vienna Tiles	http://stockcenter.vdrc.at/control/vtlibrary	Jenett *et al.* (2012)
Fluorescence protein trap	Cooley and Chia laboratories	http://flytrap.med.yale.edu/ http://kyotofly.kit.jp/cgi-bin/stocks/data_search.cgi#FlyTrap	Kelso *et al.* (2004); Morin *et al.* (2001); Yamamoto (2010)
Fluorescence protein trap	Cambridge Protein Trap Insertion	http://kyotofly.kit.jp/stocks/documents/CPTI.html	Yamamoto (2010)
Fluorescence protein trap	Carnegie Protein Trap Library	http://flytrap.med.yale.edu/	Buszczak *et al.* (2007)
Fluorescence protein trap for dominant phenotypes	Edwards laboratory collection		Singari *et al.* (2014)
MiMIC insertions (for fluorescence trap, *etc.*)	*Drosophila* Gene Disruption Project	http://flypush.imgen.bcm.tmc.edu/pscreen/mimic.html	Venken *et al.* (2011)
Human disease relevant (curated)	Community at large	http://flystocks.bio.indiana.edu/Browse/HD/HDintro.htm	Cook *et al.* (2010a)
Human disease-relevant fly stocks (RNAi)	TRiP	http://www.flyrnai.org/hudis-trip http://www.flyrnai.org/diopt-dist (after a search, click the button "find RNAi reagents" to carry the list to UP-TORR)	Hu *et al.* (2011); Ni *et al.* (2011)
CRISPR/Cas9 system	Kondo and Ueda	http://www.shigen.nig.ac.jp/fly/nigfly/cas9/	Kondo and Ueda (2013)
Population genomics and analysis of quantitative traits	*Drosophila* Genetic Reference Panel	http://mackay.gnets.ncsu.edu/MackaySite/DGRP.html http://dgrp.gnets.ncsu.edu/php	Mackay *et al.* (2012)
Non-*melanogaster* species	Community at large	http://kyotofly.kit.jp/cgi-bin/ehime/index.cgi	

Table 2 Resources for *Drosophila* cell-based and molecular approaches (cell lines, plasmid clones, design tools, *etc.*)

Reagent type	Facility or lab group	Recommended online entry point(s)	References (PMID)
Cultured cell lines	Community at large	https://dgrc.cgb.indiana.edu	Cherbas and Gong (2014)
Cell-based RNAi libraries and reagents	DRSC	http://www.flyrnai.org/up-torr/ http://www.flyrnai.org/	Flockhart *et al.* (2012); Hu *et al.* (2013a)
Cell-based RNAi reagent design	DRSC	www.flyrnai.org/snapdragon	Flockhart *et al.* (2012)
Cell-based RNAi reagent design	DKFZ/Boutros lab	www.dkfz.de/signaling/e-rnai3 http://b110-wiki.dkfz.de/signaling/wiki/display/nextrnai/NEXT-RNAi	Arziman *et al.* (2005); Horn *et al.* (2010)
cDNAs and ORFs	BDGP	https://dgrc.cgb.indiana.edu	Stapleton *et al.* (2002)
pUAS-ORFs	BDGP and DRSC	http://plasmid.med.harvard.edu	Zuo *et al.* (2007)
Empty vectors, additional cDNAs or ORFs, *etc.*	Community at large	https://dgrc.cgb.indiana.edu http://www.addgene.org/ http://plasmid.med.harvard.edu https://dnasu.org/DNASU/Home.do	Herscovitch *et al.* (2012); Stapleton *et al.* (2002) Seiler *et al.* (2014); Zuo *et al.* (2007)
Plasmids for making CRISPR genome engineering constructs	AddGene	http://www.addgene.org/CRISPR/	Herscovitch *et al.* (2012)
Plasmids for making TALEN genome engineering constructs	AddGene	http://www.addgene.org/talengineering/	Herscovitch *et al.* (2012)
Primer designs for qPCR	DRSC	http://www.flyrnai.org/flyprimerbank	Hu *et al.* (2013b)
shRNA/siRNA design	DSIR	http://biodev.extra.cea.fr/DSIR/	Vert *et al.* (2006)

DSIR, Designer of Small Interfering RNA

primers, or clustered regularly spaced short palindromic repeat (CRISPR) short guide (sg) RNAs (Table 2 and Table 3). The study of protein function in particular is reliant on the availability of appropriate resources, namely, high-quality plasmid clones that contain the ORFs in a format compatible with protein expression. The "Gold Collection" built by Berkeley *Drosophila* Genome Project (BDGP) using cDNA libraries from a variety of tissues and developmental stages (Stapleton *et al.* 2002) comprises ~11,000 sequence-verified full-length cDNA clones representing ~10,300 genes (74% of predicted *D. melanogaster* genes). To facilitate high-throughput subcloning, as well as ensure maximum flexibility, such as different tagging options and expression levels, it is important to use recombinational cloning vectors and include only the coding sequences, with all untranslated sequences removed. Successful large-scale use of the BDGP ORF collections was demonstrated in the DPiM project (Guruharsha *et al.* 2011). In this large-scale study, 4273 "bait" ORFs were cloned into a vector that fuses the ORF to a carboxyl-terminal FLAG-HA tag cassette under the control of a metallothionein promoter. Each clone was transiently transfected into S2R+ cells. The associated protein complexes were then affinity purified and detected by mass spectrometry analysis, which led to the generation of a map encompassing 556 protein complexes.

cDNA, ORF, and genomic clones for functional genomics

Gold collection clones are available for search or order from the *Drosophila* Genomics Resource Center (Table 2). Currently most recombinational ORFs are in CloneTech Creator system vectors, with some smaller collections in Invitrogen Gateway cloning system vectors. The UAS-ORF collections generated by the Basler group are in a barcoded vector, pUASg.attB, for stop codon-containing clones, and pUASg-HA.attB, for ORFs with HA tags at their C-terminal ends. Using these clones together with the site-specific PhiC31 integrase, the group created a collection of 1149 UAS-ORF transgenic fly strains for *in vivo* overexpression screens (Bischof *et al.* 2013). Relevant to RNAi studies is the availability of genomic clones from non-*melanogaster* species that can be used for RNAi rescue approaches, as these distant species' genes are often different enough at the nucleotide level to evade RNAi but similar enough at the regulation and protein sequence level so as to confer function. Available collections include a *D. persimilis* library distributed by the DRSC (flyrnai.org/cgi-bin/RNAi_find_rescue_compl.pl) useful for building constructs for testing in cells or *in vivo* (Kondo *et al.* 2009) and as mentioned above, the FlyFos *D. pseudoobscura* collection as described in Ejsmont *et al.* (2009) and Langer *et al.* (2010).

Online tools for reagent search and information

As shown in Table 1 and Table 2, most large-scale reagent collections are supported by one or more online site for search, view, and request of reagents. For most resources, it is easiest (and in some cases, only possible) to search for information related to one gene at a time. Moreover, the sites often support search of only one type or collection of reagents. For large-scale studies, however, it can often be easier to search in batch mode, as well as to search all available reagents of a type rather than only a specific collection. Another practical option for functional genomics studies, where long lists of genes are being managed, is to have the ability to download results in tab-delimited or similar file formats that are compatible with Excel and other spreadsheet or database software programs. Although FlyBase provides access to information about many different reagent collections, it is not easy for the novice user to search with more than one gene at a time or limit searches to a particular type of resource (for help with FlyBase searches see McQuilton *et al.* 2012; St Pierre *et al.* 2014). Conversely, although the

Table 3 Resources for genome engineering (TAL and CRISPR approaches)

Reagent type	Facility or lab group	Recommended online entry point(s)	References (PMID)
sgRNA designs for CRISPR genome engineering	DRSC	http://www.flyrnai.org/crispr/	Ren *et al.* (2013)
	O'Connor-Giles lab	http://tools.flycrispr.molbio.wisc.edu/targetFinder/	Sander *et al.* (2010)
	DKFZ/Boutros lab	http://www.e-crisp.org/E-CRISP/designcrispr.html	Heigwer *et al.* (2014)
	NIG-FLY/Ueda lab	www.shigen.nig.ac.jp/fly/nigfly/cas9/index.jsp	Kondo and Ueda (2013)
	Center for Bioinformatics, PKU	http://cas9.cbi.pku.edu.cn/	Ma *et al.* (2013)
TALEN design	DKFZ/Boutros lab	http://www.e-talen.org/E-TALEN/	Heigwer *et al.* (2013)
	Cornell University	https://tale-nt.cac.cornell.edu/	Doyle *et al.* (2012); Cermak *et al.* (2011)
	Ekker lab	www.talendesign.org	Neff *et al.* (2013); Sander *et al.* (2010, 2007)
	Zinc Finger Consortium	http://zifit.partners.org/ZiFiT/	
TALE and CRISPR information	Joung lab	http://talengineering.org/	
	Zhang lab	www.genome-engineering.org/taleffectors/	
	Zhang lab	www.genome-engineering.org/crispr	
	Liu lab	http://groups.mrcfgu.ox.ac.uk/liu-group/useful-links/oxfcrispr/oxfcrispr	
	Bullock lab	http://flycrispr.molbio.wisc.edu/tools	

See Table 2 for plasmid constructs related to genome engineering.

metasite FlyMine (Lyne *et al.* 2007) makes it relatively easy to create and view information pertaining to long lists of genes, FlyMine appears to capture only limited information about available fly stocks and other reagents.

Identification of RNAi reagents

To facilitate the identification of cell-based and *in vivo* RNAi reagents, we recently developed UP-TORR (updated targets of RNAi reagents). UP-TORR facilitates search with more than one gene identifier at a time, allows for simultaneous search of major public fly RNAi reagent collections [*Drosophila* RNAi Screening Center (DRSC), German Cancer Research Center (DKFZ), TRiP, VDRC, and NIG-Japan], and gives the option to download results as an Excel-compatible file. UP-TORR also addresses another common problem with reagent and other online search tools, that is, the need to keep up to date with new gene annotation releases, including changes to gene identifiers, gene names, and synonyms. Keeping up with new gene annotations is particularly important for RNAi reagents, for which the predicted targets and off targets can change significantly over time, but the issue is generally relevant as well. Indeed, over time, many sites get out of synch with updates to FlyBase identifiers (FBgns), gene symbols, *etc*., such that searches can produce false negative and/or false positive results. Ideally, all sites would build automated detection and import of updates by FlyBase or other relevant information sources into their online tools, such that those tools automatically keep up to date. However, this is not a trivial task and still requires maintenance over time, as the information sources are likely to add to or change the format of data releases, such that the automated approach would have to be revised. One practical approach researchers can use to address the problem that different sites are built around different FlyBase releases is to process input and output lists through a user account at FlyMine. FlyMine detects "retired" gene symbols or identifiers in a list and allows the user to update a list based on suggested new identifiers, both at the time a list is created and when a registered user signs in and views a list already saved to his or her account. Another useful tool is FlyBase's Upload/Convert IDs tool (flybase.org/static_pages/downloads/IDConv.html).

