

Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells

Nadire Ramadan¹, Ian Flockhart¹, Matthew Booker¹, Norbert Perrimon^{1,2} & Bernard Mathey-Prevot¹

¹Department of Genetics and *Drosophila* RNAi Screening Center, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. ²Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. Correspondence should be addressed to B.M.-P. (bprevot@receptor.med.harvard.edu).

Published online 6 September 2007; doi:10.1038/nprot.2007.250

This protocol describes the various steps and considerations involved in planning and carrying out RNA interference (RNAi) genome-wide screens in cultured *Drosophila* cells. We focus largely on the procedures that have been modified as a result of our experience over the past 3 years and of our better understanding of the underlying technology. Specifically, our protocol offers a set of suggestions and considerations for screen optimization and a step-by-step description of the procedures successfully used at the *Drosophila* RNAi Screening Center for screen implementation, data collection and analysis to identify potential hits. In addition, this protocol briefly covers postscreen analysis approaches that are often needed to finalize the hit list. Depending on the scope of the screen and subsequent analysis and validation involved, the full protocol can take anywhere from 3 months to 2 years to complete.

INTRODUCTION

Over the past two decades, great strides in genome sequencing and new technologies for high-throughput analysis have revolutionized studies of gene function. The completion of the *Drosophila* genome sequence in 2000 (ref. 1) provided the opportunity to apply genome-wide approaches to systematically explore gene function in flies. RNA interference (RNAi), usually referred to as the ability of double-stranded RNAs (dsRNAs), small interfering RNAs (siRNAs) or small hairpin RNAs (shRNAs) to silence a target gene through the specific destruction of that gene's mRNA^{2–5} offered a clear-cut way to systematically interrogate the role of all individual genes predicted from genomic sequencing. The observation that RNAi could be experimentally induced in *Drosophila* cell lines^{6,7} and the subsequent success of early large-scale screens^{8–12} led to a proliferation of genome-wide RNAi screens performed in *Drosophila* cells and the creation of the *Drosophila* RNAi Screening Center (DRSC) where interested groups can apply to perform their own screens (http://flyrnai.org/app_policy.html).

Two different philosophies, which are not mutually exclusive, have motivated these screens. The first was to use genome-wide RNAi screens as a gene-discovery tool, much in the same vein as for any traditional genetic screen. This approach is very well suited to situations when the identity of a single key component or even the whole gene network regulating a biological process is unknown or has failed to be identified by other means. For instance, a critical regulator of store-operated Ca²⁺ entry, which had been postulated to exist but had escaped detection, was identified independently by three different groups using slightly different assays in their RNAi screens to capture it^{13–15}. The other motivation for genome-wide screens is to apply a systems-like approach to a particular process to derive a comprehensive picture of all the genes involved and to link them into a functional network. This goal, however, is more difficult to achieve as there are technical hurdles to overcome (such as minimizing the rates of false positives and false negatives) before one can attempt to propose definitive answers. Nonetheless, the screens performed so far have contributed important information to our understanding of biology. They have examined a variety of biological processes, such as the dissection of signaling pathways,

the discovery of genes associated with human diseases or involved in host–pathogen interactions, the identification of genes involved in general cellular functions (such as cell viability, secretion, chromosome segregation) and, more recently, the study of developmental programs that can be duly recapitulated in primary embryonic cultures such as muscle growth and differentiation or neuronal outgrowth (J. Bai, K. Sepp and N.P., unpublished data). For a more thorough discussion of the types of questions addressed in *Drosophila* RNAi screens and a list of primary references, see ref. 16.

Until recently, systematic RNAi screens in *Drosophila* have been limited to a cell culture approach, even though robust RNAi had been achieved *in vivo* by injecting dsRNAs into embryos¹⁷ or through expression during fly development of transgenes encoding extended hairpin-loop interfering RNAs¹⁸. Although large-scale screens using direct injection of dsRNAs into embryos have been done^{19,20}, a number of issues including the amount of reagents needed, the transient nature of the expression of the interfering RNAs and the occasional nonspecific phenotypes resulting from the procedure have limited this approach. Systematic RNAi through expression of hairpin constructs via the UAS/Gal4 system requires genome-wide collections of transgenic flies, which until recently were not available. However, such collections have just been assembled at the Vienna *Drosophila* RNAi Center (<http://stockcenter.vdrc.at/control/main>) and the National Institute of Genetics (NIG)-Fly Stock Center (<http://www.shigen.nig.ac.jp/fly/nigfly/>). As access to these collections is only recent, there is no published information regarding the feasibility of performing genome-wide screens *in vivo* with these collections. Although such screens will be far more labor-intensive and time-consuming than performing similar screens in tissue culture, they will have the advantage of being able to examine a wide array of complex phenotypes (including behavioral), in the whole animal or in a particular tissue depending on the choice of Gal4 line used to drive expression of the UAS-hairpin constructs. One great benefit that these collections will have is that they will provide a rapid means of validating *in vivo* the involvement of candidate genes identified in

BOX 1 | DETECTING POTENTIAL OFF-TARGET SEQUENCES IN DOUBLE-STRANDED RNAs

Off-target effects (OTEs), due to the knock-down of unintended transcripts by cross-hybridizing sequences present in some *Drosophila* RNAi Screening Center (DRSC) 1.0 double-stranded RNAs (dsRNAs), have been reported to affect the rate of false positives in some of the early screens^{21,22}. Because it is often difficult to validate thoroughly all the hits identified in a primary screen due to the very large investment in time and effort required, the occurrence of OTEs loomed as a significant issue that needed to be addressed urgently. The recommendation was to eliminate (to the extent possible) any similar sequence in dsRNAs that had the potential to knock-down transcripts other than its intended target, and to confirm any phenotype obtained with a particular dsRNA with a second, independent dsRNA targeting the same gene²³. To help design dsRNAs free of small sequences identical to other transcripts, informatics approaches have been developed. Freely available web-based tools, such as that at http://www.flyrnai.org/RNAi_find_frag_free.html, are used to find potential OTEs in PCR DNA templates (amplicons) used to generate dsRNAs. Once this 'Predicted Off-Target Free Sequence Regions' tool is opened on a browser, copy and paste one or more nucleotide sequences into the submission box. The tool determines all possible primary targets of the sequence and highlights gene-targeting regions in blue. Stretches in the sequence that cause potential OTEs are highlighted in red.

cell-based RNAi screens. However, RNAi *in vivo* is still an emerging technology that will undoubtedly evolve as the field becomes more mature, just as it has been the case for RNAi in tissue culture.

Indeed, the accumulated experience of the past 3 years has given us a better understanding of how to apply this technology, but has also revealed some inherent limitations associated with it^{16,21,22}. In particular, the observation that off-target effects (OTEs, see **Box 1**) can take place with the use of long dsRNAs in *Drosophila* cells threw a cautionary note in the interpretation of early screens. OTEs arise from dsRNAs that contain simple sequence repeats or sequence matches to nontarget transcripts^{21,22}, and are a source of false positives. In the DRSC 1.0 collection (used in early screens performed at our facility), 39% of the dsRNAs contained at least one predicted simple repeat or perfect sequence identity to another gene. A statistical analysis of the results from 30 screens revealed that such dsRNAs had the tendency to score more often than predicted by chance alone, and were likely a source of false positives²². Ma *et al.* independently arrived at the same conclusion, documenting in particular that CAN repeats needed to be absolutely avoided in the design of dsRNAs to minimize the prevalence of OTEs²¹. Following these reports, new guidelines were proposed²³, including the release of improved libraries²⁴. However, the issue of OTEs is not the only risk of artifacts in high-throughput screens (HTSs). Compounding the biological variability associated with the assay employed in a screen are the random errors that cannot be prevented. Sources of random errors can be due to biological, instrumental and human-related errors. They arise through occasional malfunctions of liquid-handling robotics, random loss of cells during the processing steps or changes in concentrations due to evaporation in certain wells and occasional defects in tissue-culture plates. Collectively, they lead to nonspecific phenotypes in particular wells and contribute to a higher rate of false positives or false negatives. A common case of positional noise is called 'edge-effects' where wells located on the most peripheral rows or columns of a plate score as positives in an assay, likely due to uneven evaporation in these wells. Given that a large number of wells are scored at most in duplicate, it is often difficult to identify and control for this kind of noise in the raw data, and care needs to be taken when interpreting the results of a screen, unless appropriate statistical treatments have been applied to the raw data.

This experimental guide incorporates the latest improvements in the technology, and describes sequentially the different steps of a screen, from wet-bench experiments to data analysis and data

mining. It does not discuss in detail the protocols that have not been significantly changed and that can be found in ref. 25 or online at http://flyrnai.org/methodical_papers.html. Rather, it elaborates on steps where technical or conceptual difficulties are commonly encountered by screeners, namely screen optimization, data collection, hit selection and lastly post-screen analysis of the results.

Scope of this protocol

Although many of the experimental steps in *Drosophila* RNAi screens, such as the introduction of dsRNAs into various *Drosophila* cells or the general design of cell-based assays, have changed little since originally reported^{6,7,25,26}, the design of dsRNAs and acquisition and treatment of data have witnessed important developments. Here we outline the various steps involved in carrying out RNAi HTSs in *Drosophila* cells (**Fig. 1**). We focus largely on the protocols that have been modified or, in some cases, drastically changed to reflect our better understanding of the RNAi process in *Drosophila* cells^{21–23} and spell out the experimental modifications and strategies put in place at the DRSC to address the most common issues faced by screeners. We also suggest a number of strategies to bring into focus the information emanating from large data sets generated by RNAi screens. Although the emphasis is on the procedures used in high-throughput applications, many of the steps in the PROCEDURE are also relevant to cases where only a few dsRNAs are being screened.

Although *Drosophila* RNAi screens are based on the same general principle of introducing long dsRNAs into cells, the source and the manner in which the dsRNA libraries have been designed and generated can vary (**Table 1**). Investigators have also used different plate formats in their assays (384-well plates versus 96- or 48-well plates^{8,27}), or relied on alternative methods to capture the signal [plate reader or microscopy versus fluorescence-activated cell sorting (FACS) analysis]^{12,28}. Lastly, each group has developed their own strategy to interpret numerical data (statistical treatment, choice of threshold)^{21,28–30}. A general outline of the methodology involved in HTSs is given in refs. 25 and 31, and a brief discussion of other experimental strategies and what has been learned through these approaches can be found in refs. 16 and 32–34.

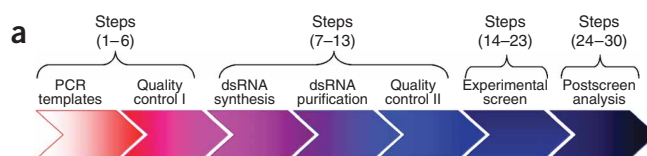
Experimental design: PCR

In contrast to vertebrate cells, *Drosophila* cells do not mount an interferon response to long dsRNAs and many cell lines will take up dsRNAs spontaneously when they are presented to the cells in

medium-lacking serum^{6,7}. Therefore, there has been very little incentive to develop siRNA or shRNA reagents for *Drosophila* cells, in part because of cost (with regard to the use of siRNAs), but also because both siRNA and shRNAs have been associated with issues of specificity and efficacy that are often more difficult to control for than in the case of dsRNAs. For these reasons, investigators have concentrated on developing methods to generate dsRNA libraries after *in vitro* transcription (IVT) of PCR templates flanked with the appropriate polymerase promoter. Two general strategies have been described for the synthesis of DNA templates for dsRNAs^{24,35} (Fig. 2). The first approach involves a two-step amplification reaction: the first amplification uses gene-specific primers tailed with unique identifier sequences; the second amplification uses primers corresponding to the unique identifier sequences (present in the initial set of primers) and tailed with the T7 promoter³⁵. The second approach, which we detail in the protocol below, eliminates the second PCR step, thereby saving labor. Please note that care should be taken to avoid cross-contamination when the PCR templates are re-amplified using only the T7 primer. For large-scale re-amplification reactions, the unique primers tailed with the T7 sequence should be used. Both approaches yield PCR templates that direct the synthesis of dsRNAs, using T7 polymerase for IVT of the templates. The dsRNAs can range in size from 250 to 900 nt or longer. Upon uptake by cells, these dsRNAs are processed by Dicer to yield siRNAs of 21–23 nt in length. In large-scale screens, it is customary to assay only one dsRNA per well. However, screens that are designed to identify synthetic activities can be performed with more than one dsRNA per well. In fact, we have been able to treat cells with up to four dsRNAs at a time and observe good knock-down efficiency for all targets (B.M.-P., unpublished data). It is generally recommended to target coding sequences of genes, and avoid intronic sequences where possible. However, in some cases, it is not always practical as genes can be very short and interrupted with very small introns. In those cases, exceptions to the rule are acceptable. Lastly, care should be taken to target common exons when multiple transcripts are expressed from a single gene due to alternative splicing.