Recent trends among online tools for reagent search and information

Two recent trends in the organization and curation of *Drosophila* reagents are notable. First, groups such as the BDSC and TRiP are making efforts to highlight fly stocks that may be particularly relevant to the study of human disease (for URLs see Table 1). A recent review by Chen and Crowther (2012) highlights online tools useful for disease-related functional genomics. Second, an increasing number of online resources are soliciting direct contribution of information by the community. FlyBase has a long history of accepting personal communications and recently, FlyBase and other sites have begun to provide additional opportunities for direct contributions as a supplement to curated information. For example, FlyBase has links to FlyGene Wiki (see "User Contributed Data" under "Summaries" at a FlyBase gene page). Specific to reagent resources, we recently created online sites that allow users to input information about reagent quality, namely, the RNAi Stock Validation and Phenotypes (RSVP) searchable online database of quality information (qPCR and phenotypes) for RNAi fly stocks (flyrnai.org/rsvp), and FlyPrimerBank, which provides sequences and information regarding primer pairs for qPCR analysis (flyrnai.org/flyprimerbank) (Hu *et al.* 2013b). At the moment, much of the reagent information in RSVP and FlyPrimerBank is from our own internal studies or curated from the literature. It remains to be seen if direct contribution by individual labs or researchers will become a significant source of information at these sites. One reason for optimism regarding direct contribution of information about reagents in particular is the fact that providing negative information (*e.g.*, an experimentally validated

Table 4 Resources for search and view of *Drosophila* 'omics datasets

Data type	Data source	Recommended online entry point(s)	References (PMID)
Phenotypes (mutant alleles and RNAi), RNA expression, *etc.*	Community at large	http://flybase.org/ http://www.flymine.org/	Lyne *et al.* (2007); McQuilton *et al.* (2012); St Pierre *et al.* (2014)
RNA expression (various tissues)	Dow lab	http://flyatlas.gla.ac.uk http://flyatlas.org/ http://flybase.org/	Chintapalli *et al.* (2007); McQuilton *et al.* (2012); Robinson *et al.* (2013); St Pierre *et al.* (2014)
RNA expression (various stages, tissues, and treatments)	modENCODE consortium	http://flybase.org/ http://intermine.modencode.org/	Contrino *et al.* (2012); McQuilton *et al.* (2012); St Pierre *et al.* (2014)
RNA expression (in cultured cells)	modENCODE consortium	www.flyrnai.org/cellexpress	Cherbas *et al.* (2011); Flockhart *et al.* (2012)
		https://dgrc.cgb.indiana.edu/cells/TilingSearch http://intermine.modencode.org/	Contrino *et al.* (2012)
RNA localization	FlyFish FlyExpress	http://fly-fish.ccbr.utoronto.ca/ http://www.flyexpress.net/	Kumar *et al.* (2011); Lecuyer *et al.* (2007)
In vivo RNAi data (screen results)	Community at large	http://www.genomernai.org/	Horn *et al.* (2007)
In vivo RNAi data (qPCR and phenotype data)	Transgenic RNAi Project (TRiP) and community at large	www.flyrnai.org/rsvp	Flockhart *et al.* (2012); Ni *et al.* (2011)
Cell-based RNAi data	Screen data from DRSC, DKFZ and other sources	www.flyrnai.org/genelookup www.flyrnai.org/screensummary www.genomernai.org	Flockhart *et al.* (2012); Horn *et al.* (2007)
Cell-based RNAi data (in protein network context)	Screen data from DRSC, DKFZ and other sources	www.flyrnai.org/SignedPPI/	Vinayagam *et al.* (2013b)
Protein interaction data (large-scale cell-based mass spec study)	DPiM Project	http://droidb.org/ https://interfly.med.harvard.edu	Guruharsha *et al.* (2011); Yu *et al.* (2008)
Protein interaction data (large-scale binary/mass spec studies)	Various large-scale and community efforts	http://droidb.org/	Yu *et al.* (2008)

lack of RNAi knockdown), which researchers might be more willing to part with prepublication, can initiate renomination of a gene for design and production of alternative reagents, providing a significant incentive for submission of the data. It will be interesting to watch over time if it proves both practical and worthwhile to move away from curation by others and toward direct contribution, either for specific types of information or more generally.

The Rapid Emergence of Resources for Genome Engineering

CRISPR system and resources

Recently emerging technologies for genome engineering such as CRISPR/Cas9 and transcription activator-like effector nucleases (TALENs) were quickly adapted to knockout or tag genes in *Drosophila*. For example, several groups have demonstrated the efficacy of the CRISPR/Cas9 method for *in vivo* modification of *Drosophila* (Bassett *et al.* 2013a; Gratz *et al.* 2013; Kondo and Ueda 2013; Ren *et al.* 2013; Sebo *et al.* 2013; Yu *et al.* 2013). These initial reports suggest that off-target (OT) issues are not much of a concern in *Drosophila* provided that the sgRNA sequence is selected carefully. We implemented a genome-wide CRISPR resource annotating all possible CRISPR designs (flyrnai.org/crispr) (Ren *et al.* 2013). This resource visually displays all possible CRISPR designs using a genome browser for the gene or the genome location of interest specified by the user. For each sgRNA design, we analyzed the on-target efficiency based on nucleotide composition as well as all potential OT sites and the genomic feature of OT locations, such as whether an OT site is located at an intergenic region, intron, exon, or coding sequence (CDS) region of other genes. We plan to regularly update the tool as new information about CRISPR design becomes available. Similar tools include flyCRISPR and e-CRISP (for URLs see Table 3), which also provide CRISPR sgRNA designs.

TALEN system and resources

In addition to CRISPR, TALENs have also been shown to be effective in generating *in vivo* modifications in *Drosophila* (Liu *et al.* 2012; Beumer *et al.* 2013; Katsuyama *et al.* 2013; Kondo *et al.* 2013). In mammalian systems, the efficiency of double-strand break generation appears to be better using CRISPR (Cong *et al.* 2013) but several groups have reported high rates of mutagenesis at OT sites (Cradick *et al.* 2013; Fu *et al.* 2013; Hsu *et al.* 2013; Mali *et al.* 2013; Pattanayak *et al.* 2013). In contrast, specificity does not appear to be a problem with the TALEN system due to the longer targeting sequence and use of heterodimeric nucleases (Mussolino *et al.* 2011; Cade *et al.* 2012; Dahlem *et al.* 2012; Ding *et al.* 2013; Hisano *et al.* 2013; Mali *et al.* 2013). In *Drosophila*, however, OT mutations associated with CRISPR appear to be less problematic with several groups reporting use of the system with no detectible OT effects (Bassett *et al.* 2013a,b; Gratz *et al.* 2013; Kondo and Ueda 2013; Ren *et al.* 2013), making the advantages of TALENs less clear. Furthermore,

Table 5 Resources for analysis, enrichment, ortholog identification, *etc.* of *Drosophila* gene lists

Data/analysis type	Tool	Recommended online entry point(s)	References (PMID)
GO term, publication, domain, pathway, *etc.* enrichment	Flymine	http://www.flymine.org/	Contrino *et al.* (2012); Lyne *et al.* (2007)
	modMine	http://intermine.modencode.org/	Dennis *et al.* (2003)
	DAVID	http://david.abcc.ncifcrf.gov/	
Pathway annotation	PANTHER	http://pantherdb.org/pathway/	Mi *et al.* (2013); Ogata *et al.* (1999)
	KEGG	http://www.genome.jp/kegg/	
	FlyReactome	http://fly.reactome.org/	Joshi-Tope *et al.* (2005)
	Reactome	http://www.reactome.org/	
Complex enrichment	COMPLEAT	www.flyrnai.org/compleat	Vinayagam *et al.* (2013a)
Phenotype data	PhenomicDB	http://phenomicdb.de/	Kahraman *et al.* (2005); McQuilton *et al.* (2012); St Pierre *et al.* (2014)
		http://flybase.org/	
Cocomplex data	DroID	http://droidb.org/	Yu *et al.* (2008)
Binary interaction data	DroID	http://droidb.org/	Yu *et al.* (2008)
Ortholog identification	DIOPT	www.flyrnai.org/diopt	Hu *et al.* (2011)
Disease-related ortholog identification	DIOPT-DIST	www.flyrnai.org/diopt-dist	Forslund *et al.* (2011); Hu *et al.* (2011)
	orthoDisease	http://orthodisease.sbc.su.se	Chien *et al.* (2002)
	Homophila	http://superfly.ucsd.edu/homophila/	
Genetic modifiers of neurodegenerative diseases	NeuroGem	http://chibi.ubc.ca/neurogem/	Na *et al.* (2013)

construction of TALENs is considerably more difficult, requiring specialized molecular biology techniques and production of separate fusion proteins to target each individual locus. Support for TALEN design and analysis is available at E-TALEN (Heigwer *et al.* 2013) (www.e-TALEN.org). This resource allows the user to simultaneously enter up to 50 gene identifiers or a specific sequence to be analyzed for possible TALEN target sites. These target sites are then assessed for possible OT effects and assigned a score indicating their likely functionality and specificity. Filters can also be applied to design TALENs for a specific purpose such as protein tagging or gene knockout. Alternatively, the user can enter the sequence of a specific TALEN pair and map this to the target genome. Further tools are also available at TALE-NT (Cermak *et al.* 2011; Doyle *et al.* 2012) (tale-nt.cac.cornell.edu). In this case, a suite of tools is available for the design of TALENs, TALE-transcription factors, or analysis of possible OTs in a range of model organism genomes or a sequence provided by the user. Finally, multiple other resources are available including databases of TALE-related information and forums for discussion (Table 3).

Next directions for genome engineering

A further application of CRISPRs and TALEs is the regulation of gene expression. As recently demonstrated, both TALEs and CRISPR can be used to target a transcription factor fusion to a gene of interest and alter its transcriptional state (Crocker and Stern 2013; Gao *et al.* 2013; Gilbert *et al.* 2013; Konermann *et al.* 2013; Maeder *et al.* 2013a,b; Mali *et al.* 2013; Perez-Pinera *et al.* 2013a,b). Although to date CRISPR-based transcription factors have not been reported for *Drosophila*, success of related approaches in mammalian cells suggests that in this case, TALEs appear to be more effective than CRISPR. Thus, at least at this relatively early stage in the technologies, it seems likely that CRISPRs will be the system of choice for genome editing, whereas TALEs may be more useful for regulation of gene expression.

Resources for Search and View of *Drosophila* 'Omics Datasets

Data search and view at metasites

Sites like FlyBase and FlyMine provide access to a wealth of data relevant to *Drosophila* from the published literature and large-scale studies. In addition, a number of other sites provide access to specific data collections or types of 'omics data (Table 4). These include large-scale surveys by the Dow lab of gene expression in various *Drosophila* tissues, the FlyAtlas project (Chintapalli *et al.* 2007), and surveys by the modENCODE consortium across various *Drosophila* cell types, life cycle stages, treatment conditions, *etc*. (Cherbas *et al.* 2011; Graveley *et al.* 2011). For an individual gene, FlyAtlas and modENCODE transcript expression data can be viewed at the corresponding FlyBase gene page (expand to view "High-Throughput Expression Data" in the "Expression Data" section). FlyAtlas data can also be navigated at FlyAtlas (Chintapalli *et al.* 2007) or FlyAtlas2 (Robinson *et al.* 2013) (Table 4), which among other things facilitates search for genes that share a common pattern of expression as compared with the input gene ("Profile" option). For large-scale studies, researchers might find it necessary to download the underlying datasets and perform their own analyses offline of the web tools, because currently, using the available portals, it is difficult to search more than one gene at a time and/or results cannot easily be downloaded. A different view of RNA is available through FlyFish (Lecuyer *et al.* 2007) (Table 4), which facilitates search and view of the subcellular localization patterns of mRNAs, as determined using high-throughput fluorescence *in situ* detection. FlyFISH and other expression data can also be viewed and navigated at FlyExpress (Kumar *et al.* 2011) (Table 4).

Table 6 Resources for prediction of *Drosophila* miRNA targets

Tool or algorithm	Recommended online entry point(s)	Reference(s)
miRBase	http://mirbase.org/	Kozomara and Griffiths-Jones (2011)
MinoTar	www.flyrnai.org/minotar	Schnall-Levin *et al.* (2010)
TargetScanFly	http://targetscan.org/fly_12/	Ruby *et al.* (2007b)
DIANA-microT-CDS	http://diana.imis.athena-innovation.gr/DianaTools	Paraskevopoulou *et al.* (2013); Reczko *et al.* (2012)
miRanda	www.microrna.org	Betel *et al.* (2010)
DIANA-TarBase	http://diana.imis.athena-innovation.gr/DianaTools	Vergoulis *et al.* (2012)
miRTarBase	http://mirtarbase.mbc.nctu.edu.tw/	Hsu *et al.* (2011)

Specific support of RNAi, proteomics, and other datasets

Several resources facilitate search and view of cell-based and *in vivo* data relevant to RNAi screens, with online entry points to various datasets available through websites such as the DRSC (Flockhart *et al.* 2012) and GenomeRNAi (Schmidt *et al.* 2013) (for URLs see Table 4). The RSVP search tool for search and view of *in vivo* RNAi data is somewhat unique in that it allows users not just to view text but when available, also to view example images of mutant phenotypes. Large-scale proteomics data are available through the DPiM (Guruharsha *et al.* 2011) project site but in our experience, are more easily navigated through the *Drosophila* Interactions Database (DroID) (Yu *et al.* 2008; Murali *et al.* 2011), where researchers can select to view only DPiM results or view them alongside other data, *e.g.*, from binary interaction studies. SignedPPI (Vinayagam *et al.* 2013b) is designed to predict activation or suppression roles for proteins that directly interact with one another, based on phenotype data from publicly available RNAi datasets and interaction data from a variety of sources. If two proteins interact with each other and always score in the same direction across multiple RNAi screens, the two proteins are predicted to be activators of one another in the SignedPPI network, whereas two proteins are predicted to be suppressors of one another if they always score in opposite directions in RNAi screens. The SignedPPI tool allows users to query the signed protein–protein network among proteins assigned signs, as well as view an extended network of interactors (Vinayagam *et al.* 2013b).

Resources for enrichment analysis

Mining existing data can be useful at multiple stages of a functional genomics project. As mentioned previously, existing data can be used to limit or expand a list. For example, at the beginning of a study, a researcher might choose to narrow down a list of initial candidates to the subset of genes expressed in the stage or tissue of interest. Conversely, the choice might be to expand a list of initial candidates to include coexpressed genes, genes encoding proteins for which 'omics or other data suggest a physical interaction, genes with common gene ontology (GO) terms, *etc*. Perhaps more importantly, after a large-scale study, researchers can use enrichment approaches to limit and expand an initial "hits" list in an effort to reduce the impact of false positive and/or false negative results. A quick scan for "favorite" or well-characterized genes predicted to be positive in the assay is usually impossible to resist. However, systematically looking for commonalities in a long list of positive results on a gene-by-gene basis can be impossible to do. In this case, researchers turn to computational approaches to find patterns or commonalities in gene lists based on existing information (Table 5), such as biological processes or pathways that are overrepresented among the scored genes. This is an effective strategy for reducing the impact of false positive and negative results inherent in high-throughput screen datasets. Moreover, the results of enrichment analysis can improve confidence in subsets of results by placing them in biological context and helping foster the development of new hypotheses for in-depth follow-up studies.

Enrichment tools for Drosophila

FlyMine, modMine, and DAVID can be used for enrichment analysis for biological pathways, gene ontology terms, protein domains, publications, and more. FlyMine summarizes data sources at their website (see "Data Sources" tab on the home page flymine.org). The modMine home page similarly lists the types of data that can be viewed (chromatin structure, copy number variation, *etc*.; see intermine.modencode.org/). For National Institutes of Health, NIH DAVID, data, gene identifiers, and approaches are discussed in their frequently asked questions (FAQs) and summarized at david.abcc.ncifcrf.gov/content.jsp?file=update.html. It is worth noting that the backend knowledge bases that these tools are built on are assembled from publicly available resources and are slightly different from one another. For example, DAVID provides both original GO terms and "slim" GO terms, whereas FlyMine and modMine only provide original GO terms. In addition to GO term analysis, DAVID also provides pathway annotation from PANTHER and KEGG, while modMine provides pathway annotation from Reactome and flyReactome.

Pathway and other specialized enrichment tools

More specific molecular pathway annotations can also be accessed directly through various online entry points (Table 5). FlyReactome (Croft *et al.* 2013) provides annotation of eight major signaling pathways, namely, the circadian clock, Hedgehog, Hippo, Imd, JAK/STAT, planar cell polarity, Toll, and Wingless pathways. KEGG (Ogata *et al.* 1999) also provides pathway annotation for fly genes, in this case with a focus on metabolism-related genes. Although GO and pathway annotations are useful, they can be either too specific or too broad. For example, annotations for pathways usually span from receptor

complexes that receive a signal to transcription factor complexes that constitute the signal readout. To help scientists identify functional modules with higher resolution, we developed the protein COMPLex Enrichment Analysis tool (COMPLEAT) (Vinayagam *et al.* 2013a). This tool annotates protein complexes using both the literature and predictions from protein–protein networks. COMPLEAT facilitates analysis and visualization based on a full dataset, without input of a predefined cutoff value(s), as well as comparison of complexes at different conditions/time points.

Importance of background in enrichment analysis

When performing enrichment analysis, it is important to select the proper background. For studies that began with a focused gene list, such as all kinases or another nonrandom subset of genes, this gene list should be used as the background, not the full genome. The default setting for most enrichment tools is to use the whole genome as the background. To specify the background, users typically need to upload the background list as a file. At FlyMine, for example, the background can be specified by uploading and saving the background gene list, uploading the list of genes scored in the experiments, and finally, after the initial enrichment is completed with default settings, clicking "change" at "background population" and selecting a background list from the dropdown menu. For information about using custom background lists with DAVID, we recommend starting with their FAQs answer on the topic (david.abcc.ncifcrf.gov/content.jsp?file=FAQs.html#22).

Similarity through shared phenotypes

For *Drosophila* studies it is also useful to create or analyze a gene list based on a specific phenotype, for example to identify genes that might be involved in the same pathway or network. Expert users can generate queries in FlyBase, which relies on an extensive controlled vocabulary, to retrieve genes with related phenotypes (for help with FlyBase queries see McQuilton *et al.* 2012; St Pierre *et al.* 2014). PhenomicDB (Kahraman *et al.* 2005; Groth *et al.* 2007) allows users to generate gene lists by phenotype and provides mapping (via HomoloGene) to related genes in other species, so that phenotypes in other species can be explored. The Phenotype-based search is an area that might benefit from further development in the future.

Ortholog Identification and Use of Orthologs to Build Predictions

Tools for finding human disease-related and other orthologs

"Homologs" can be defined as genes that share a common origin, with the subcategory "paralogs" used to refer to genes of common origin within a species (such as due to gene duplication events) and "orthologs" used to refer to genes of common origin in different species (see figures 1–25 in Alberts 2002). Mapping gene orthologs among species serves an important role in functional genomics by allowing researchers to develop hypotheses about gene function in one organism based on what is known in another. Many algorithms for predicting orthologs have been developed but no one algorithm can hold itself up as providing perfect sensitivity and specificity. Moreover, some web interfaces for these algorithms are easier than others to navigate, and many do not support search of more than one gene at a time. To address some of these concerns, we developed the DRSC Integrative Ortholog Prediction Tool (DIOPT) (Hu *et al.* 2011) (Table 5), which combines results from 10 different ortholog prediction tools among flies, human, mouse, zebrafish, *Xenopus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. To help support large-scale studies, DIOPT facilitates search of multiple genes and results can be downloaded in an Excel-compatible format. Moreover, in an effort to integrate our tools, we recently added a feature that makes it possible to click a button ("find RNAi reagents") on the output page that will "carry" results from DIOPT to UP-TORR. Another tool based on the DIOPT approach, DIOPT Diseases and Traits (DIOPT-DIST) (Hu *et al.* 2011) (Table 5), integrates search for human gene orthologs with information from National Center for Biotechnology Information, NCBI Online Mendelian Inheritance in Man (OMIM) (Hamosh *et al.* 2000; Baxevanis 2012) or genome-wide association studies (GWAS; see genome.gov/gwastudies) (Hindorff *et al.* 2009) that link human genes to specific diseases or traits. Previously developed disease-related ortholog databases include Homophila (Chien *et al.* 2002) and OrthoDisease (O'Brien *et al.* 2004; Forslund *et al.* 2011), which support disease genes annotated in OMIM. With any of these tools, researchers should keep aware of when they were last updated (see below). It is worth mentioning that NeuroGem (Na *et al.* 2013) provides the first comprehensive knowledge base of integrated information on genetic modifiers of nine different neurodegenerative diseases, as identified in *Drosophila*, *C. elegans*, or yeast. This type of resource might prove useful for scientists who study the biological mechanisms underlying specific diseases.