Experimental design: cell culture

When choosing a *Drosophila* cell system, there are a number of issues to consider: established cell lines versus primary cells, ease of culture and susceptibility to RNAi and suitability to the biological question under study. For instance, is the pathway under study functional in the cells, or are the cells a good model for the biological process of interest; do they need to be modified to express a specific gene required for the assay read-out (e.g., reporter gene) or to activate or modify the pathway or process of interest? Is the desired phenotype robust and penetrant in the chosen cell system? To address these various questions, prior testing will be necessary. There are a few available resources to help identify appropriate candidate cell lines. Information on the general characteristics of cell lines can be found at <https://dgrc.cgb.indiana.edu/cells/store/catalog.html> (see also Table 2). Although systematic microarray analyses are not available for all *Drosophila* cell lines, an excellent starting point to check for the expression of particular genes that might be required or influence an assay is the FLIGHT database (http://flight.licr.org/browse/browse_cell_lines.jsp).



b



B1. Robotically array collection of dsRNAs against *Drosophila* genes in 384-well plates (Step 13).



B2. Dispense cells on pre-arrayed plates, and place in incubator (25 °C, 3–8 d) for RNAi treatment (Step 14).



B3. Use automated liquid handler dispenser to process cells for assay read-out (Step 14 and Box 3).



B4. Monitor read-out on high-throughput detection platforms: plate readers or microscopes (Step 15).

Figure 1 | General outline of a genome-wide RNA interference (RNAi) screen. (a) The outline is broken down among its major experimental and data analysis sections. Each section comprises several steps that are numbered according to our protocol (see text for details). (b) Outline of the experimental part only of the screen is schematized together with typical equipment used (B1: Assay TekBench robot (TekCel), B2: WellMate cell dispenser; B3: μ Fill liquid dispenser; B4 top left, clockwise: Aerius scanner/reader, Analyst GT plate reader, Evotec microscope).



PROTOCOL

TABLE 1 | *Drosophila* libraries available for RNA interference screens.

Collection and coverage	Description	Comments
DRSC 2.0: genome-wide; currently in use at the <i>Drosophila</i> RNAi Screening Center (DRSC) (replaces DRSC 1.0) (http://flyrnai.com)	22,490 double-stranded RNA (dsRNAs) with average length of 400 bp	94.3% of all annotated genes are targeted by at least one dsRNA free of off-target (OT). Approximately 7,000 genes targeted by more than one dsRNA
Ambion: genome-wide library (based on version 3.0 genome annotation)	dsRNAs to 13,071 Genes, ranging in length from 300 to 800 bp	No public data available regarding OT sequences. Several screens reported to have been successful
Open Biosystems: genome-wide; Release 1.0 and 2.0	15,881 dsRNAs with average length of 200–800 bp	No public data available regarding OT sequences
BDGP: genes represented in cDNA set 1 collection	4,923 dsRNAs of variable size	No public data available regarding OT sequences
BDGP: DGC1 and DGC2; covers 70% of the annotated genes	Approximately 500-bp PCR fragments tailed with T7 promoter were amplified from these collections to serve as template for dsRNA synthesis	No public data available regarding OT sequences
DKFZ: genome-wide library (http://rna.dkfz.de)	Bioinformatics approach to predict specific and efficient target sites. Two-step PCR amplification of PCR templates	Controlled for OT sequences and optimized for RNA interference (RNAi) efficiency. New release

Additional details for each of the nonDRSC collections are available from the respective web pages: Ambion library: (<http://www.ambion.com/catalog/ProdGrp.html?fkApp=25&fkProdGrp=326>); BDGP cDNA collections: (<http://www.fruitfly.org/DGC/index.html>); DKFZ new release: (T. Horn and M. Boutros, personal communication; <http://rna.dkfz.de>); Open Systems library: (<http://www.openbiosystems.com/RNAi/DrosophilaRNAicollections>).

What follows below is a brief guide regarding what to be aware of when using *Drosophila* cell lines.

- Many *Drosophila* cell lines and useful expression vectors can be obtained from the *Drosophila* Genomic Resource Center (<https://dgrc.cgb.indiana.edu/>).
- Drosophila* cells grow between 23 and 25 °C, and do not require a CO₂ or humidified incubator for growth. However, it is best to limit evaporation of the medium by growing cells in a somewhat moist environment.
- Most *Drosophila* cell lines use the same medium, but some lines will require additional factors for growth (e.g., fly extracts) as in the case of clone 8 cells. Detailed protocols to prepare fly extracts can be found at <http://biology.st-andrews.ac.uk/sites/flycell/flyextract.html>, or on our website, http://flyrnai.org/all_protocols.html#Medium. Be aware that commercial lots of *Drosophila* medium can vary significantly in their ability to support growth of certain cell lines (like Kc₁₆₇), even though they might be appropriate for other common cell lines (such as S2 cells). Furthermore, there can be significant differences in the chemical composition of *Drosophila* media sold under the same name among different suppliers. Careful testing of media lots should always be done before launching a large-scale experiment, and provision needs to be made such that the selected lot of medium will not run out during the experiment.
- FCS lots need to be checked carefully as some can be toxic to *Drosophila* cells. Most FCS providers do not test their lots against *Drosophila* cells, and the occurrence of toxicity associated with serum is significantly more of an issue than it is for mammalian cells.

- Cell lines commonly used in screens are S2 cells (and their many derivatives, e.g., S2R+), Kc₁₆₇ and clone 8 cells. S2 and Kc₁₆₇ cells are derived from embryonic hemocytes (akin to macrophages), and grow as semi-adherent or adherent cells. These cells often have a non-normal mode of chromosomes. However, they offer the advantage that they will efficiently take up dsRNA in serum-free medium without requiring any transfection reagents (bathing method).
- To ensure reproducibility, it is best to keep cells on a rigorous schedule for passaging and also to avoid cell overgrowth. Overgrown cultures are less susceptible to RNAi, even if they are plated at the correct density during the course of the assay. Also, avoid using cells kept in culture for more than 6 months. It is best to freeze cells at early passages, and thaw aliquots as needed.
- Recently, RNAi screens have been successfully performed in primary cultures derived from dissociated, post-gastrulation

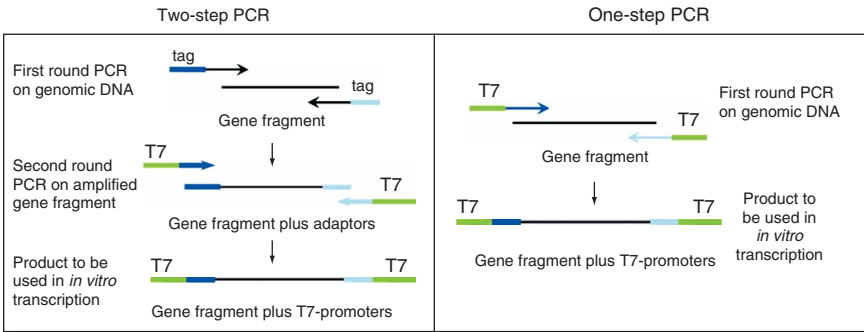


Figure 2 | Comparison of one-step and two-step PCR. This schematic depicts the difference between the one-step and two-step PCR amplification process for generating DNA templates that can be transcribed *in vitro* into double-stranded RNAs (dsRNAs). In the two-step PCR, gene-specific primers with a unique identifier tag (represented in blue and cyan) added at their ends are first used to amplify genomic DNA. During the second step, primers containing the complementary sequence to the unique tag and the T7 promoter sequence (green) are used, generating DNA templates that will be flanked by the T7 promoter. In the one-step PCR, the T7 promoter sequence is added directly to the 5-end of each gene-specific primer. The blue and cyan portions represent the gene-specific portions of the primer set.

TABLE 2 | *Drosophila* cell lines available for high-throughput screen interference RNA screening. This table is taken, with permission, from ref. 25.

Line	Origin	Reference	Characteristics
Schneider's Line S2—(S2)	Dissociated embryos, near hatching (<i>Oregon R</i>)	53	Hemocyte-like gene expression, phagocytic, semi-adherent in colonies, round, granular cytoplasm
Schneider's S2—(S2 ^{star})	Dissociated embryos, near hatching (<i>Oregon R</i>)	53	Subclone of S2 cells that is more responsive to lipopolysaccharide, particularly after priming with 1 μ M 20-hydroxy ecdysone
Schneider's S2—(S2C)	Dissociated embryos, near hatching (<i>Oregon R</i>)	53	Hemocyte-like gene expression, semi-adherent in colonies, round, granular cytoplasm
Schneider's S2—(S2-R+)	Dissociated embryos, near hatching (<i>Oregon R</i>)	53, 54	Hemocyte-like gene expression, phagocytic, adherent, flat cells; Fz+ and Wg-responsive
Schneider's S2—(DL2)	Dissociated embryos, near hatching (<i>Oregon R</i>)	53	Hemocyte-like gene expression, phagocytic, adherent monolayer of uniformly round, smooth cells
Schneider's Line S3	Dissociated embryos, near hatching (<i>Oregon R</i>)	53	Adherent, spindle-shaped cells, ecdysone-responsive, grow in clumps
Kc (Kc 167)	Dissociated embryos, 8–12 h (F2 <i>ebony</i> \times <i>sepia</i>)	55	Hemocyte-like gene expression, phagocytic, uniformly round, clump in sheets, ecdysone-responsive into adherent, bipolar spindle-shaped cells
l(2)mbn	Third instar larvae tumorous hemocytes, (<i>l(2)malignant blood neoplasm</i>)	56	larger cells, larger granular, complex cytoplasm, phagocytic, aneuploid, heterogenous size and shape
ML-DmBG2	Dissociated third instar larvae brain and ventral ganglia (<i>y v f mal</i>)	57	Acetylcholine, horseradish peroxidase expression, neuronal-like processes
ML-DmBG6	Dissociated third instar larvae brain and ventral ganglia (<i>y v f mal</i>)	57	Acetylcholine, HRP expression, neuronal-like processes
Clone 8	Third instar larvae wing imaginal discs	58	Derived from wing columnar epithelial cells, adherent, will form multiple layers, conserved signaling pathways

embryos (4–6 h after egg laying). A protocol for deriving such cultures, as well as an example of a screen illustrating promising biological applications, will be presented elsewhere (J. Bai and N.P., unpublished data).

Experimental design: screening considerations

1. *The assay read-out needs to be optimized to the plate-format of the screen:* the fact that an assay can be easily read in a 96-well plate does not necessarily translate into the same behavior when it is performed in the 384-well plate format. Often times, the signal-to-noise ratio drops off dramatically, making readings much more difficult to interpret than in a 96-well format.
2. *Favor multiple read-outs:* whenever possible, develop an assay read-out that measures many different features of interest and include, if possible, a marker for cell number and/or viability. Apply extreme care when interpreting results with dsRNAs causing cell viability phenotypes, as many assays tend to identify them as 'hits'. In reality, they are more often false positives.
3. *Automate the various steps:* experimental noise is the enemy of high-throughput assays. To minimize it, we recommend automation for seeding cells, adding reagents and aspirating medium. See EQUIPMENT SETUP for further details.
4. *Minimize medium loss in wells during the assay:* the lids provided with most plates will not seal the plates effectively, causing the liquid in the wells to evaporate if the environment is too dry. To safeguard against this, the air in the incubation chamber or incubator should be kept somewhat saturated with water vapor. An easy and cost-effective way to achieve this is by placing the assay plates in a plastic container lined with wet paper towels at the bottom. The lid of the container

should be loosely closed before the container is placed in a temperature-controlled incubator or environment. The added moisture will decrease the loss of liquid volume by keeping the experimental environment humid.