Challenges to ortholog prediction that impact users

For all ortholog tools, it is a significant challenge to keep up to date with current gene annotations, and for tools like DIOPT-DIST (Hu *et al.* 2011), OrthoDisease (O'Brien *et al.* 2004), and Homophila (Chien *et al.* 2002) (Table 5), there is the added challenge of keeping up with human disease-gene information, including new GWAS data, for which new deposits are frequent. We do not know of any tool that is set up to remain up to date with all relevant data. Updates to gene annotations and disease gene information could be automated to some extent, such as by using a strategy similar to the approach we used for auto-update of UP-TORR (Hu *et al.* 2013a). However, given the number of different data sources involved (*i.e.*, various genome annotation and disease-gene association sources), full automation of updates does not currently seem like an achievable goal. Further standardization and centralization could help this problem in the future. In the meantime, users of ortholog and

disease-gene ortholog search tools are cautioned to keep aware of versions and update schedules associated with ortholog search tools. In addition, for DIOPT, DIOPT-DIST, and individual ortholog search tools, search results might not be limited to true orthologs (*e.g.*, the results for a given gene might include an ortholog as well as one or more related paralogs in the output species). Many additional tools are useful for researchers specifically interested in phylogeny (see evolution.genetics.washington.edu/phylip/software.html).

Orthologs and metadata sites

Some tools use underlying ortholog mapping to build predictions, as in the case of interlogs, which are putative interactors assigned based on interactions identified for orthologs in other species. *Drosophila* interlogs are included as an information source in DroID (Murali *et al.* 2011). Thus, if a user also maps interactions based on orthologs, such as using DIOPT or InParanoid, interologs from DroID should not be counted as a unique piece of evidence, as the same underlying information (InParanoid predictions) is already included. It is also worth mentioning that typically, built-in ortholog mapping in resources with another main focus are usually based on results from a single or a few ortholog prediction tools. As mentioned above, DroID uses InParanoid to map orthologs. In addition, PhenomicDB uses HomoloGene, and modMine uses InParanoid and TreeFam. For more comprehensive and accurate mapping of orthologs among species, the most cautious and comprehensive approach would be to download data and/or annotations from the original species only and then use a multialgorithm search tool like DIOPT to find predicted orthologs.

Integration and Visualization of Large-Scale Datasets and Networks

Visualization of protein networks

We increasingly view genes, proteins, small RNAs, *etc*. not as acting alone but as acting as parts of interconnected pathways and networks. Following a phenotypic study, an increasingly common and useful next step is to look at the relationship of genes with shared phenotypes or interactions to known or predicted protein interactions (Table 5). DroID (Yu *et al.* 2008; Murali *et al.* 2011) allows researchers to retrieve interactors identified in one or more of several large-scale protein complex and binary interaction studies, as well as to build visualizations of these networks. SignedPPI (Vinayagam *et al.* 2013b) allows users to build a "signed" network, *i.e.*, a protein–protein interaction network with edges of different colors indicating the activator/suppressor nature of any given protein pair. This can be done for a given gene list, and then the user can double-click on a given protein to extend the network and view additional interactors. For visualization of various types of biological data, many developers rely on the open source platform Cytoscape (Shannon *et al.* 2003; Kohl *et al.* 2011). For example, Cytoscape makes it possible to accommodate many configurations of nodes and edges, and thus it is used for visualizations at COMPLETE and SignedPPI. Downloads, tutorials, publications, *etc*. related to Cytoscape are available from their home page (http://www.cytoscape.org/) but we note that significant training and expertise is required to utilize the platform.

Visualization of mRNA expression data

Visualizations are also useful in viewing mRNA expression data. FlyBase (McQuilton *et al.* 2012; St Pierre *et al.* 2014) and modMine (Contrino *et al.* 2012) provide visualizations of expression data from various sources, easily accessed in gene views (FlyBase) and also viewable in the context of a genome browser (FlyBase and modMine). As mentioned previously, FlyAtlas (Chintapalli *et al.* 2007) expression array data across various tissues can be searched or viewed at FlyAtlas (Chintapalli *et al.* 2007) or FlyAtlas2 (Robinson *et al.* 2013), as well as at FlyBase and FlyMine (Lyne *et al.* 2007) (Table 4). To help plan or analyze the results of cell-based studies, the DRSC provides tabular output with color coding regarding expression in commonly used cultured cell lines for single- or multigene searches (flyrnai.org/cellexpress) (Flockhart *et al.* 2012) and the DGRC provides single-gene search access to data from 25 cell lines (dgrc.cgb.indiana.edu/cells/TilingSearch). The modMine site (Contrino *et al.* 2012) also provides visualizations in the form of heat maps based on their extensive RNAseq expression datasets, for single- or multigene searches (follow links from their home page modmine.org to view an example or initiate a search).

Resources for Interrogation of miRNA Function

miRNA sequence and expression database

Since the discovery of the first microRNAs (miRNAs), evidence has revealed the importance of miRNAs in numerous processes (Dai and Ahmed 2011; Huntzinger and Izaurralde 2011; Wang and Peng 2011). Many online tools assist in miRNA identification and target prediction (Table 6). We will discuss some of the online resources that are currently available, with an emphasis on tools that are available for *Drosophila* and among the most frequently utilized. The miRBase tool (www.mirbase.org) has become the primary online reference for all miRNA-related research, providing indexes of known miRNAs for a given organism, experimentally validated sequences, genomic locations, and miRNA transcript expression in various tissues and cell lines (Kozomara and Griffiths-Jones 2011). Release 20 of miRBase provides 24,521 precursor miRNA sequences and 30,424 mature miRNA sequences from 206 species. In general, a single miRNA precursor duplex produces one to two functional mature miRNA strands, while a rare third functional strand has been reported from the terminal loops of some miRNA precursors (Okamura *et al.*

2013). The number of precursor miRNAs and mature sequences for *D. melanogaster* has reached 238 and 426, respectively. Each miRBase miRNA entry page provides predicted hairpin structures and sequences, mature miRNA sequences, links to related miRNA families and clusters, reads from deep sequencing experiments containing each miRNA, links to some of the most widely used miRNA target prediction algorithms (discussed below), and supporting references for each miRNA. Many miRBase miRNA entry pages now also include a community annotation section, which provides information about specific miRNA families and functions taken directly from the free online encyclopedia, Wikipedia. This encourages miRBase users and miRNA experts to contribute their knowledge in the form of Wikipedia edits and new pages.

miRNA target prediction algorithms

Now that genome-wide information regarding what miRNAs are encoded is available for *Drosophila*, identification of target genes for a given miRNA is arguably the most important information for moving miRNA studies forward. However, even for extensively studied organisms miRNA targets are largely unknown. Initial target prediction algorithms based on simple Watson–Crick base-pairing rules generated a high number of false positive results. Advancements in target prediction algorithms have been made in parallel with increasing knowledge about miRNAs. Nevertheless, no "perfect" set of rules has been identified, and the results provided by various updated prediction algorithms are still largely different, due to differences in the criteria they apply. These differences include stringency of base pairing, target site conservation, and target site accessibility. Below, we discuss three of the most frequently used online target prediction algorithms and the criteria each uses to generate target predictions, typically based on 3′-untranslated regions (UTRs) of genes (for URLs see Table 6). An additional tool, minoTar, facilitates identification of potential miRNA binding sites in coding sequences (flyrnai.org/minotar) (Schnall-Levin *et al.* 2010).

TargetScanFly (targetscan.org/fly_12/) (Ruby *et al.* 2007b) is one of the pioneering target prediction algorithms presenting high sensitivity and precision. TargetScanFly Release 6.2 displays predicted regulatory targets of 148 *D. melanogaster* miRNAs (Ruby *et al.* 2007b), including mirtronic miRNAs (Ruby *et al.* 2007a). Targets are predicted using the TargetScanS algorithm (Lewis *et al.* 2005) and conservation cutoffs calculated using branch length scores (Kheradpour *et al.* 2007). TargetScanS requires complementarity of at least 6 nucleotides in the 5′ end of the miRNA. TargetScanFly also offers users the option to search for target sites within ORFs (TargetScanFly ORFs) and to input a seed sequence (nucleotides 2–8) for a small RNA to search matching conserved sites (TargetScanFly Custom).

Another commonly used tool is DIANA-microT-CDS (microrna.gr/webServer) (Reczko *et al.* 2012; Paraskevopoulou *et al.* 2013). In its fifth version, the microT algorithm searches for target sites within both the 3′-UTR and coding regions of a gene. DIANA-microT-CDS relies first on thermodynamic stability between miRNAs and their potential target sites. The algorithm also takes into account features such as binding category, conservation among species, accessibility of target sites as calculated using the Sfold algorithm (Chan *et al.* 2005), and AU content of regions flanking the target sites.

The miRanda tool (microrna.org) identifies target sites using features such as sequence complementarity and binding energy between the mature miRNA and the target site (Betel *et al.* 2010). Predicted duplexes are given an miRNA support vector regression (mirSVR) score, which computes a weighted sum of context and sequence features such as AU content flanking the target site, target site accessibility and position, 3′-UTR length, base pairing, and a conservation score. Conservation score is calculated based on two features, evolutionary conservation of the target site sequence and site position in aligned 3′-UTRs of homologous sequences.

Experimentally validated targets of miRNAs

To date, miRNA target prediction algorithms have helped identify targets of miRNAs. These predictions were experimentally validated using various techniques, including reporter gene assays, evaluation of miRNA *vs.* mRNA expression using qPCR or Northern blotting, and assessment of the effects of miRNA expression on protein levels of a target gene using standard techniques such as immunoblotting. These results are now being curated into databases of experimentally validated miRNA targets. DIANA-TarBase (microrna.gr/tarbase) is currently the largest manually curated target database, with an index of >65,000 miRNA–target interactions from many species (Vergoulis *et al.* 2012). The second largest manually curated database of miRNA targets is miRTarBase (mirtarbase.mbc.nctu.edu.tw), with 51,460 miRNA–target interactions from 18 species (Hsu *et al.* 2011). Both databases offer extensive information on the techniques used for validation, target gene information, and miRNA binding sites, as well as links to supporting publications.