5. *Normalization of data:* include an independent measure of either cell number or cell viability when conducting a reporter assay. Since results can vary with changes in cell numbers, this provides an effective means of normalizing results.
6. *Duplicates:* assay duplicate plates on the same day when doing a screen in duplicate. This will reduce variability between plates as the same reagents, cells and conditions are used.
7. *Controls:* whenever possible, positive and negative controls should be incorporated into the screen. For positive controls, dsRNAs targeting known components of a pathway or biological process are chosen to ensure that RNAi treatment has been successful and leads to a measurable phenotype. In cases where the identity of potential genes involved in the process of interest is unknown, addition of a drug or an experimental treatment resulting in a phenotype of interest should be included to make sure that the assay read-out is reliably detected in high throughput. Negative controls consist of dsRNAs (e.g., LacZ or GFP dsRNAs) that do not target any gene expressed in the cell system used in the assay. These dsRNAs should have no noticeable effect on the read-out.

Experimental design: data collection

Parts of this protocol will necessarily be generic, as each high-throughput RNAi experiment is in some way different from the ones performed earlier. Methods of data collection will vary widely based on the scientific question being asked and the experimental approach chosen (Fig. 3). As a result, screens will generate different

types of data. Screens based on luminescence or fluorescence signals detected by a plate reader will produce one or more numeric data points for each well. Automated microscopy experiments will generate hundreds of gigabytes of image data, which may in turn be run through various automated image analysis programs to produce quantitative data. Other assay-specific instrumentation that defies these broad classifications will inevitably be encountered. Ideally, the goal, in as much as it is possible, is to end up with a set of numeric data representative of the phenotypes or measurements observed in each well. This format allows the data to be processed through various statistical analyses in order to identify wells of interest where values differ significantly from the value measured in control wells (no dsRNAs or control dsRNAs).

Data treatment and analysis

Once all the numerical data from a screen have been entered into the database, analysis of the results can proceed. As the data sets generated by HTSs are quite large, they require various visualization tools and statistical treatments to identify the wells of interests or ‘hits’^{30,36,37}. There are a variety of open source (e.g., Bioconductor, cellHTS) or commercial (e.g., MatLab, Spotfire) software solutions available that can be used to facilitate the statistical treatment and visualization of RNAi HTS, but we will not review them here. Instead, we outline various data treatment and analysis tools from the DRSC, which can be freely accessed from our website (http://flyrnai.org/RNAi_tools.html).

Statistical normality versus normalization

It is important to distinguish between the terms ‘statistical normality’ and ‘normalization’. Determining statistical normality means examining the distribution of a single data set to see if the results show bias or are skewed in some way. Testing for statistical normality is an essential step in the analysis of HTS data sets. Most statistical tests that are applied to large data sets work under the assumption that the data set to be tested is statistically normal. Most data sets contain some bias, which can be compensated for through the use of various statistical techniques. ‘Normalization’ refers to the manipulation of data from multiple sets to make data from different plates comparable³⁸.

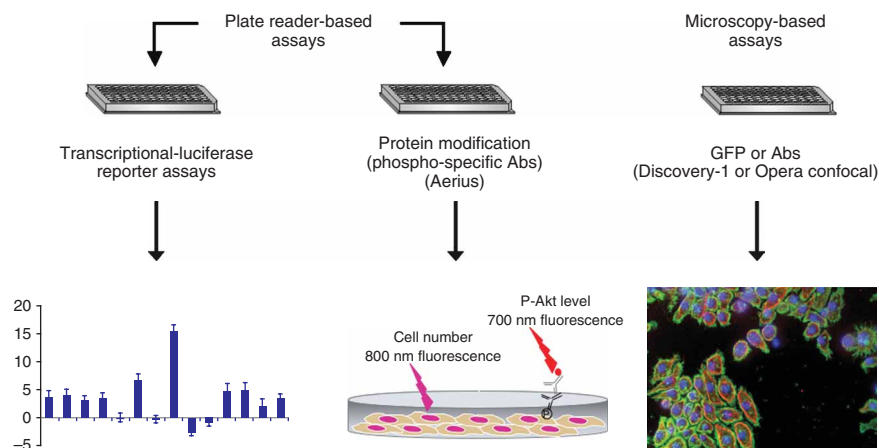


Figure 3 | Common approaches for assay detection in RNA interference (RNAi) screens. High-throughput RNAi screens commonly rely on plate-reader/scanner- and microscopy-based assays. Plate readers are used to detect luminescent and fluorescent read-outs. The theoretical bar graph (shown underneath) displays relative Z-scores calculated from normalized values following individual double-stranded RNA (dsRNA) perturbations. A recent alternative is the use of plate scanners (such as the Aerius, Li-Cor) to detect fluorescence in the far red, typically in Ab-based assays that measure protein modification(s). In the diagram shown below this approach, the red arrow represents emission at 700 nm from a conjugated secondary antibody binding to a phospho-Akt-specific Ab. The reddish purple arrow indicates far-red emission fluorescence emitted by a dye that binds DNA (used for calculating cell numbers after RNAi treatment). The other common approach in RNAi screens is to perform a visual screen, which requires the use of a high-throughput epifluorescent or confocal microscope (depicted here is an image taken with the Evotec confocal microscope, Dresden, Germany). The displayed image was obtained in high-throughput mode ($\times 60$) with the Evotec microscope, and shows Kc167 cells in a 384-well stained for DNA (4',6-diamidino-2-phenylindole), F-actin (phalloidin) and tubulin (conjugated anti-tubulin Ab). The decision to choose between these approaches should rest on experimental evidence for the most robust and reliable way of detecting a phenotype in the chosen plate-format.

The thresholds to select wells of interest are often arbitrary. Usually, one attempts to reach an appropriate balance between sensitivity and selectivity.

Postscreen analysis and filtering of candidate hits for validation

The experience at the DRSC is that the number of genes represented by positive dsRNAs resulting from a genome-wide primary screen varies considerably, ranging from 100 to more than 2,500. Even at the lower end of this range, the number of potential candidates requires a great deal of work and effort to perform a postscreen analysis of the results and to decide which genes ought to be selected for further validation. No single strategy exists to ensure the correct way of solving this critical issue, as the approach will vary depending on the ultimate goal(s) of the screen. However, we include a series of decision-tree options (see Fig. 4) that can be used to review primary data with the goal of curating a primary hit list for further validation. The postdata collection analysis and validation steps of a screen take much longer than screeners often assume.

MATERIALS

REAGENTS

- Wild-type Oregon-R *Drosophila* (source of genomic DNA) **▲ CRITICAL** We have tried different reagent kits and have found the kits listed to work best for our purposes.
- Choice-Taq Mastermix (Denville Scientific, cat. no. CB4070), includes dNTPs, Taq polymerase and reaction buffer
- 1-kb Plus DNA ladder (Invitrogen, cat. no. 10787-018)

- T7 MEGascript Kit (Applied Biosystems, cat. no. AMB1334), includes ATP, GTP, CTP, UTP, T7 polymerase, T7 reaction buffer, DNase and nuclease-free water
- Qiagen RNeasy mini kit (Qiagen, cat. no. 74104) includes RLT and RPE buffers
- RNase/DNase-free water (Sigma, cat. no. W4502)
- Buffer A (see REAGENT SETUP)
- Buffer B (see REAGENT SETUP)

- Isopropanol (Sigma) **! CAUTION** Flammable.
- Ethanol (EtOH; Sigma) **! CAUTION** Flammable.
- Tris-EDTA (Sigma)
- β -Mercaptoethanol (β -ME; Sigma) **! CAUTION** Flammable and toxic if inhaled.
- Agarose (Sigma)
- 1 \times TBE buffer (see REAGENT SETUP)
- DNA molecular weight ladder (Invitrogen)
- Effectene transfection reagent (Qiagen, cat. no. 301425) (contains effectene, enhancer and buffer EC)
- Dual-Glo luciferase (Promega)
- CellTiter-Glo (Promega)
- Steady-Glo luciferase (Promega)
- The Drosophila Resource Center (<https://dgrc.cgb.indiana.edu/>). Many useful DNA vectors, expression constructs for *Drosophila* tissue-culture cells and *Drosophila* cell lines can be requested there
- MetaMorph (Molecular Devices, Sunnyvale, CA). This commercial software includes an image analysis program
- Primer3 (<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi>). We use this program to optimize the design of primers for amplification of DNA templates. We use the default settings (including 60 T_m and 20–80% GC content)

EQUIPMENT

- Thermocycler
- TekBench robot to array dsRNAs and transfer PCR templates (TekCel, Hopkinton, MA)
- Liquid Dispenser such as WellMate dispenser (Matrix Technologies, Hudson, NH) and μ Fill microplate dispenser (BioTek, Winooski, VT)
- MultiScreen PCR₉₆ filter plate (Millipore, Billerica, MA)
- MultiScreen vacuum manifold (Millipore, Billerica, MA)
- Multiwell plates (various suppliers) (see EQUIPMENT SETUP)
- Aspiration wand (VP Scientific, San Diego, CA)
- Plate reader with stacker such as the Analyst GT (Molecular Devices)
- High-throughput automated epifluorescent microscope such as the ImageXpress screening system (Molecular Devices, Sunnyvale, CA) or the Opera high content screening system (Evotec Technologies, Hamburg, Germany)
- Spectrophotometer
- Microscope plate-loading robotic arm such as the BioRobot Twister II (Qiagen Inc., Valencia, CA)
- Plastic ‘Tupperware’ container (various suppliers)
- Sorvall refrigerated centrifuge (Thermo Fisher Scientific, Inc, Waltham, MA)
- 0.22- μ m filter unit (Nalge Nunc International, Rochester, NY)
- Cell scraper (Nalge Nunc International, Rochester, NY)
- DRSC tools (http://flyrnai.org/RNAi_tools.html). This webpage lists a number of bioinformatics tools primarily developed for the analysis of RNAi screens in *Drosophila*
- OTE tool (http://www.flyrnai.org/RNAi_find_frag_free.html). This DRSC tool is particularly useful for selecting regions free of predicted off-targets
- The Heatmap tool (http://flyRNAi.org/cgi-bin/RNAi_heatmap_public.pl). This tool helps visualize the numeric data generated for a plate. Data displayed on this page is organized into grids by plate and well location, and can be color-coded by relative value. There is an internal version that works with data stored in our database and an external version; both versions accept data formatted as a text or Excel (.xls) file.
- FLIGHT (<http://flight.licr.org>). Excellent database designed to facilitate the integration of data from high-throughput experiments carried out in *Drosophila* cell culture³⁹
- FlyBase (<http://www.flybase.org>). Main database for information about *Drosophila*

REAGENT SETUP

Buffer A 100 mM Tris-HCl (Sigma) pH 7.5, 100 mM EDTA (Sigma), 100 mM NaCl (Sigma), 0.5% SDS (wt/vol) (Sigma).

Buffer B 1 Part 5 M KOAc (Sigma), 2 parts 6 M LiCl (Sigma).

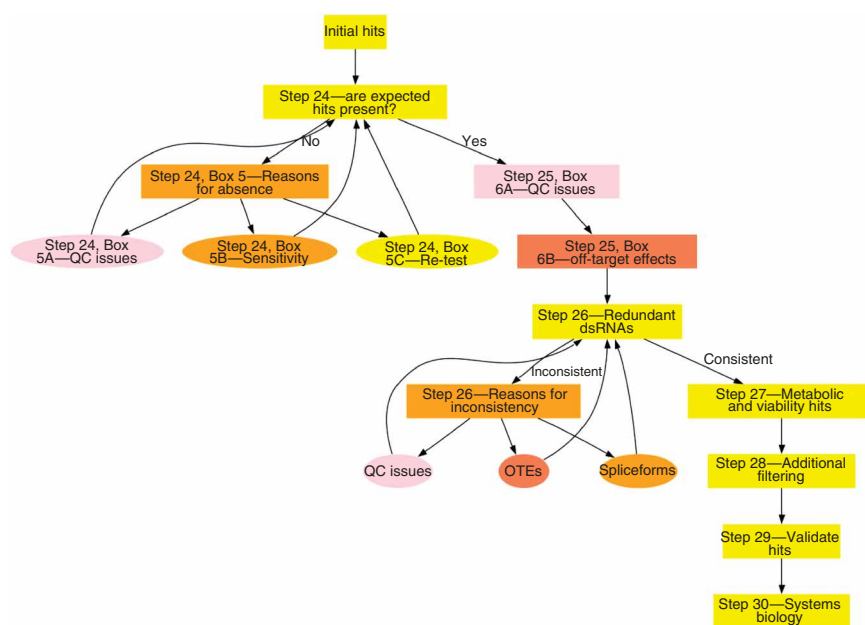


Figure 4 | Postscreen strategy to evaluate initial hits in a screen. Directed diagram showing important decision points based on a postscreen analysis of the data to determine the relevance and specificity of the large number of candidate genes identified in genome-wide RNA interference (RNAi) screens. Protocol Steps 24–30 describe this approach in more detail.

1 \times TBE buffer 0.09 M Tris base (Sigma), 0.09 M boric acid (Sigma), 2.5 mM EDTA, pH 8.3 (Sigma).

EQUIPMENT SETUP

Instrumentation to add cells and sterile reagents Dispensers such as the WellMate or the Multidrop (Thermo-Electron, Waltham, MA) are convenient. A highly reproducible volume of cells or reagents is rapidly dispensed into each well, thereby limiting errors encountered by hand pipetting, and speeding up the process considerably. Furthermore, these dispensers can be kept sterile, a necessity in a cell-based assay.

Instrumentation to handle liquid aspiration To remove any liquid from the plates, an aspiration wand is used. When using a wand, set the vacuum power and the depth of aspiration to appropriate levels. For the latter, tape the wand on both ends (before the first tip and after the last) to provide a cushion between the plate and the wand, and also prevent the vacuum tips from touching the bottom of the well. **▲ CRITICAL** Avoid having the vacuum tips touch the bottom of the wells, as it will cause significant cell loss during aspiration and introduce a large variance in cell number. Before the addition or removal of medium or reagents, spin down the plates in a table-top centrifuge (2 min at 290g is more than sufficient). This will cause any floating or semi-adherent cells to reattach to the bottom of the well, thereby limiting cell count variability in wells, particularly after aspiration.

Multi-well plates There are many different multi-well plates to choose from for performing assays. One consideration in selecting a plate format is the economy of scale. Plates with 384 wells will need much less dsRNA per well than 48-well plates. However, signal-to-noise ratios in 384-well plates can be much worse than in 96- or 48-well plates. Once a particular format has been chosen, select the plate type that is appropriate for that particular assay. For luminescence assays, a white plate with a solid bottom is best. This will maximize the luminescence reading as light will be reflected by the white surface and escape from the surface of the well directly into the optic sensors of the plate reader. A clear bottom plate with the bottom sealed with an opaque sticker is not recommended for this use, as light will leak over to adjacent wells creating well-to-well crosstalk and a decreased reading. Fluorescent screens require a black-walled, clear bottom plate. When measuring fluorescence, it is best to read from the bottom of the plate as this reduces the signal-to-noise ratio by measuring the fluorescence closer to the source of emission. Visual screens that are routinely based on fluorescence signals are best carried out in the same type of plates. It is best to use plates with a thin plastic bottom that is optically clear to produce good images and to reduce time taken for focusing.



PROTOCOL

PROCEDURE

Generation of sequence-optimized templates for IVT of dsRNAs ● **TIMING** Variable according to number of amplicons: 1 d for 50, 3 months or more for genome-wide library

1| Select DNA regions to serve as templates for dsRNAs. The region (usually 400 bp or longer) corresponding to a gene of interest should be common to all splice-forms of the gene, should avoid introns and, to minimize potential OTEs, the target sequence should not contain any 19-nt sequences with perfect identity to other genes²². To facilitate this step, we developed a tool⁴⁰ based on an algorithm that predicts the occurrence of short sequence matches (16–50 bp) to all annotated *Drosophila* genes (for more details see **Box 1**). If using a library that is preassembled (see **Table 1** for available libraries from commercial sources or screening centers), proceed directly to Step 14.

2| Design PCR primer pairs (20 nt in length) corresponding to the 5'- and 3'-ends of each region using Primer3 (ref. 41) set to its default settings.

▲ **CRITICAL STEP** Check with the basic local alignment search tool (BLAST) accessible from FlyBase that the primers are unique in the fly genome to rule out mis-priming events in which sequences are amplified in addition to the intended target.

3| Append a T7 promoter sequence (5'-TAATACGACTCACTATAGGG) to the 5'-end of each primer pair. In this manner, each PCR fragment will be flanked by the T7 promoter, which will allow binding of the T7 polymerase used for dsRNA synthesis.

4| Use the resulting primer pairs to amplify each target sequence from genomic DNA isolated from wild-type flies (e.g., Oregon R, Canton S or yw flies) (see **Box 2**). Set up 100-μl reactions, as detailed below, typically in a 96-well plate containing 96 different specific primer pairs.

Component	Amount per reaction (μl)	Final
Genomic DNA (0.35 mg ml ⁻¹)	1	350 ng
Forward primer (10 μM)	2	20 μM
Reverse primer (10 μM)	2	20 μM
Choice- <i>Taq</i> Mastermix	50	—
DNase/RNase-free water	45	100 μl (final volume)

▲ **CRITICAL STEP** Take care not to cross-contaminate reactions. New tips should be used between reactions to ensure purity of sample. If necessary, add controls to the plate to confirm that the conditions are working. For a positive control, use a sample that has worked well in past amplifications. As a negative control, use water in place of genomic DNA.

BOX 2 | ISOLATION OF *DROSOPHILA* GENOMIC DNA

● **TIMING**: Approximately 2 h plus an overnight incubation.

1. Freeze 30 wild-type flies at -80 °C for 5 min.
2. Add 200 μl of buffer A to frozen flies; grind flies in a dounce tissue homogenizer.
3. Add another 200 μl buffer A used to lyse fly tissue; grind until only the cuticle is left.
4. Incubate at 65 °C for 30 min.
5. Add 800 μl of buffer B to precipitate any debris.
6. Place on ice for 10 min.
7. Spin at 11,200g for 15 min.
8. Transfer supernatant into two new microfuge tubes
9. Add 700 μl of isopropanol for every 1 ml of supernatant.
- ! **CAUTION**: Isopropanol is flammable.
10. Spin at 11,200g for 15 min.
11. Remove and discard supernatant.
12. Wash DNA pellet with 1 ml of cold 70% (vol/vol) ethanol (EtOH).
- ! **CAUTION**: EtOH is flammable.
13. Spin 11,200g for 5 min.
14. Remove and discard supernatant; allow DNA pellet to air-dry.

▲ **CRITICAL STEP**: Evaporate all ethanol to ensure good-quality DNA; residual EtOH will hinder PCRs.

15. Add 200 μl of Tris-EDTA; let sit overnight at room temperature to allow DNA to resuspend fully.
16. Run a 5 μl aliquot on an agarose gel. Expect to see a high molecular weight band (bright intensity). Additional bands migrating approximately 1,000 bp can sometimes be observed. This is acceptable. If a smear is present, the DNA preparation is not good.

! **CAUTION**: Agarose gels contain ethidium bromide (carcinogenic).

5| Run the following touchdown PCR program.

Cycle number	Denature	Anneal	Extend	Hold
1	94 °C, 3 min	—	—	—
2–11	94 °C, 30 s	60 °C (–1 °C/cyc), 45 s	72 °C, 1 min	—
12–34	94 °C, 30 s	50 °C, 45 s	72 °C, 1 min	—
35	—	—	72 °C, 10 min	4 °C

▲ CRITICAL STEP Since the optimal annealing temperature for each primer pair may differ, it may not be possible to obtain products when using a single annealing temperature. To alleviate this problem, perform a touchdown PCR that will cater to most primer designs.

6| Run 5 µl of each PCR template on an agarose gel (0.7–1%) to check for size and quantity. When the reaction is successful, a single band of the expected size is observed. Quantity can be roughly estimated by comparing the brightness of the band to the brightness of a similarly migrating DNA marker band since the concentration of DNA marker bands should be known. We normally run the 1-kb Plus DNA ladder in the outside lanes of our gels.

! CAUTION Agarose gel contains ethidium bromide.

■ PAUSE POINT PCR products can be stored at –20 °C for up to 1 year. For longer storage, place at –80 °C.

? TROUBLESHOOTING

IVT of dsRNAs ● **TIMING** 10 min plus overnight incubation

7| Set up the IVT reaction as detailed in the table below. We routinely use the T7 MEGAscript kit from Applied Biosystems (formerly Ambion).

Component	Reagent provided in T7 MEGAscript kit	Amount per reaction	Final
ATP (75 mM)	Yes	4 µl	300 mM
UTP (75 mM)	Yes	4 µl	300 mM
CTP (75 mM)	Yes	4 µl	300 mM
GTP (75 mM)	Yes	4 µl	300 mM
T7 Enzyme	Yes	4 µl	4 µl
T7 reaction buffer	Yes	4 µl	4 µl
Nuclease-free water	Yes	6 µl	6 µl
PCR product	No	10 µl (15–20 µl if PCR yield is low)	10 µl
Total reaction volume		40 µl (45–50 µl if PCR yield is low)	

8| Incubate at 37 °C for 16 h or overnight.

Purification of dsRNAs ● **TIMING** 60–115 min

9| Add 2 µl of [2 U µl^{–1}] DNase (also included in the T7 MEGAscript kit)—incubate for 45 min at 37 °C.

▲ CRITICAL STEP The DNase treatment is an essential step and is required to remove all DNA from solution. Residual DNA can hinder RNAi effectiveness.

■ PAUSE POINT dsRNA can be stored at –20 °C for up to 1 year.

10| Purify the dsRNAs. For a small number of dsRNAs (less than 96), use option A. For purification of dsRNAs in 96-well plates, use a filter plate (option B).

(A) Purification with Qiagen RNeasy kit (for fewer than 96 samples) ● **TIMING** 15 min

(i) Bring volume up to 100 µl with RNase/DNase-free water.

(ii) Add 350 µl of RLT/β-ME; mix well.

(iii) Add 250 µl 100% EtOH.

! CAUTION EtOH is flammable.

(iv) Transfer dsRNA to a spin column; centrifuge at 8,000g (10,000 r.p.m.) at room temperature (approximately 23 °C) for 30 s in a table-top centrifuge.

(v) Add 500 µl RPE to wash column; centrifuge at 8,000g (10,000 r.p.m.) at room temperature for 30 s. Move spin column to a new collection tube.

(vi) Add 500 µl RPE to wash column; Centrifuge at 8,000g (10,000 r.p.m.) at room temperature for 2 min to dry column. Move spin column to a new microfuge tube.

(vii) Resuspend dsRNA in 30-µl nuclease-free water, and centrifuge at maximum speed at room temperature for 5 min.

PROTOCOL

(viii) Add 30 μL of additional nuclease-free water; centrifuge at maximum speed at room temperature for 5 min.

▲ **CRITICAL STEP** This last centrifugation is critical for proper elution of all dsRNA left in the spin column.

(B) Purification with filter plate (for more than 96 samples) ● **TIMING** 70 min

(i) Bring volume in each well up to 150 μL with RNase/DNase-free water.

(ii) Transfer dsRNA into a Millipore filter/purification plate.

(iii) Attach plate to a vacuum Millipore manifold.

(iv) Vacuum out all liquid until filter membrane is mostly dry (approximately 30 min).

▲ **CRITICAL STEP** Do not allow vacuum pressure to exceed 8 psi (check gauge on vacuum manifold). Vacuum pressure of greater than 10 psi will result in loss of product.

(v) Add 200 μL nuclease-free water to each well. Seal plate with a foil seal to avoid cross-contamination.

(vi) Place plate on a shaker at a low speed (when using a vortexer, setting should not exceed 3) for 30 min to resuspend dsRNA back into solution.

▲ **CRITICAL STEP** Shaking too fast will cause dsRNA to splash out of the plate.

(vii) Transfer resuspended dsRNA to a new 96-well storage plate.

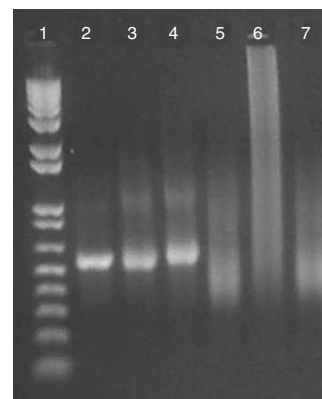


Figure 5 | Examples of acceptable and unacceptable double-stranded RNA (dsRNA) products. Illustration of acceptable and unacceptable dsRNA products run on a 1% (wt/vol) agarose gel stained with ethidium bromide. Lane 1: DNA markers (1-kb Plus DNA ladder, Invitrogen). Lanes 2–4: examples of acceptable dsRNA products. The dsRNAs run as discrete bands. A faint, slower migrating band is commonly observed under these conditions. Lanes 5–7: examples of unacceptable dsRNA products (smearing, no discrete bands).