Reagents and fly stocks for miRNA functional studies: loss of function

As for protein-coding genes, loss-of-function analysis is an appropriate route to study the endogenous function of miRNA-encoding genes. However, there are relatively few loss-of-function mutant strains available for individual miRNAs. Engineered transgenic constructs with multiple miRNA antisense binding sites, which should sequester miRNA and are known as "sponges," have been described as an alternative to knockouts (Kumar *et al.* 2008; Ebert and Sharp 2010; Papapetrou *et al.* 2010). In *Drosophila*, miRNA sponges have been shown to accurately copy the phenotype of loss-of-function mutants (Loya *et al.* 2009). Because miRNA sponges are under UAS control, tissue-specific Gal4 drivers can be used to manipulate spatiotemporal miRNA levels in intact flies. Furthermore, by introducing multiple different miRNA binding sites, sponge technology can also be used to study the role of different miRNAs simultaneously. Recently, Fulga and colleagues completed

production of a transgenic library of conditional miRNA sponges (T. A. Fulga, E. M. McNeill, R. Binari, J. Yelick, A. Blanche, M. Booker, M. Schnall-Levin, Y. Zhao, T. DeLucca, F. Bejarano, Z. Han, E. C. Lai, D. Wall, N. Perrimon, and D. Van Vactor, unpublished data). The plasmids can also be used for high-throughput screening of miRNA function in cell culture.

Reagents and fly stocks for miRNA functional studies: gain of function

Phenotype-driven gain-of-function miRNA screens are another approach to identifying the activities of miRNAs. Overexpression of miRNAs can produce phenotypes due to ectopic repression of one or more targets, lending insights into miRNA–target relationships. However, a significant caveat to this approach is that mis- and/or overexpression of some miRNAs might lead to binding of nonphysiological targets. Currently, there are four collections of conditional miRNA overexpression fly lines (Bejarano *et al.* 2012; Schertel *et al.* 2012; Szuplewski *et al.* 2012). Bejarano *et al.* (2012) constructed two collections of conditional miRNA overexpression lines. The first was built using vectors that coexpress DsRed as a marker and were randomly inserted into the genome. This collection includes 665 lines, comprising 165 different transgenes that cover 149 distinct miRNA hairpins. Because random insertion can lead to position effects, a second collection of 107 miRNA overexpression inserts was built using a site-directed attP vector. The second set of constructs also replaced DsRed with luciferase. Szuplewski *et al.* (2012) and Schertel *et al.* (2012) also generated collections using site-specific integration. Szuplewski *et al.* (2012) generated 109 miRNA overexpression lines and Schertel *et al.* (2012) generated 180 miRNA overexpression lines. The majority of these fly stocks can be directly obtained from their respective laboratories, and the 107 miRNA overexpression fly stocks from Bejarano *et al.* (2012) can be obtained from BDSC. As mentioned above, the plasmids used to construct miRNA overexpression fly stocks can be used to screen miRNA function in cell culture using high-throughput screening techniques (Silver *et al.* 2007).

Example Functional Genomics Workflows

To illustrate how some of the tools that we highlight in this review can be integrated in the context of network studies, we briefly describe two recent studies from our group aimed to identify new components involved in autophagy and Hippo signaling. Many other examples and references of large-scale screens, especially starting from genome-wide RNAi screens, are available on the "Publications" page at flyrnai.org (see also Mohr *et al.* 2010).

In workflow one (Figure 1), (J. Zirin, J. Nieuwenhuis, A. Samsonova, R. Tao, and N. Perrimon, unpublished data) assembled a list of genes representing the *Drosophila* orthologs of a mammalian autophagy network (Behrends *et al.* 2010). *Drosophila* proteins related to the mammalian proteins were identified using DIOPT. Such a gene list can be further supplemented with additional genes using searches for genes with common expression patterns, functions, disease links, *etc.*, using tools like FlyAtlas, modMINE, FlyMine, and/or DIOPT-DIST. Next, a small library of dsRNAs for cell-based RNAi with multiple reagent-per-gene coverage was assembled based on the DRSC collection of RNAi reagents, supplemented with new designs generated using SnapDragon (Flockhart *et al.* 2012). The library was then screened using a robust autophagy primary muscle cell-based assay. Following an initial statistical analysis of the screen data, tools like COMPLEAT and DAVID were used to help identify high-confidence hits, as well as determine what functions are enriched for positive hits as compared with the starting set of genes in the library. Tools like UP-TORR and FlyPrimerBank can be used to help identify *in vivo* fly stocks for follow-up studies and design primers for qPCR validation of knockdown. Finally, after publication, the screen data are shared with the community through the DRSC (flyrnai.org), where they are subsequently captured by GenomeRNAi (Schmidt *et al.* 2013) and other metasites.

In workflow two (Figure 2), Kwon *et al.* (2013) performed a large-scale mass spec study followed by functional validation in both cell lines and *in vivo*. Specifically, the study started with a pathway of interest (Hippo signaling), such as might be explored at FlyReactome (Croft *et al.* 2013). Next, clones were identified and tagged fusion proteins developed for known pathway members. These were then expressed in fly cells, followed by pulldown, purification of complexes, and mass spec to identify complex members (Kwon *et al.* 2013). An initial analysis of mass spec results can be used to generate gene lists from each pulldown experiment. Tools like DroID (Yu *et al.* 2008; Murali *et al.* 2011), COMPLEAT (Vinayagam *et al.* 2013a), and DAVID (Dennis *et al.* 2003) can be used to assess quality and look for supporting evidence for potential new interactions. Cell-based reagents were identified from the DRSC collection using UP-TORR (Hu *et al.* 2013a) and newly identified genes were tested functionally using RNAi. For workflows like this one, tools like SignedPPI (Vinayagam *et al.* 2013b) and FlyBase phenotype annotations for individual genes can also provide a picture of potential functional roles and relationships among complex members.

Concluding Remarks

In this review, we describe many of the resources and online tools available to *Drosophila* researchers for functional genomics approaches. Specifically, we list resources useful for three fundamental steps, *i.e.*, to identify or design reagents; to narrow down lists of candidate or initial positive-scoring genes or proteins, such as following a large-scale study; and to expand lists of candidate genes. Numerous online tools and community resources have been developed. These serve the *Drosophila* community by providing genome-scale fly stock collections and other reagent for genetics, transcriptomics, and proteomics, as well as online tools for genome annotation, computational predictions, access to previous data, ortholog predictions, and data analysis, integration, and visualization.

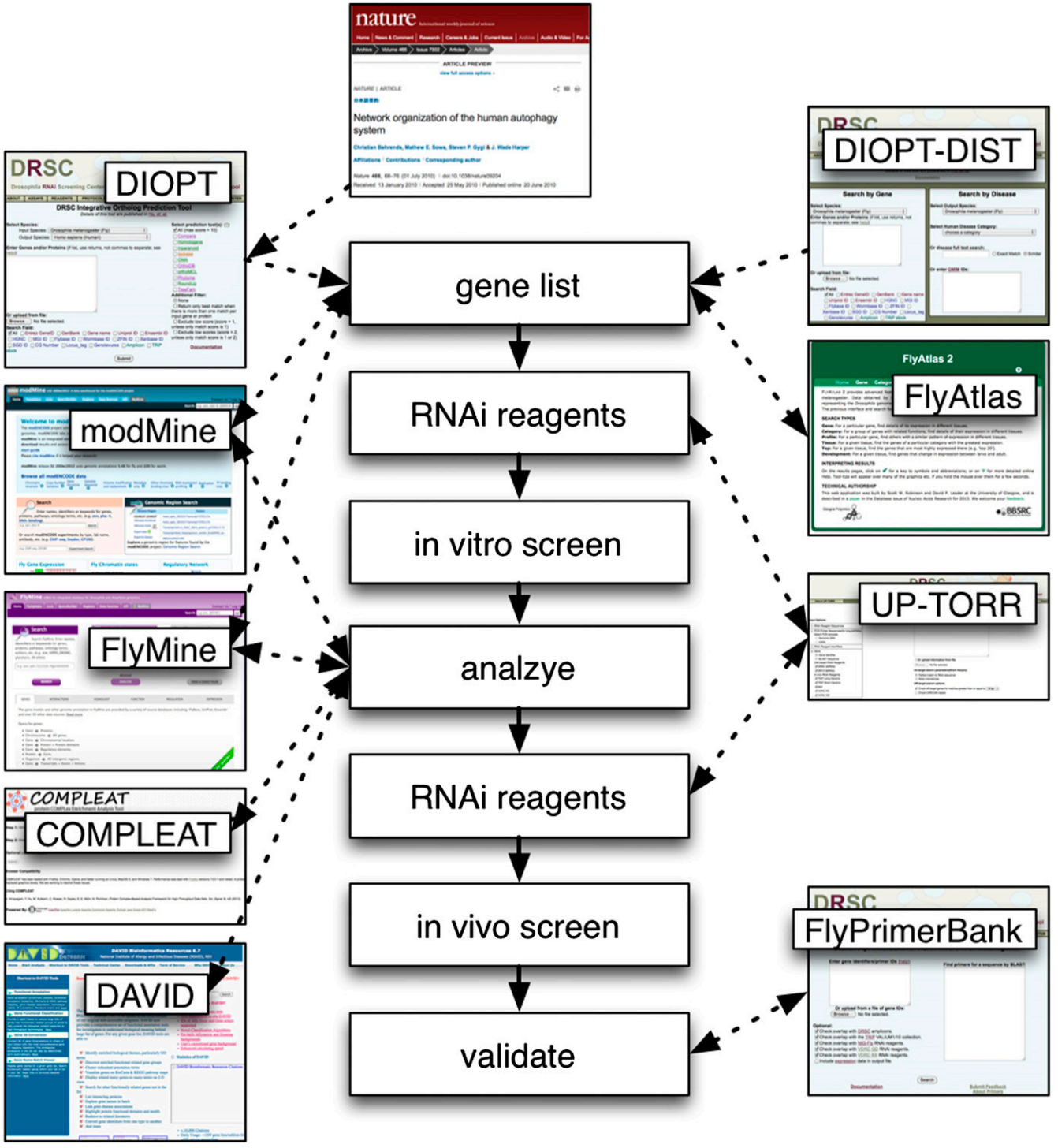

Figure 1 Functional genomics workflow example one. In this workflow based on J. Zirin, J. Nieuwenhuis, A. Samsonova, R. Tao, and N. Perrimon (unpublished data), results from a human proteomics study were converted to fly orthologs, which were then used to create an RNAi library for cell-based screening. Results were analyzed and integrated using enrichment and protein complex-based approaches.