11 | Run 5 μL of each dsRNA on a 0.7–1.0% agarose gel, and check for proper size and integrity of dsRNA bands. They should match the size of their corresponding PCR templates. They may, on occasion, appear a bit more smeared than their PCR counterparts. This is acceptable as long as the dominant band is of the correct size (see **Fig. 5**).

! **CAUTION** Agarose gel contains ethidium bromide (carcinogenic).

■ **PAUSE POINT** dsRNA can be stored at $-20\text{ }^{\circ}\text{C}$ for up to 1 year.

? **TROUBLESHOOTING**

12 | Determine each dsRNA concentration (C) using a spectrophotometer as follows:

$$C(\text{in } \mu\text{g } \mu\text{L}^{-1}) = (\text{OD}_{260})(45)(\text{dilution factor})/(1,000),$$

where 45 is the value for the extinction coefficient used for dsRNA.

13 | Dilute dsRNA to working concentration with RNase/DNase-free water ($0.05\text{ } \mu\text{g } \mu\text{L}^{-1}$ for bathing and $0.016\text{ } \mu\text{g } \mu\text{L}^{-1}$ for transfection). For a 384-well plate, robotically add 250 ng (5 μL of working stock) of dsRNA per well for the bathing method, or 80 ng (5 μL of working stock) of dsRNA per well for the transfection method.

▲ **CRITICAL STEP** Different plates need to be used according to whether the read-out will be done in a plate reader or a microscope (see **EQUIPMENT SETUP**). See our website for further suggestions (http://flyrnai.org/RNAi_screen_tips.html).

RNAi treatment

14 | The RNAi treatment can be carried out using either a bathing method (option A, e.g., for S2 and derivatives, or Kc₁₆₇ cells) or by transfection (option B, e.g., for clone 8 cells).

(A) Bathing protocol ● **TIMING** 3–6 d

(i) Plate 10,000–40,000 cells in 10 μL of serum-free medium per well in 384-well plates.

(ii) Incubate cells with 250 ng dsRNA at room temperature for 30 min to 1 h.

▲ **CRITICAL STEP** Less than 30 min may not allow enough time for complete uptake of dsRNAs by the cells, and more than 1 h can lead to cell death due to prolonged serum starvation.

(iii) Centrifuge plate at 290g for 2 min.

(iv) Add 30 μL of complete media to each well. Centrifuge plate at 290g for 2 min.

(v) Seal or incubate plates in humid chamber to prevent evaporation in the wells of the plate. This evaporation could lead to edge-effects, which are observed in rows and columns situated at the periphery of the plate. Results may be affected by this phenomenon. Use of the Heatmap tool for the analysis of plate data (see Step 16) will readily document edge-effects when they occur.

(vi) Incubate at 23–25 $^{\circ}\text{C}$ for 3–5 d and then process cells for assay read-out (see **Box 3**).

BOX 3 | PROCESSING CELLS BEFORE DATA ACQUISITION

Depending on the assay method involved in the screen, cells will need to be processed differently at Step 14. Possible approaches for processing cells for plate reader- and microscopy-based assays are given below.

Luminescence-based plate reader screens

- After 3–5 d incubation, use either the Dual-Glo luciferase, Steady-Glo luciferase or CellTiter-Glo kit (all from Promega).
- To ensure proper cell lysis, make sure to shake the plates on an orbital shaker for 2 min at a low setting.
- Plate should be centrifuged (290g for 2 min) after each reagent addition to ensure proper mixing of cells and reagents.
- Plates should be kept in dark conditions after the addition of luciferase reagent for 10 min to stabilize signal.
- Temperature can affect signal intensity and the rate of decay. For this purpose, plates should be equilibrated to ambient temperature before performing the luciferase assay.

Fluorescence-based plate-reader or microscopy screens

(A) Fixation and antibody staining

- After 3–5 d incubation, cells should be fixed with a formaldehyde solution (3–12% (vol/vol)) or other fixative, depending on optimization.
- Ensure proper removal of fixative by repeated washing steps with PBS (at least three wash cycles should be done for each washing step).
- Add conjugated Abs (as in the case of the in cell-western assay with the LI-COR plate scanner) and proceed according to optimized antibody staining protocol.
- Plate should be centrifuged (290g for 2 min) after each reagent addition to avoid excess cell loss.
- For microscopy screens, plates should be stored at 4 °C until plate is imaged. Signal should be stable for at least 1 week after fixation.

(B) Live cell imaging

- After 3–5 d incubation, the necessary reagents should be added to cells that will cause the fluorescent read-out for the particular assay.
- Plate should be centrifuged (290g for 2 min) after each reagent addition to avoid excess cell loss.

(B) Transfection protocol ● TIMING 3–6 d

- Prepare EC + DNA Mastermix in a microfuge tube. All DNA constructs that need to be introduced into cells (e.g., an experimental reporter, a ligand expression vector or a normalization reporter) and EC buffer (a component of the effectene) are added to the master mix for a 384-well plate. Dilute the DNA constructs into 8.75 µl EC buffer.
▲ **CRITICAL STEP** Keep the nucleic acid concentration equal in all wells and use no more than 175 ng of total nucleic acid (DNA + dsRNA) per 384 well. Mix EC buffer and DNA thoroughly by inversion or pipetting.
- Add enough enhancer to the EC + DNA master mix, such that there is 1.0 µl of enhancer per well. Mix well by inversion. Incubate at room temperature for 2–3 min.
- Add appropriate amount of effectene to the enhancer + EC + DNA master mix such that there is 0.3 µl of effectene per well. Mix final mix by pipetting up and down five to ten times or vortexing on high for 10 s. Immediately begin dispensing 10 µl of the final mix into each well of a 384-well plate pre-arrayed with 80-ng dsRNA per well.
- Incubate the dsRNA and transfection mix for 4–8 min at room temperature.
- Plate 40 µl of medium containing 10,000–40,000 cells in each well of the plate. Centrifuge plate at 290g for 2 min.
▲ **CRITICAL STEP** Add cells to the transfection mix within 20 min of its preparation, or the transfection efficiency begins to fall rapidly.
- Incubate plates in a humidified chamber in a tissue culture incubator for 3–5 d at 23–25 °C, then process cells for assay read-out (see **Box 3**).

? TROUBLESHOOTING

Data collection

15| Collect the data into a common location or a database. Screen data can be collected on a plate reader (option A), automated microscope (option B) or other assay-specific instrumentation (option C). We use an SQL database to track experimental data. Although this is not strictly necessary, it makes manipulation and presentation of the data considerably easier. The important thing is to be able to easily retrieve all of the data for a plate, by well.

(A) Plate reader data ● TIMING The plate reader typically takes approximately 2 min per plate

- Load the treated plates into an automated plate reader (e.g., the Analyst GT) for analysis.
- Select the measurement parameters (luminescence or fluorescence). Also input any wavelengths, integration time, etc. that are pertinent to your particular assay. Refer to the manual of your particular plate reader for information on running the machine. The output will be one or more files of signals detected in the various wells.
- Edit each data file carefully to make sure that plates are correctly identified before unloading the deck of the plate reader stacker.

▲ **CRITICAL STEP** Although most plate readers have a barcode reader, the barcode of plates loaded onto the plate reader can become illegible due to experimental manipulation.

? TROUBLESHOOTING

(B) Image data ● **TIMING** Depending on the microscope being used, the exposure parameters used and the number of sites and wavelengths being examined per well, the time to complete a plate can vary from 30 min to 4 h or more

- (i) Load the plates onto the staging tray of an automated microscope.

▲ **CRITICAL STEP** Confirm that the microscope will be able to run without interruption for enough time to complete the data collection. Confirm that the plates are loaded in the correct order.

- (ii) Run the automated image collection. Set the parameters for your particular assay regarding wavelength, exposures, magnification, etc. Imaging at a magnification of $\times 20$ will result in the acquisition of a larger image field, and will allow the camera to capture more cells for each site imaged. For a $\times 20$ screen, four to five fields should be sufficient for visualizing a representative sample of cells to be analyzed. The advantage of imaging at a $\times 20$ resolution is the speed of acquisition (given that only four to five sites are imaged per well) and a relatively low computer storage requirement for the complete set of images. However, depending on the assay, $\times 20$ imaging will yield images at too low resolution to be useful. In this case, magnifications of $\times 40$ or $\times 60$ are used. At these higher magnifications, fewer cells will be captured requiring more sites to be imaged in each well. This will result in an increase in total acquisition time, and the larger number of images will add to the storage requirement. A $\times 40$ field is approximately one-fourth the area of a $\times 20$ field; therefore, 16–20 fields may have to be imaged to get the same image area as that of a $\times 20$. High-resolution screens, where three different fluorophores are followed, necessitate approximately 1–2 TB of storage under these conditions.

Time for imaging a plate will depend on the microscope and the sophistication of the imaging protocol. However, we find that eight 384-well plates can be routinely imaged during an overnight session. Most fluorescent signals are stable for at least a week after fixation. However, it is advisable to test the kinetics of signal decay before undertaking a large screen. To avoid any significant drop in signal intensities, make sure to coordinate processing of the plates for imaging with the imaging sessions.

Refer to the manual of your particular microscope for image acquisition. Create a specific folder in a long-term storage system for each screener.

▲ **CRITICAL STEP** High-throughput image-based screens produce large (hundreds of gigabytes to 1–2 TB) quantities of data. We have chosen to use a large RAID disk array for image storage available across our local area network via server message block (Windows file sharing). Transfer image files temporarily stored on the microscope computer to this folder. The time needed for transfer should not be underestimated. The quantity of data being moved will take considerable time, even on the fastest networks. On our 100-Mb network, transferring a full data set can take 5–10 h.

- (iii) Link to the image data in such a way that it will be remotely accessible. In the case of the DRSC, links to the images are added to our database so that screeners may access their images through pages on our website. Whenever possible, images should be analyzed with an automated image analysis program. A number of commercial (e.g., MetaMorph, or Acapella bundled with the Opera microscope) or open-source packages (e.g., CellProfiler⁴²) are available and will generate quantitative metrics corresponding to the phenotype(s) of interest for each image. This is a transformation that greatly facilitates the subsequent analysis of the data. Many microscope platforms come with their own bundled image analysis software. Although these programs can be reliably used for routine analyses, more sophisticated needs may require the development of a specific algorithm to accurately detect a screen-specific phenotype. Often, this will entail collaborating with an imaging specialist. If that is not possible or the expertise is not readily available, screen-specific phenotype(s) can be reviewed by visual examination. While the eye can discern very subtle phenotypes, this approach is extremely time-consuming and suffers from being susceptible to human bias.

? TROUBLESHOOTING

(C) Other data

- (i) Store the data in a format that is flexible enough to allow nonstandard means of data acquisition or further manipulations, but keep the plate and well information firmly coupled with the measurement or processed data.