Many of the online resources have become essential to the day-to-day workflow of *Drosophila* research. Nevertheless, there remain opportunities for refinement of existing tools and development of new ones. As mentioned previously, approaches for identifying genes with shared phenotypes is one area that might benefit from new development. Related to this is the development of new text-mining tools. Due to the long history of scientific publication, the majority of data exists in free-text format in the literature. Text-mining tools might help automate curation at FlyBase (McQuilton 2012)

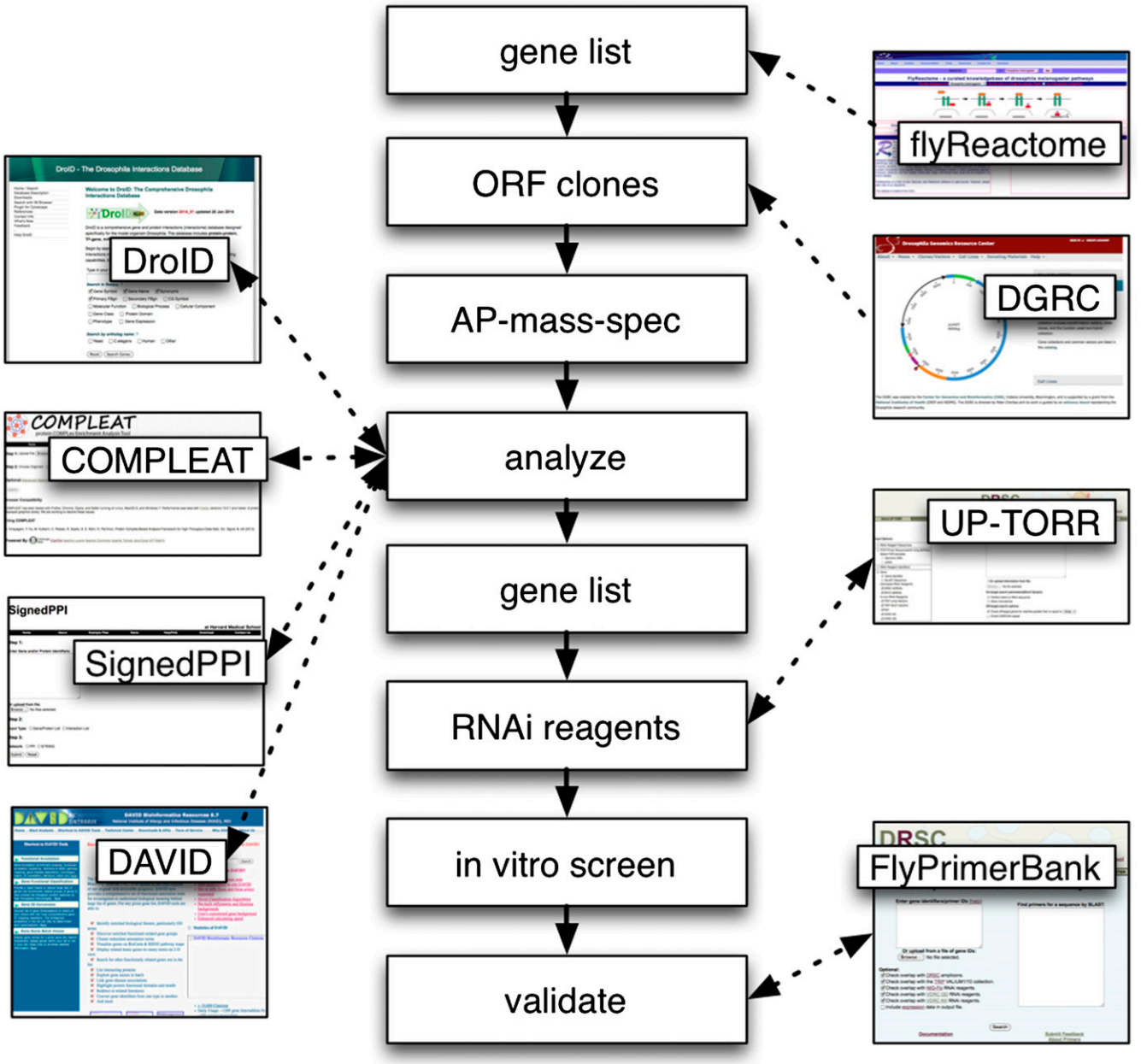

Figure 2 Functional genomics workflow example two. In this workflow based on Kwon *et al.* (2013), known components of the Hippo signaling transduction pathway were tagged and used for mass spectrometry-based discovery of protein complexes. Results were then compared with existing complex and interaction data to assess quality and functionally tested using RNAi.

as well as help catalog previously identified gene–phenotype relationships, protein–protein interactions, *etc*. Another area for improvement is in annotation of regulatory sequences. Good prediction algorithms and experimental data are available for annotation of coding sequences. However, the locations and extents of UTRs, promoters, and enhancers are more difficult to annotate. For mammalian genes, the TRANSFAC database (Wingender *et al.* 1996) reports experimentally validated transcription factor binding sites, consensus binding sequences, *etc*. The availability of various relevant data from modENCODE (Contrino *et al.* 2012) should have a positive impact on annotation of regulatory regions in the future. Moreover, new databases relevant to fly transcriptomics are already emerging, *e.g.*, OnTheFly (Shazman *et al.* 2013) and REDfly (Gallo *et al.* 2011). Additionally, although there is a wealth of information about signaling and biochemical pathways in *Drosophila*, the current high-quality pathway annotations grossly undersample the available information. Extensive pathway annotation for mammalian genes is available at BioCarta (biocarta.com) and Reactome (reactome.org),

as well as commercial sources like Ingenuity. By contrast, for the fly, flyReactome (Croft *et al.* 2013) appears to be the only source, and includes only eight pathways.

We expect that data visualization tools will be increasingly important as more and more large-scale studies are performed and the data made available. These tools can help researchers view and compare data, leading to development of new hypotheses. It can also be helpful to view experimental images associated with small- or large-scale studies, *e.g.*, light microscopy views of wing defects or fluorescence micrographs showing defects identified in a cell-based study. There are significant challenges in making image-based data available for view and download, including the basic problem of how to store, manage, and provide access to what tend to be large files. For both data visualization and images, it is tempting to look beyond what is being developed specifically for bioinformatics and begin to ask if we can learn from other areas, such as 3D and interactive video gaming and social networking sites like Flickr or Instagram that routinely handle large image datasets.

Altogether, we conclude that the *Drosophila* field is well served by existing functional genomics resources and online tools, and that with careful consideration of which tools to use and how to use them, the resources can have significant positive impact on study design, analysis, and integration. Moreover, given a continued commitment to update and maintenance of existing tools, as well as continued development of new resources, support for *Drosophila* functional genomics should continue to improve in the future.

Acknowledgments

We extend our thanks to Kathy Matthews and Annette Parks at the Bloomington *Drosophila* Stock Center and to Bill Gelbart, Susan St. Pierre, and curators at FlyBase for helpful suggestions. We gratefully acknowledge support of the *Drosophila* RNAi Screening Center and Transgenic RNAi Project by National Institutes of Health–National Institute of General Medical Sciences (NIH–NIGMS) R01 GM067761, NIH-NIGMS R01 GM084947, and NIH–Office of Research Infrastructure Programs R24 RR032668. S.E.M. is also supported in part by the Dana Farber/Harvard Cancer Center and N.P. is an investigator of the Howard Hughes Medical Institute.

Literature Cited

Alberts, B., 2002 *Molecular Biology of the Cell*, Garland Science, New York.

Arziman, Z., T. Horn, and M. Boutros, 2005 E-RNAi: a web application to design optimized RNAi constructs. Nucleic Acids Res. 33: W582–W588.

Bassett, A. R., C. Tibbit, C. P. Ponting, and J. L. Liu, 2013a Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep. 4: 220–228.

Bassett, A. R., C. Tibbit, C. P. Ponting, and J. L. Liu, 2013b Mutagenesis and homologous recombination in Drosophila cell lines using CRISPR/Cas9. Biol. Open 3: 42–49.

Baxevanis, A. D., 2012 Searching Online Mendelian Inheritance in Man (OMIM) for information on genetic loci involved in human disease. Curr. Protoc. Hum Genet. Chapter 9: Unit 9.13.

Behrends, C., M. E. Sowa, S. P. Gygi, and J. W. Harper, 2010 Network organization of the human autophagy system. Nature 466: 68–76.

Bejarano, F., D. Bortolamiol-Becet, Q. Dai, K. Sun, A. Saj *et al.*, 2012 A genome-wide transgenic resource for conditional expression of Drosophila microRNAs. Development 139: 2821–2831.

Bellen, H. J., R. W. Levis, G. Liao, Y. He, J. W. Carlson *et al.*, 2004 The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics 167: 761–781.

Bellen, H. J., R. W. Levis, Y. He, J. W. Carlson, M. Evans-Holm *et al.*, 2011 The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics 188: 731–743.

Betel, D., A. Koppal, P. Agius, C. Sander, and C. Leslie, 2010 Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. 11: R90.

Beumer, K. J., J. K. Trautman, M. Christian, T. J. Dahlem, C. M. Lake *et al.*, 2013 Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in Drosophila. G3 3: 1717–1725.

Bischof, J., M. Bjorklund, E. Furger, C. Schertel, J. Taipale *et al.*, 2013 A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development 140: 2434–2442.

Buszczak, M., S. Paterno, D. Lighthouse, J. Bachman, J. Planck *et al.*, 2007 The carnegie protein trap library: a versatile tool for Drosophila developmental studies. Genetics 175: 1505–1531.

Cade, L., D. Reyon, W. Y. Hwang, S. Q. Tsai, S. Patel *et al.*, 2012 Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. Nucleic Acids Res. 40: 8001–8010.

Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang *et al.*, 2011 Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39: e82.

Chan, C. Y., C. E. Lawrence, and Y. Ding, 2005 Structure clustering features on the Sfold Web server. Bioinformatics 21: 3926–3928.

Chen, K. F., and D. C. Crowther, 2012 Functional genomics in Drosophila models of human disease. Brief. Funct. Genomics 11: 405–415.

Cherbas, L., and L. Gong, 2014 Cell lines. Methods DOI: 10.1016/j.ymeth.2014.01.006.

Cherbas, L., A. Willingham, D. Zhang, L. Yang, Y. Zou *et al.*, 2011 The transcriptional diversity of 25 Drosophila cell lines. Genome Res. 21: 301–314.

Chien, S., L. T. Reiter, E. Bier, and M. Gribskov, 2002 Homophila: human disease gene cognates in Drosophila. Nucleic Acids Res. 30: 149–151.

Chintapalli, V. R., J. Wang, and J. A. Dow, 2007 Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat. Genet. 39: 715–720.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto *et al.*, 2013 Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819–823.

Contrino, S., R. N. Smith, D. Butano, A. Carr, F. Hu *et al.*, 2012 modMine: flexible access to modENCODE data. Nucleic Acids Res. 40: D1082–D1088.

Cook, K. R., A. L. Parks, L. M. Jacobus, T. C. Kaufman, and K. A. Matthews, 2010a New research resources at the Bloomington Drosophila Stock Center. Fly (Austin) 4: 88–91.

Cook, R. K., S. J. Christensen, J. A. Deal, R. A. Coburn, M. E. Deal *et al.*, 2012 The generation of chromosomal deletions to provide

extensive coverage and subdivision of the Drosophila melanogaster genome. Genome Biol. 13: R21.
Cook, R. K., M. E. Deal, J. A. Deal, R. D. Garton, C. A. Brown *et al.*, 2010b A new resource for characterizing X-linked genes in Drosophila melanogaster: systematic coverage and subdivision of the X chromosome with nested, Y-linked duplications. Genetics 186: 1095–1109.
Cradick, T. J., E. J. Fine, C. J. Antico, and G. Bao, 2013 CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. 41: 9584–9592.
Crocker, J., and D. L. Stern, 2013 TALE-mediated modulation of transcriptional enhancers in vivo. Nat. Methods 10: 762–767.
Croft, D., A. F. Mundo, R. Haw, M. Milacic, J. Weiser *et al.*, 2013 The Reactome pathway knowledgebase. Nucleic Acids Res. 42: D472–D477.
Dahlem, T. J., K. Hoshijima, M. J. Jurynec, D. Gunther, C. G. Starker *et al.*, 2012 Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. PLoS Genet. 8: e1002861.
Dai, R., and S. A. Ahmed, 2011 MicroRNA, a new paradigm for understanding immunoregulation, inflammation, and autoimmune diseases. Transl. Res. 157: 163–179.
Dennis, G. Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao *et al.*, 2003 DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4: 3.
Dietzl, G., D. Chen, F. Schnorrer, K. C. Su, Y. Barinova *et al.*, 2007 A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151–156.
Ding, Q., Y. K. Lee, E. A. Schaefer, D. T. Peters, A. Veres *et al.*, 2013 A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12: 238–251.
Doyle, E. L., N. J. Booher, D. S. Standage, D. F. Voytas, V. P. Brendel *et al.*, 2012 TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res. 40: W117–W122.
Ebert, M. S., and P. A. Sharp, 2010 MicroRNA sponges: progress and possibilities. RNA 16: 2043–2050.
Ejsmont, R. K., M. Sarov, S. Winkler, K. A. Lipinski, and P. Tomancak, 2009 A toolkit for high-throughput, cross-species gene engineering in Drosophila. Nat. Methods 6: 435–437.
Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis *et al.*, 2013 Ensembl 2014. Nucleic Acids Res. 42: D749–D755.
Flockhart, I. T., M. Booker, Y. Hu, B. McElvany, Q. Gilly *et al.*, 2012 FlyRNAi.org–the database of the Drosophila RNAi screening center: 2012 update. Nucleic Acids Res. 40: D715–D719.
Forslund, K., F. Schreiber, N. Thanintorn, and E. L. Sonnhammer, 2011 OrthoDisease: tracking disease gene orthologs across 100 species. Brief. Bioinform. 12: 463–473.
Fu, Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon *et al.*, 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31: 822–826.
Gallo, S. M., D. T. Gerrard, D. Miner, M. Simich, B. Des Soye *et al.*, 2011 REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic Acids Res. 39: D118–D123.
Gao, X., J. Yang, J. C. Tsang, J. Ooi, D. Wu *et al.*, 2013 Reprogramming to Pluripotency Using Designer TALE Transcription Factors Targeting Enhancers. Stem Cell Rep. 1: 183–197.
Gilbert, L. A., M. H. Larson, L. Morsut, Z. Liu, G. A. Brar *et al.*, 2013 CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154: 442–451.
Gratz, S. J., A. M. Cummings, J. N. Nguyen, D. C. Hamm, L. K. Donohue *et al.*, 2013 Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194: 1029–1035.
Graveley, B. R., A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011 The developmental transcriptome of Drosophila melanogaster. Nature 471: 473–479.
Groth, P., N. Pavlova, I. Kalev, S. Tonov, G. Georgiev *et al.*, 2007 PhenomicDB: a new cross-species genotype/phenotype resource. Nucleic Acids Res. 35: D696–D699.
Guruharsha, K. G., J. F. Rual, B. Zhai, J. Mintseris, P. Vaidya *et al.*, 2011 A protein complex network of Drosophila melanogaster. Cell 147: 690–703.
Hamosh, A., A. F. Scott, J. Amberger, D. Valle, and V. A. McKusick, 2000 Online Mendelian Inheritance in Man (OMIM). Hum. Mutat. 15: 57–61.
Heigwer, F., G. Kerr, and M. Boutros, 2014 E-CRISP: fast CRISPR target site identification. Nat. Methods 11: 122–123.
Heigwer, F., G. Kerr, N. Walther, K. Glaeser, O. Pelz *et al.*, 2013 E-TALEN: a web tool to design TALENs for genome engineering. Nucleic Acids Res. 41: e190.
Herscovitch, M., E. Perkins, A. Baltus, and M. Fan, 2012 Addgene provides an open forum for plasmid sharing. Nat. Biotechnol. 30: 316–317.
Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta *et al.*, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA 106: 9362–9367.
Hisano, Y., S. Ota, K. Arakawa, M. Muraki, N. Kono *et al.*, 2013 Quantitative assay for TALEN activity at endogenous genomic loci. Biol. Open 2: 363–367.
Horn, T., Z. Arziman, J. Berger, and M. Boutros, 2007 GenomeRNAi: a database for cell-based RNAi phenotypes. Nucleic Acids Res. 35: D492–D497.
Horn, T., T. Sandmann, and M. Boutros, 2010 Design and evaluation of genome-wide libraries for RNA interference screens. Genome Biol. 11: R61.
Hsu, P. D., D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann *et al.*, 2013 DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31: 827–832.
Hsu, S. D., F. M. Lin, W. Y. Wu, C. Liang, W. C. Huang *et al.*, 2011 miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 39: D163–D169.
Hu, Y., I. Flockhart, A. Vinayagam, C. Bergwitz, B. Berger *et al.*, 2011 An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics 12: 357.
Hu, Y., C. Roesel, I. Flockhart, L. Perkins, N. Perrimon *et al.*, 2013a UP-TORR: online tool for accurate and Up-to-Date annotation of RNAi Reagents. Genetics 195: 37–45.
Hu, Y., R. Sopko, M. Foos, C. Kelley, I. Flockhart *et al.*, 2013b FlyPrimerBank: an online database for Drosophila melanogaster gene expression analysis and knockdown evaluation of RNAi reagents. G3 3: 1607–1616.
Huntzinger, E., and E. Izaurralde, 2011 Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. 12: 99–110.
Jenett, A., G. M. Rubin, T. T. Ngo, D. Shepherd, C. Murphy *et al.*, 2012 A GAL4-driver line resource for Drosophila neurobiology. Cell Rep. 2: 991–1001.
Joshi-Tope, G., M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt *et al.*, 2005 Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 33: D428–D432.
Kahraman, A., A. Avramov, L. G. Nashev, D. Popov, R. Ternes *et al.*, 2005 PhenomicDB: a multi-species genotype/phenotype database for comparative phenomics. Bioinformatics 21: 418–420.
Katsuyama, T., A. Akmammedov, M. Seimiya, S. C. Hess, C. Sievers *et al.*, 2013 An efficient strategy for TALEN-mediated genome engineering in Drosophila. Nucleic Acids Res. 41: e163.
Kelso, R. J., M. Buszczak, A. T. Quinones, C. Castiblanco, S. Mazzalupo *et al.*, 2004 Flytrap, a database documenting a GFP protein-trap insertion screen in Drosophila melanogaster. Nucleic Acids Res. 32: D418–D420.