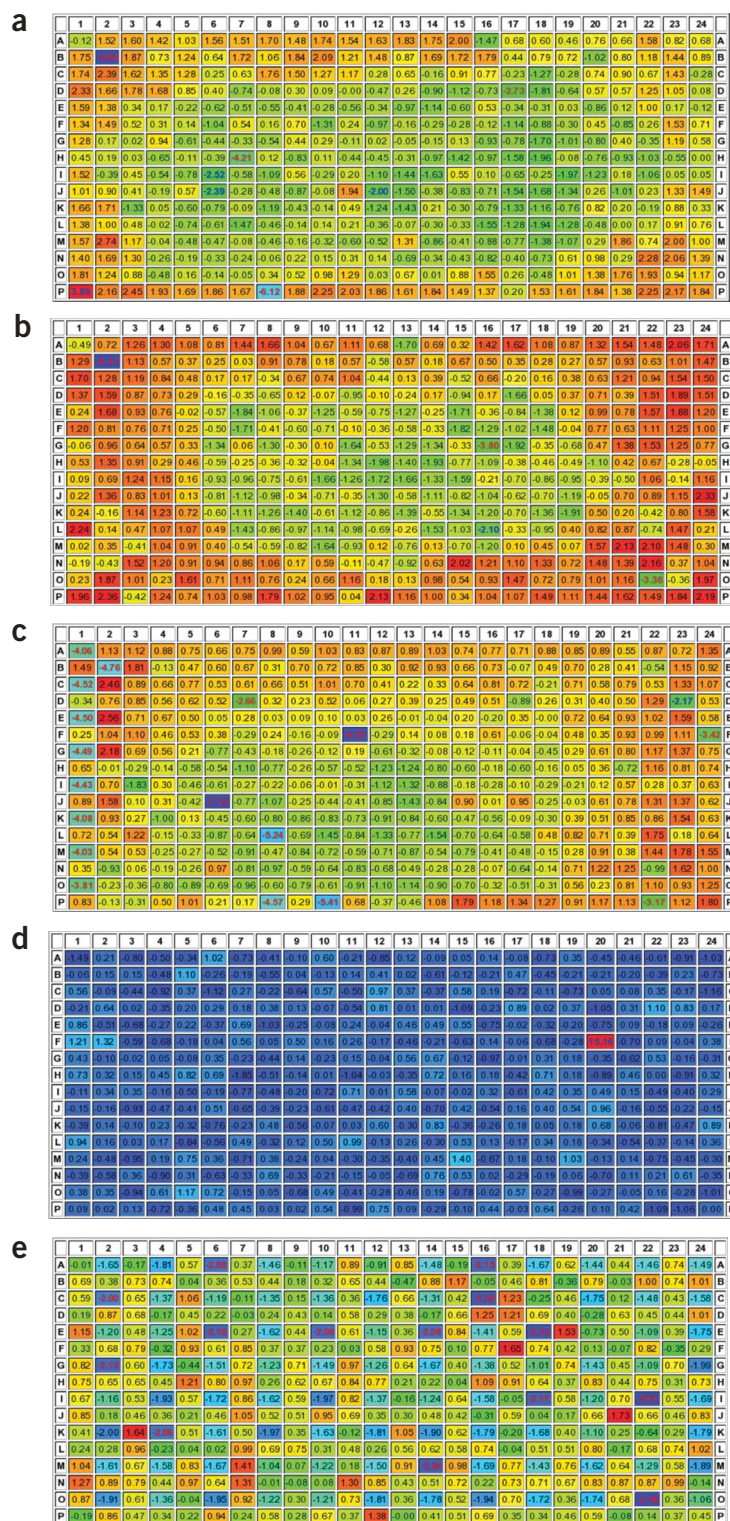
■ **PAUSE POINT** Data acquisition is usually run in batches, with a set of plates being prepared or at mid-point of the screening protocol as another set is being read. The opportunity to pause arises between batches and, to some extent, while a plate reader or automated microscope is running.

16| Use the Heatmap tool (see EQUIPMENT) to visualize numeric data generated for a plate. Refer to ANTICIPATED RESULTS for examples. Enter plate data by browsing for the file to be submitted (information on the correct data format and sample data is available on the Heatmap web page (http://flyRNAi.org/cgi-bin/RNAi_heatmap_public.pl)). Select raw data for the computation format.

■ **PAUSE POINT** Pausing at any point from Step 16 on is possible as these steps describe data analysis tasks with no critical time components beyond the desire to finish the project.

17| Visually inspect the plate-by-plate data to identify experimental artifacts and other problems that will limit the precision of later statistical analysis, such as strong variation between values in different parts of the plate (e.g., edge-effects, **Fig. 6a** and **b**), an off-color row or column (probable pipetting error, **Fig. 6c**), problematic extreme outliers (**Fig. 6d**), any regular pattern (probable aliquotting or source material problem, **Fig. 6e**), etc. (see ANTICIPATED RESULTS).

Figure 6 | The Heatmap tool: Heatmap views of plates showing experimental artifacts. Panel (a) shows a plate with a classic narrow edge-effect—the bottom row of the plate shows an obvious color difference from the rows above it. This is in contrast to what is observed in panel (b), where the distortion is not limited to the outside edge. In this example, well values generally decrease toward the plate center, perhaps as a result of uneven evaporation or plate defects leading to this artifact. Panel (c) shows artifacts caused by a pipetting error of some kind. Column 1 is clearly defective in all of the wells that came from the first of the two 96-well plates that populate that column. Panel (d) illustrates an example where a single, extreme outlier is masking the variation in the rest of the plate (which aside from that outlier shows a normal distribution). Panel (e) illustrates a case where the wells originating from one of the four 96-well plates that served as source material for this 384-well plate is showing consistently lower values, suggesting that the material on that source plate was damaged or aliquoted incorrectly. The color key for all five plates is determined using a relative scale where the highest value on each plate is assigned red, the lowest value is assigned blue and all other well values are colored based on where they lie between the two on a linear scale. Data for these examples are part of a sample set available from the public Heatmap page (http://flyRNAi.org/cgi-bin/RNAi_heatmap_public.pl), and the images were generated from that page. Data used in panels (a) through (c) are taken from Boutros *et al.*¹¹, and data for panels (d) through (e) were generated artificially to clearly illustrate those specific conditions.



20 While selecting a normalization method, also select a cut-off method to identify those wells that vary significantly from the plate mean or median.

▲ CRITICAL STEP Repeating this step multiple times to find a cut-off that accurately predicts known targets while reducing noise is recommended. The proper cut-off for a screen is variable and strongly dependent on the characteristics of the individual assay. Each researcher is encouraged to experiment extensively with different normalization methods and cut-off values.

21 Redisplay the Heatmap data using the chosen normalization method and cut-off. The calculated values may be saved to a file from this page. Wells beyond the cut-off value are highlighted for selection as wells of interest.

Plateset comparison ● TIMING 1–5 d

22 Compare the plate sets. Separate the wells of interest based on which wells scored beyond your cut-off in all cases, which wells scored in some cases but were close to the cut-off in others, and which wells varied widely between sets (see ANTICIPATED RESULTS).

23 Combine the data from the plate sets in a manner that compensates for experimental variation between the two data sets. Plates are affected globally by environmental conditions such as exact temperature and humidity. Duplicate plates (arrayed with identical dsRNAs) run through the identical assay procedure on different days, or even on the same day, will likely have different mean well values and different variance. Quantile normalization of data sets, which is supported at the DRSC, has shown particular promise in resolving this issue and will be described elsewhere (I.F. and B.M.-P., unpublished data).

Postscreen analysis of the results

24 To begin the postscreen analysis of the results, check whether genes that are expected or projected to be identified in a screen are part of the hit list. This allows the sensitivity of the screen to be determined (Fig. 4). For instance, in a screen for receptor tyrosine kinase and extracellular-signal-regulated kinases (RTK/ERK signaling), a rather permissive threshold (Z-score ± 1.5) was chosen to identify hits. The reason for this relatively low threshold was that the authors wanted to maximize the identification of novel regulators before they subjected these hits to various filters to gain knowledge of the systems-level network at play⁴⁴. A screen with a more narrow focus aimed at discovering a missing component (gene discovery) should use a more stringent threshold.

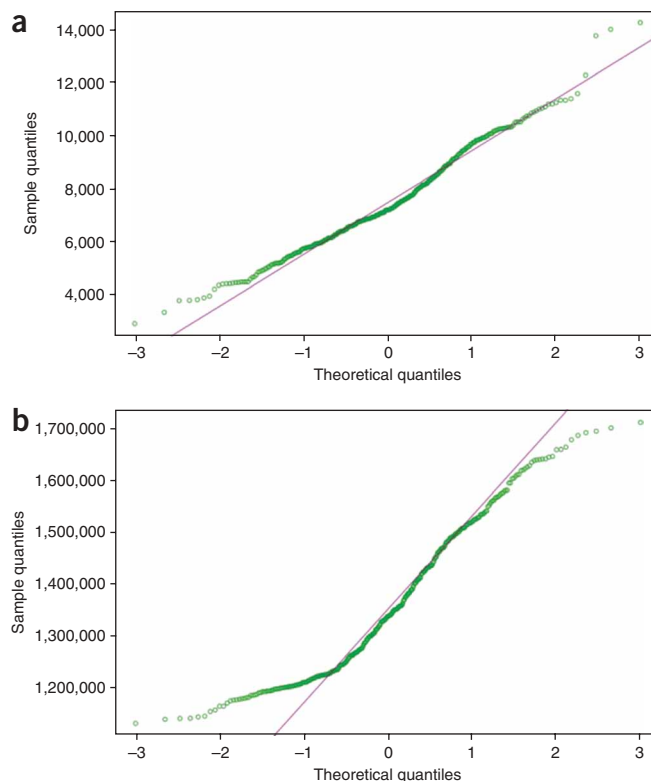


Figure 7 | Quantile-Quantile plots. Panel (a) illustrates an acceptable plot of normal experimental data. The experimental values, when plotted against values from an artificial normal curve, conform roughly to a straight line. A few (generally less than five) deviations at either end, as shown, are acceptable and expected, as both hits and many experimental artifacts (such as a dry well) will produce values away from the normal curve in many experiments. Panel (b) shows a plot of data that is not conforming to a normal curve. The first and fourth quarters of the data deviate from the normal curve in opposite directions. Data for these plots were taken from DasGupta *et al.*⁵² (a) and Boutros *et al.*¹¹ (b).

BOX 4 | NORMALIZATION METHODS

Log[value/average]: This is a simple measure of variation on a logarithmic scale. This is useful when the variation on the plate follows a normal distribution on a logarithmic scale.

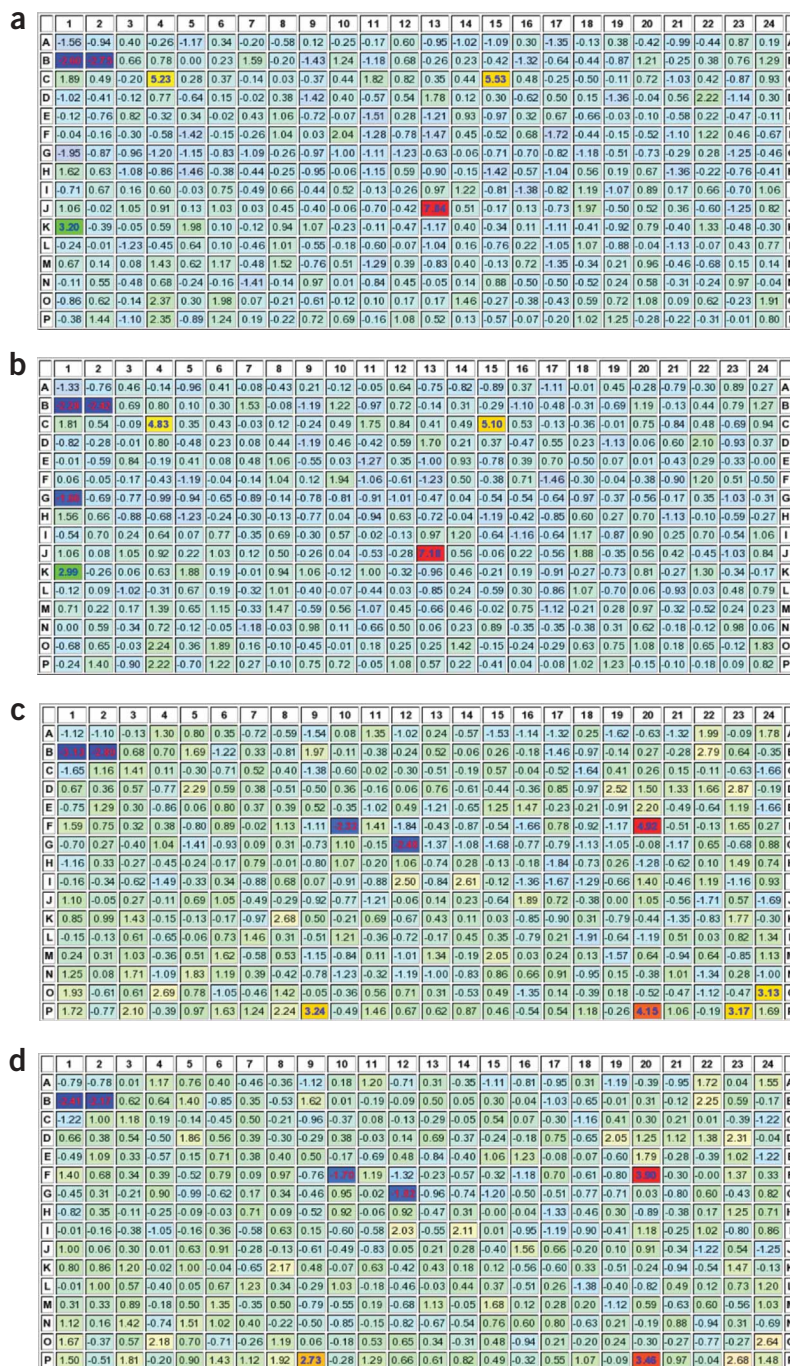
Z-score: This is a measure of distance, in SDs, from the plate mean. A well with a Z-score of 0 has the same value as the plate mean. A well with a Z-score of 1.0 is exactly one SD above the plate mean, and a Z-score of -0.5 is half a SD below the plate mean.