Kheradpour, P., A. Stark, S. Roy, and M. Kellis, 2007 Reliable prediction of regulator targets using 12 Drosophila genomes. Genome Res. 17: 1919–1931.
Kohl, M., S. Wiese, and B. Warscheid, 2011 Cytoscape: software for visualization and analysis of biological networks. Methods Mol. Biol. 696: 291–303.
Kondo, S., and R. Ueda, 2013 Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics 195: 715–721.
Kondo, S., M. Booker, and N. Perrimon, 2009 Cross-species RNAi rescue platform in Drosophila melanogaster. Genetics 183: 1165–1173.
Kondo, T., T. Sakuma, H. Wada, A. Akimoto-Kato, T. Yamamoto *et al.*, 2014 TALEN-induced gene knock out in Drosophila. Dev. Growth Differ. 56: 86–91.
Konermann, S., M. D. Brigham, A. E. Trevino, P. D. Hsu, M. Heidenreich *et al.*, 2013 Optical control of mammalian endogenous transcription and epigenetic states. Nature 500: 472–476.
Kozomara, A., and S. Griffiths-Jones, 2011 miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39: D152–D157.
Kumar, M. S., S. J. Erkeland, R. E. Pester, C. Y. Chen, M. S. Ebert *et al.*, 2008 Suppression of non-small cell lung tumor development by the let-7 microRNA family. Proc. Natl. Acad. Sci. USA 105: 3903–3908.
Kumar, S., C. Konikoff, B. Van Emden, C. Busick, K. T. Davis *et al.*, 2011 FlyExpress: visual mining of spatiotemporal patterns for genes and publications in Drosophila embryogenesis. Bioinformatics 27: 3319–3320.
Kwon, Y., A. Vinayagam, X. Sun, N. Dephoure, S. P. Gygi *et al.*, 2013 The Hippo signaling pathway interactome. Science 342: 737–740.
Langer, C. C., R. K. Ejsmont, C. Schonbauer, F. Schnorrer, and P. Tomancak, 2010 In vivo RNAi rescue in Drosophila melanogaster with genomic transgenes from Drosophila pseudoobscura. PLoS ONE 5: e8928.
Lecuyer, E., H. Yoshida, N. Parthasarathy, C. Alm, T. Babak *et al.*, 2007 Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell 131: 174–187.
Lewis, B. P., C. B. Burge, and D. P. Bartel, 2005 Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15–20.
Liu, J., C. Li, Z. Yu, P. Huang, H. Wu *et al.*, 2012 Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J. Genet. Genomics 39: 209–215.
Loya, C. M., C. S. Lu, D. Van Vactor, and T. A. Fulga, 2009 Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. Nat. Methods 6: 897–903.
Lyne, R., R. Smith, K. Rutherford, M. Wakeling, A. Varley *et al.*, 2007 FlyMine: an integrated database for Drosophila and Anopheles genomics. Genome Biol. 8: R129.
Ma, M., A. Y. Ye, W. Zheng, and L. Kong, 2013 A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes. BioMed Res. Int. 2013: 270805.
Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The Drosophila melanogaster Genetic Reference Panel. Nature 482: 173–178.
Maeder, M. L., S. J. Linder, V. M. Cascio, Y. Fu, Q. H. Ho *et al.*, 2013a CRISPR RNA-guided activation of endogenous human genes. Nat. Methods 10: 977–979.
Maeder, M. L., S. J. Linder, D. Reyon, J. F. Angstman, Y. Fu *et al.*, 2013b Robust, synergistic regulation of human gene expression using TALE activators. Nat. Methods 10: 243–245.
Mali, P., J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner *et al.*, 2013 CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31: 833–838.
McQuilton, P., 2012 Opportunities for text mining in the FlyBase genetic literature curation workflow. Database (Oxford) bas039.
McQuilton, P., S. E. St Pierre, and J. Thurmond, 2012 FlyBase 101—the basics of navigating FlyBase. Nucleic Acids Res. 40: D706–D714.
Mi, H., A. Muruganujan, and P. D. Thomas, 2013 PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 41: D377–D386.
Mohr, S., C. Bakal, and N. Perrimon, 2010 Genomic screening with RNAi: results and challenges. Annu. Rev. Biochem. 79: 37–64.
Morin, X., R. Daneman, M. Zavortink, and W. Chia, 2001 A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. Proc. Natl. Acad. Sci. USA 98: 15050–15055.
Murali, T., S. Pacifico, J. Yu, S. Guest, G. G. Roberts, 3rd *et al.*, 2011 DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. Nucleic Acids Res. 39: D736–D743.
Mussolino, C., R. Morbitzer, F. Lutge, N. Dannemann, T. Lahaye *et al.*, 2011 A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 39: 9283–9293.
Na, D., and M. Rouf, C. J. O'Kane, D. C. Rubinsztein, and J. Gsponer, 2013 NeuroGeM, a knowledgebase of genetic modifiers in neurodegenerative diseases. BMC Med. Genomics 6: 52.
NCBI Resource Coordinators, 2013 Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 41: D8–D20.
Neff, K. L., D. P. Argue, A. C. Ma, H. B. Lee, K. J. Clark *et al.*, 2013 Mojo Hand, a TALEN design tool for genome editing applications. BMC Bioinformatics 14: 1.
Ni, J. Q., R. Zhou, B. Czech, L. P. Liu, L. Holderbaum *et al.*, 2011 A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat. Methods 8: 405–407.
O'Brien, K. P., I. Westerlund, and E. L. Sonnhammer, 2004 OrthoDisease: a database of human disease orthologs. Hum. Mutat. 24: 112–119.
Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono *et al.*, 1999 KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27: 29–34.
Okamura, K., E. Ladewig, L. Zhou, and E. C. Lai, 2013 Functional small RNAs are generated from select miRNA hairpin loops in flies and mammals. Genes Dev. 27: 778–792.
Papapetrou, E. P., J. E. Korkola, and M. Sadelain, 2010 A genetic strategy for single and combinatorial analysis of miRNA function in mammalian hematopoietic stem cells. Stem Cells 28: 287–296.
Paraskevopoulou, M. D., G. Georgakilas, N. Kostoulas, I. S. Vlachos, T. Vergoulis *et al.*, 2013 DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. 41: W169–W173.
Pattanayak, V., S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna *et al.*, 2013 High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31: 839–843.
Perez-Pinera, P., D. D. Kocak, C. M. Vockley, A. F. Adler, A. M. Kabadi *et al.*, 2013a RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods 10: 973–976.
Perez-Pinera, P., D. G. Ousterout, J. M. Brunger, A. M. Farin, K. A. Glass *et al.*, 2013b Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat. Methods 10: 239–242.
Perrimon, N., J. Q. Ni, and L. Perkins, 2010 In vivo RNAi: today and tomorrow. Cold Spring Harb. Perspect. Biol. 2: a003640.
Pfeiffer, B. D., A. Jenett, A. S. Hammonds, T. T. Ngo, S. Misra *et al.*, 2008 Tools for neuroanatomy and neurogenetics in Drosophila. Proc. Natl. Acad. Sci. USA 105: 9715–9720.

Reczko, M., M. Maragkakis, P. Alexiou, I. Grosse, and A. G. Hatzigeorgiou, 2012 Functional microRNA targets in protein coding sequences. Bioinformatics 28: 771–776.

Ren, X., J. Sun, B. E. Housden, Y. Hu, C. Roesel *et al.*, 2013 Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proc. Natl. Acad. Sci. USA 110: 19012–19017.

Robinson, S. W., P. Herzyk, J. A. Dow, and D. P. Leader, 2013 FlyAtlas: database of gene expression in the tissues of Drosophila melanogaster. Nucleic Acids Res. 41: D744–D750.

Ruby, J. G., C. H. Jan, and D. P. Bartel, 2007a Intronic microRNA precursors that bypass Drosha processing. Nature 448: 83–86.

Ruby, J. G., A. Stark, W. K. Johnston, M. Kellis, D. P. Bartel *et al.*, 2007b Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. 17: 1850–1864.

Sander, J. D., P. Zaback, J. K. Joung, D. F. Voytas, and D. Dobbs, 2007 Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res. 35: W599–W605.

Sander, J. D., M. L. Maeder, D. Reyon, D. F. Voytas, J. K. Joung *et al.*, 2010 ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. 38: W462–W468.

Schertel, C., T. Rutishauser, K. Forstemann, and K. Basler, 2012 Functional characterization of Drosophila microRNAs by a novel in vivo library. Genetics 192: 1543–1552.

Schmidt, E. E., O. Pelz, S. Buhlmann, G. Kerr, T. Horn *et al.*, 2013 GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. Nucleic Acids Res. 41: D1021–D1026.

Schnall-Levin, M., Y. Zhao, N. Perrimon, and B. Berger, 2010 Conserved microRNA targeting in Drosophila is as widespread in coding regions as in 3′UTRs. Proc. Natl. Acad. Sci. USA 107: 15751–15756.

Sebo, Z. L., H. B. Lee, Y. Peng, and Y. Guo, 2013 A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. Fly (Austin) 8: 1.

Seiler, C. Y., J. G. Park, A. Sharma, P. Hunter, P. Surapaneni *et al.*, 2014 DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. Nucleic Acids Res. 42: D1253–D1260.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang *et al.*, 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13: 2498–2504.

Shazman, S., H. Lee, Y. Socol, R. S. Mann, and B. Honig, 2013 OnTheFly: a database of Drosophila melanogaster transcription factors and their binding sites. Nucleic Acids Res. 42: D167–171.

Silver, S. J., J. W. Hagen, K. Okamura, N. Perrimon, and E. C. Lai, 2007 Functional screening identifies miR-315 as a potent activator of Wingless signaling. Proc. Natl. Acad. Sci. USA 104: 18151–18156.

Singari, S., N. Javeed, N. J. Tardi, S. Marada, J. C. Carlson *et al.*, 2014 Inducible protein traps with dominant phenotypes for functional analysis of the Drosophila genome. Genetics 196: 91–105.

St Pierre, S. E., L. Ponting, R. Stefancsik, and P. McQuilton, 2014 FlyBase 102–advanced approaches to interrogating FlyBase. Nucleic Acids Res. 42: D780–D788.

Stapleton, M., J. Carlson, P. Brokstein, C. Yu, M. Champe *et al.*, 2002 A Drosophila full-length cDNA resource. Genome Biol. 3: RESEARCH0080.

Szuplewski, S., J. M. Kugler, S. F. Lim, P. Verma, Y. W. Chen *et al.*, 2012 MicroRNA transgene overexpression complements deficiency-based modifier screens in Drosophila. Genetics 190: 617–626.

Venken, K. J., K. L. Schulze, N. A. Haelterman, H. Pan, Y. He *et al.*, 2011 MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat. Methods 8: 737–743.

Vergoulis, T., I. S. Vlachos, P. Alexiou, G. Georgakilas, M. Maragkakis *et al.*, 2012 TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res. 40: D222–D229.

Vert, J. P., N. Foveau, C. Lajaunie, and Y. Vandenbrouck, 2006 An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinformatics 7: 520.

Vinayagam, A., Y. Hu, M. Kulkarni, C. Roesel, R. Sopko *et al.*, 2013a Protein complex-based analysis framework for high-throughput data sets. Sci. Signal. 6: rs5.

Vinayagam, A., J. Zirin, C. Roesel, Y. Hu, B. Yilmazel *et al.*, 2013b Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. Nat. Methods 11: 94–99.

Wang, H., and D. Q. Peng, 2011 New insights into the mechanism of low high-density lipoprotein cholesterol in obesity. Lipids Health Dis. 10: 176.

Wilson, R. J., J. L. Goodman, and V. B. Strelets, 2008 FlyBase: integration and improvements to query tools. Nucleic Acids Res. 36: D588–D593.

Wingender, E., P. Dietze, H. Karas, and R. Knuppel, 1996 TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. 24: 238–241.

Yamamoto, M. T., 2010 Drosophila Genetic Resource and Stock Center; The National BioResource Project. Experimental Animals/Japanese Association for Laboratory Animal Science 59: 125–138.

Yu, J., S. Pacifico, G. Liu, and R. L. Finley, Jr.., 2008 DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics 9: 461.

Yu, Z., M. Ren, Z. Wang, B. Zhang, Y. S. Rong *et al.*, 2013 Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. Genetics 195: 289–291.

Zuo, D., S. E. Mohr, Y. Hu, E. Taycher, A. Rolfs *et al.*, 2007 PlasmID: a centralized repository for plasmid clone information and distribution. Nucleic Acids Res. 35: D680–D684.

Communicating editor: P. Sengupta