[Well – median]/interquartile range: Also called variation/IQR, this is a measure of distance from a middle value similar to Z-score, but it is based on interquartile range (IQR—the difference between the first and third quartiles; the first quartile is the median of the lower half of the data and the third is the median of the upper half) instead of SD. Variation/IQR is used because it can be less sensitive to the effects of extreme outliers than SD. Note that the median is used for a middle value instead of the mean (see Fig. 8 for a comparison of Z-score versus variation/IQR).

Robust Z-score: This is a transformation of the variation/IQR where the IQR value is multiplied by a constant to bring the numbers into rough equivalence with the corresponding Z-scores.

B-score: The B (Brideau) score is a normalization that accounts for row and column variations (such as edge-effects). It tends to be less precise than other measures in general, but very effective in the presence of row and column effects^{36,37}. B-score normalization is an anticipated feature of the Heatmap page by the time of this publication.

Figure 8 | The Heatmap tool: comparison of two normalization methods. The choice of normalization methods can affect the identification of outliers. Here we see two plates, each normalized two different ways. Panels (a) and (b) illustrate the same plate with the same cut-off setting. In panel (a), the plate is displayed using Z-scores, in panel (b) the same data are displayed using (well value—plate median)/interquartile range (also known as variation/IQR). Note that in panel (b) an additional well (G1) is identified. Conversely, panel (c) shows the Z-score version of a plate, which has two outliers (O24 and P23), which are not identified when the same plate is displayed using Variation/IQR in panel (d). Data for these examples were obtained from the sample set by DasGupta *et al.*⁵² analyzed on the public Heatmap page (http://flyRNAi.org/cgi-bin/RNAi_heatmap_public.pl). The illustrations were generated using the Heatmap tool with an option to strongly highlight outliers and use faded colors for other wells. All four panels are colored using a relative scheme. The highest value on the plate is colored red, the lowest value blue and all other values are colored based on where they lie between the two on a linear scale. To highlight the outliers, outlier wells on each plate are color-coded using the left key (more intense coloring) at the bottom, whereas all other wells are color-coded using the right key. The numeric range of the two keys is identical.



Ultimately, the screener must use his or her own judgment. If the expected genes are included using the chosen threshold or criteria for selection, proceed to Step 25. If not, investigate the possible reasons for this as outlined in **Box 5** before proceeding.

25 | TROUBLESHOOTING

25| Once reasons for absent hits have been resolved to the extent possible, assess the quality control (QC) information for each positive dsRNA (**Fig. 4**) as detailed in **Box 6**.

26 | TROUBLESHOOTING

26| Assess results with dsRNAs that target the same gene. Due to annotation revisions or other historical reasons, many screening libraries are partially redundant in that, in a number of cases, two or more distinct dsRNAs that target the same gene are present in the library. Therefore, for each dsRNA associated with a phenotype, one should check whether any additional dsRNAs targeting the same gene are present in the screened library and, if so, how they behaved in the screen (**Fig. 4**). When such cases arise, concordance in the effects associated with the related dsRNAs will serve to validate the results. In the absence of concordance, the following explanations should be examined (**Fig. 4**): QC issues and potential OT sequences (see Steps 24 and 25) might be responsible; also, the regions of a gene targeted by the various dsRNAs should be carefully examined as they may target different sets of splice forms, a finding that could legitimately yield different results.

27| Assess importance of metabolic and viability hits. Signaling pathways or biological processes of interest that are interrelated during a screen are intimately linked to general cell metabolism. Therefore, a large number of dsRNAs that score in any



BOX 5 | ASSESSING REASONS FOR THE ABSENCE OF EXPECTED HITS

There are many reasons for false negatives. All screens have inherent noise due to experimental variation such as positional effects (edge effects), or small systematic errors (instrumental or operator related). We have reported a 15% error rate in screen results once predicted off-target effects were accounted for^{22,44}. In addition, other possible explanations need to be examined (**Fig. 4**).

(A) Examine whether the absence of phenotypes associated with genes expected to score in the assay is due to quality issues of the double-stranded RNAs (dsRNAs) targeting them (integrity, size, concentration, presence of multiple bands, etc.) (**Fig. 4**). This information is available for most libraries and should be reviewed whenever possible.

(B) If the dsRNAs appear of good quality and of the expected concentration, it may be a sign that the assay is not as sensitive as originally thought. Consider changing the threshold chosen to select hits. Threshold selection is often empirical: dsRNAs targeting genes expected to score as weak hits might result in experimental values that are just above, at or below the threshold depending on the stringency chosen. Reviewing the scores or values obtained for all expected hits might indicate a need to adjust the threshold (**Fig. 4**).

(C) Test additional dsRNAs against the few expected genes that did not score in the primary assay. If again no phenotypes are detected with the new dsRNAs under the assay conditions, consider repeating the screen, but only after further assay optimization using these genes as positive controls (last resort).

▲ **CRITICAL STEP** A balance must be maintained between screen sensitivity and selectivity. A threshold that is set too low will increase the number of weak, spurious or artifactual positives.

given screen will target genes that function in basic cell metabolism and homeostasis (**Fig. 4** and **Table 3**). These are usually legitimate hits under the conditions of the assay, but they may not be pertinent to the biological question under consideration. As such, the level of interest to be given to such hits will vary and the decision to validate all or only a few of them further will depend on the goals of a screen. For instance, only a few select representative members of the hits targeting general metabolic pathways were chosen for additional validation in the RTK/ERK signaling screen⁴⁴. However, it may be judicious to pursue further the role of such genes in a process of interest when only one or two representative genes have been identified, while none of the other genes that are part of the same metabolic machine have yielded a phenotype when knocked down by RNAi.

28| Decide whether to further prune the primary list by concentrating on genes that have mammalian orthologs. Computational resources (such as Homologene⁴⁵ and InParanoid⁴⁶) based on reciprocal BLAST or other similar methodologies^{47,48} are typically used to identify *Drosophila* gene orthologs. One approach to evaluating particularly large sets of hits is to determine whether there is a specific enrichment in the set compared with the whole transcriptome. In the RTK/ERK signaling screen, one criterion for selection for further validation was to be all inclusive of the hits corresponding to the genes known to be involved in RTK signaling as well as for evolutionary conserved genes implicated in human diseases⁴⁴.

29| Confirm results with independent dsRNAs.

▲ **CRITICAL STEP** The finalized hit list needs to be validated by using dsRNAs that are different from those used in the primary screen (**Fig. 4**). At least two independent dsRNAs must show the same phenotype for a gene to be included in the final list in the absence of any other means of validation (biochemical, genetic, etc.)²³.

? TROUBLESHOOTING

30| Assess relevance of hits with information mining. Once hits are validated, the screener can attempt to organize and link their hits by deriving a network of interactions that captures the biological process under study (**Fig. 4**). A thorough discussion of this approach is beyond the scope of this protocol, but excellent starting points are gene orthology and pathway information. Many genes are poorly characterized in flies, but may be better characterized in other organisms. Expression pattern,

BOX 6 | ASSESSMENT OF POSITIVE DOUBLE-STRANDED RNAs

Typical screens identify a large number of hits that, depending on the threshold or criteria chosen, can exceed 1,000, for example, ref. 44. It is therefore important for a screener to assess quickly which of these hits they want to pursue. When using large collections of double-stranded RNAs (dsRNAs), the first priority should be to identify all hits that might represent false positives. For instance, in the RTK/ERK signaling screen, the initial list of hits was first pruned by removing all dsRNAs with greater than ten predicted off-target (OT) sequences (see **Box 1**). This yielded a list of 1,644 dsRNAs targeting 1,168 genes⁴⁴. To minimize the risk of false positives, the first step is to assess the dsRNAs causing a phenotype in a screen.

(A) Check the physical integrity of dsRNAs: discrepancies in size, number or appearance of gel bands corresponding to dsRNAs causing a phenotype should alert a screener to possible artifacts, and additional validation will be required such as retesting with independent dsRNAs as described in Step 29.

(B) Assess off-target sequences.

▲ **CRITICAL STEP** Special emphasis should be placed on whether the dsRNAs might contain sequences with potential off-target effects. Therefore, it is imperative to check for such sequences in each dsRNA associated with a phenotype so that potential false positives can be avoided (**Fig. 4**). For this reason, screeners need to have access to the sequence of each dsRNA so that they (and the public) can make their own independent evaluation (see **Box 1**).

TABLE 3 | Gene ontology terms enriched in screens.

Gene ontology (GO)	Term	Number of dsRNAs in set	Number of hits	Enrichment (1.00 is nonenriched)
GO:0005842	Cytosolic large ribosomal subunit	53	720	9.26
GO:0008540	Proteasome regulatory particle–base subcomplex	12	154	8.75
GO:0005838	Proteasome regulatory particle	23	277	8.21
GO:0005843	Cytosolic small ribosomal subunit	41	491	8.17
GO:0000502	Proteasome complex	22	242	7.50
GO:0004175	Endopeptidase activity	49	510	7.10
GO:0005839	Proteasome core complex	24	238	6.76
GO:0005840	Ribosome	67	659	6.71
GO:0005686	snRNP U2	14	136	6.62
GO:0005665	DNA-directed RNA polymerase II–core complex	11	105	6.51
GO:0050875	Cellular physiological process	56	510	6.21
GO:0000082	G ₁ /S transition of mitotic cell cycle	12	106	6.02
GO:0006510	ATP-dependent proteolysis	20	175	5.97
GO:0000086	G ₂ /M transition of mitotic cell cycle	10	81	5.52
GO:0031202	RNA splicing factor activity—transesterification mechanism	27	210	5.30
GO:0007307	Chorion gene amplification	16	116	4.94
GO:0003735	Structural constituent of ribosome	190	1363	4.89
GO:0045298	Tubulin complex	14	95	4.63
GO:0005852	Eukaryotic translation initiation factor 3 complex	15	101	4.59
GO:0006412	Protein biosynthesis	230	1510	4.48
GO:0005681	Spliceosome complex	77	499	4.42
GO:0006888	Endoplasmic reticulum to Golgi vesicle-mediated transport	14	90	4.38
GO:0007088	Regulation of mitosis	16	98	4.18
GO:0015631	Tubulin binding	16	97	4.13
GO:0000786	Nucleosome	17	101	4.05

This analysis includes dsRNAs from version 1 of the *Drosophila* RNAi Screening Center (DRSC) screening library. Only those dsRNAs that target an annotated gene and have ten or more 19-bp matches to genes other than the intended target are included. The analysis only includes GO terms that have five or more genes with dsRNAs meeting those criteria, and also 13,500 dsRNAs with 19,796 hits are included. 'Number of Hits' indicates the total number of times that dsRNAs targeting genes with the corresponding GO term score in a screen. The enrichment score was calculated by taking the number of hits and dividing it by the expected number of hits, based on the assumption that all dsRNAs are equally likely to score as a hit.

literature-culled information, protein–protein interactions and other experimental or inferential data can be superimposed to the RNAi data set to model an interactive network. There are too many resources to list them all, but a good starting point is Flybase, which provides information on genetic pathways^{49,50} and has link-outs to other databases, such as Wormbase, National Center for Biotechnology Information, PubMed. Another great resource is the IM Browser provided by the Finley lab⁵¹.

● TIMING

Steps 1–6, generation of PCR templates (amplicons) for the synthesis of dsRNAs. Time will vary according to the number of amplicons planned: 1 d for 50, 3 months or more for a genome-wide library

Steps 7–12, synthesis and purification of dsRNAs. Again time will vary according to the number of dsRNAs, 1–2 d for 50 dsRNAs: 2 months or more for a genome-wide library

Step 13, array dsRNA to plates: 30 min for 50 dsRNAs, 1 d for full library (using robotics)

Step 14, RNAi treatment and processing for read-out: 3–6 d or longer

Step 15, data collection: 1–14 d (can be longer with visual screens). Option A: Most plate readers will not take 60+ plates at one time and laboratory conditions usually make preparing all of the plates simultaneously difficult, so usually batches of 10–20 plates are run at a time over a period of several days; Option B: Software-aided automated image analysis will add anywhere from 30 min to several hours per plate, depending on the level of sophistication and the number of images per well to analyze. As with the plate reader (option A), data acquisition and analysis can be done in batches of plates over a period of several days to a few weeks. Note that some phenotypes may not be easily processed by existing commercial or open-source automated imaging software, and will either require the development of custom-made algorithms or involve annotation by eye for each image taken. In either case, this will add significantly to the time necessary to analyze the data.

Steps 16–17, data visualization: 1–2 h (for Heatmap). The Heatmap tool will take approximately 20 min to display the data from a complete pair of plate sets

Step 18, statistical normality: 4–8 h

Steps 19–21, well-of-interest normalization and analysis: 4–24 h

Steps 22–23, plateset comparison: 1–5 d

Steps 24–28, post-screen analysis of hits: 1–4 months

Step 29, hit validation: 2–6 months
Step 30, information mining: 2 months to 1 year

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 4**.

TABLE 4 | Troubleshooting table.

Step	Problem	Possible reason	Solution
6	No PCR product detected	Annealing temperature too high	Use a program with a lower annealing temperature (e.g., 47 °C)
		GC-rich primer sequence	Add 1.5% DMSO into the PCR mix with a 61 °C annealing temperature
	PCR product appears 'smeary' on gel	Too many cycles	Use 1 µl of this PCR product as your template and 10 µM T7 primer in another PCR. Perform another touchdown reaction with only five touchdown cycles and 20 cycles annealing at 55 °C
11	Weak double-stranded RNA (dsRNA) product detected	PCR product used was weak	Add more PCR product to the <i>in vitro</i> transcription reaction
	No dsRNA product detected	Product lost during purification (96-well plate)	Make sure vacuum pressure is not set any higher than 8 psi
14B	Low or inconsistent signal observed after transient transfection of reporter genes	Transfection efficiency or low expression level of DNA construct	Make a stable cell line and screen for high expression
15A	Fluorescence levels are not as high as expected	UV lamp problems	Make sure lamp has been warmed up for at least 20 min before performing measurements
		Fluorophore excitation peak falls outside optimal emission range of UV lamp	Switch to a different fluorophore or dye. Consider using a different type of UV lamp
15B	No image	Wrong filter set may be selected	Make sure your image acquisition settings are correct
	Dim image	Exposure time set too low	Increase the exposure time for particular wavelength
	Image too bright	Exposure time set too high	Decrease exposure time for particular wavelength
	Not many cells in field	Cell loss due to aspiration, washes, etc.	Select fields away from the point of aspiration. Centrifuge plates at 290g for 2 min after any addition and before any aspiration
24	After assessing quality control (QC) and threshold issues, expected hit is still missing	Many possibilities. dsRNA may have unknown QC issues.	Cherry-pick targeting dsRNA and re-test. Or, test with a new dsRNA
25	Unable to determine predicted off-targets	dsRNA sequence is unknown or unavailable	Validate hits with dsRNAs with known sequence
29	Validation with independent dsRNAs give conflicting results	False positives and false negatives can occur	Validate hits with additional independent dsRNAs and replicates

ANTICIPATED RESULTS

Screening and raw data

The successful completion of a high-throughput RNAi screen depends greatly on the care spent during the optimization phase. However, it can be difficult to anticipate issues that might arise from automation and the use of high-throughput equipment that may not be available to a laboratory when the optimization is being done. Fortunately, such problems become apparent very quickly, and usually can be fixed after a few experimental modifications.

Successful screens often identify 2–4% of the *Drosophila* genes tested as potential hits. In some cases, screens have identified between 1,000 and 2,500 initial hits⁴⁴. Such large numbers may reflect the subtlety of a particular read-out or the highly quantitative nature of an assay. Genes that serve a basic cellular or metabolic function (housekeeping genes) will score as hits in a great number of screens (see **Table 3**). In addition, a high number of hits might correlate with the permissiveness

of the assay employed in a screen. For instance, viability screens are highly susceptible to false positives and, in addition, will be affected by experimental artifacts (edge-effects, general health of the culture, etc.) that are not related to the RNAi treatment. The rate of false negatives in a screen can be difficult to measure, particularly if the identity of a number of genes that should score in the assay is not known in advance. In this regard, it is useful to take advantage of redundancy in some libraries (i.e., more than one independent dsRNAs targeting the same gene) to monitor how the different dsRNAs scored in the assay.

Data treatment

During initial data collection, the goal should be to detect and eliminate system-wide or gross errors to produce an initial list of amplicons to analyze further. These include (but are not limited to) aliquoting errors (**Fig. 6c** and **e**), edge-effects (**Fig. 6a** and **b**), loss or mislabeling of results, extreme (usually suspect) outliers that distort the evaluation of other wells on a plate (**Fig. 6d**), plates with little or no variation, data that do not roughly follow a normal curve (**Fig. 7b**) or duplicate experimental runs with little or no correspondence between them. For example, in **Figure 6b**, where we should see a uniform coloration trend across the plate and most wells close to the mean color (light green in this case), we see instead a bias toward positive scores resulting from an extreme negative outlier. In **Figure 6b**, we see clustering of higher values toward the outer edge of the plate, suggesting an issue with edge-effects. Problems that can be identified using the additional statistics provided on the Heatmap page include narrow edge-effects, distortions from normal data or unusually high or low variance. **Figure 7** shows a Q-Q plot of normal data and a plot of data that is distorted from a normal curve. Q-Q plots are a visual measure of normality generated by ordering the values in a data set and plotting them against a like number of values in a statistically normal data set generated by formula. If both data sets are normal, they should form a straight line diagonally across the plot. The more distortion seen in the line, the further the data set deviates from normal. Other problems that are readily apparent in the Heatmap view are periodic patterns that often result from technical problems related to automation steps (uneven cell distribution, defects in dsRNA arraying (four 96-well source plates are used to array the dsRNAs into the 384-well test plate), etc. (see **Fig. 6e**). This type of pattern should alert a screener to re-screen that particular plate.

Post-screen analysis

After the wet lab phase of the screen is done and a first-pass analysis has been performed, a screener will have a large number of positive results to consider. This situation creates the daunting task of processing and prioritizing hits for further validation and analysis. Many screeners fail to anticipate this issue, and underestimate the time involved in the follow-up postscreen analysis and necessary validation phase. Unfortunately, the skills involved in performing the postscreen analysis of a screen are not as familiar to biologists and can be challenging, particularly if one is not trained in bioinformatics or statistics. The failure to recognize the challenge of dealing with large data sets and the reluctance to get involved in this phase of the screen significantly undermine the ability to extract novel information from genome-wide screens. For this reason, we have included the analysis and post-screen analysis steps as an integral part of the protocol.

Many screeners will have a large number of positive dsRNAs to contend with, and that number can be too large to easily validate with independent dsRNAs as described in Step 29 (time- and resource-wise). Therefore, every effort at postscreen analysis (Steps 24–28) is all the more important, as it provides a rationale for prioritizing the positive results for validation and future study.

OTEs have been reported to be a problem in screens with long dsRNAs^{21,22}. However, the full set of rules for off-targets is not understood. Likewise, all possible sources of noise in a screen are not known. For these reasons, screeners need to validate their results using a secondary screen with independent dsRNAs.

The expertise required in sifting through other information databases and making full use of bioinformatics tools to analyze and prioritize screen results is often outside of the traditional training received by a biologist. However, the effort required to acquire these new skills will often pay off, by saving both time and resources in follow-up studies.

ACKNOWLEDGMENTS We thank all members of the Perrimon laboratory as well as all past and present DRSC screeners for sharing their experience with us. None of the protocols described here would have been possible without their contributions. Work at the DRSC is supported by grant R01 GM067761 from the NIGMS. N.P. is an investigator of the Howard Hughes Medical School.

COMPETING INTERESTS STATEMENT The authors declare no competing financial interests.

Published online at <http://www.natureprotocols.com>

Rights and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- Adams, M.D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Hannon, G.J. RNA interference. *Nature* **418**, 244–251 (2002).

- Dykxhoorn, D.M., Novina, C.D. & Sharp, P.A. Killing the messenger: short RNAs that silence gene expression. *Nat. Rev. Mol. Cell Biol.* **4**, 457–467 (2003).
- Tomari, Y. & Zamore, P.D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
- Moffat, J. & Sabatini, D.M. Building mammalian signalling pathways with RNAi screens. *Nat. Rev. Mol. Cell Biol.* **7**, 177–187 (2006).
- Clemens, J.C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl. Acad. Sci. USA* **97**, 6499–6503 (2000).
- Hammond, S.M., Bernstein, E., Beach, D. & Hannon, G.J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
- Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).

9. Kiger, A.A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27 (2003).
10. Foley, E. & O'Farrell, P.H. Functional dissection of an innate immune response by a genome-wide RNAi screen. *PLoS Biol.* **2**, e203 (2004).
11. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
12. Ramet, M., Manfrulli, P., Pearson, A., Mathey-Prevot, B. & Ezekowitz, R.A. Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* **416**, 644–648 (2002).
13. Feske, S. *et al.* A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).
14. Vig, M. *et al.* CRACM1 is a plasma membrane protein essential for store-operated Ca^{2+} entry. *Science* **312**, 1220–1223 (2006).
15. Zhang, S.L. *et al.* Genome-wide RNAi screen of Ca^{2+} influx identifies genes that regulate Ca^{2+} release-activated Ca^{2+} channel activity. *Proc. Natl. Acad. Sci. USA* **103**, 9357–9362 (2006).
16. Perrimon, N. & Mathey-Prevot, B. Applications of high-throughput RNA interference screens to problems in cell and developmental biology. *Genetics* **175**, 7–16 (2007).
17. Kennerdell, J.R. & Carthew, R.W. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
18. Kennerdell, J.R. & Carthew, R.W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol.* **18**, 896–898 (2000).
19. Kim, Y.O., Park, S.J., Balaban, R.S., Nirenberg, M. & Kim, Y. A functional genomic screen for cardiogenic genes using RNA interference in developing *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* **101**, 159–164 (2004).
20. Koizumi, K. *et al.* RNA interference screen to identify genes required for *Drosophila* embryonic nervous system development. *Proc. Natl. Acad. Sci. USA* **104**, 5626–5631 (2007).
21. Ma, Y., Creanga, A., Lum, L. & Beachy, P.A. Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature* **443**, 359–363 (2006).
22. Kulkarni, M.M. *et al.* Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat. Methods* **3**, 833–838 (2006).
23. Echeverri, C.J. *et al.* Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat. Methods* **3**, 777–779 (2006).
24. Perrimon, N. & Mathey-Prevot, B. Matter arising: off targets and genome scale RNAi screens in *Drosophila*. *Fly* **1**, e1–e5 (2007).
25. Armknecht, S. *et al.* High-throughput RNA interference screens in *Drosophila* tissue culture cells. *Methods Enzymol.* **392**, 55–73 (2005).
26. Friedman, A. & Perrimon, N. High-throughput approaches to dissecting MAPK signaling pathways. *Methods* **40**, 262–271 (2006).
27. Kallio, J. *et al.* Functional analysis of immune response genes in *Drosophila* identifies JNK pathway as a regulator of antimicrobial peptide gene expression in S2 cells. *Microbes Infect.* **7**, 811–819 (2005).
28. Björklund, M. *et al.* Identification of pathways regulating cell size and cell-cycle progression by RNAi. *Nature* **439**, 1009–1013 (2006).
29. Boutros, M., Bras, L. & Huber, W. Analysis of cell-based RNAi screens. *Genome Biol.* **7**, R66 (2006).
30. Zhang, X.D. *et al.* Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Pharmacogenomics* **7**, 299–309 (2006).
31. Fuchs, F. & Boutros, M. Cellular phenotyping by RNAi. *Brief Funct. Genomic. Proteomic.* **5**, 52–56 (2006).
32. Mathey-Prevot, B. & Perrimon, N. *Regulatory RNAs: Cold Spring Harbor Symposia on Quantitative Biology Vol. LXXI.* (eds. Stillman B. & Stewart D.) (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007).
33. Cully, M.J. & Leivers, S.J. RNA interference pinpoints regulators of cell size and the cell cycle. *Genome Biol.* **7**, 219 (2006).
34. Echeverri, C.J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat. Rev. Genet.* **7**, 373–384 (2006).
35. Hild, M. *et al.* An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5**, R3 (2003).
36. Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J. & Nadon, R. Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.* **24**, 167–175 (2006).
37. Makarenkov, V. *et al.* An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics* **23**, 1648–1657 (2007).
38. Gentleman, R., Carey, V.J., Huber, W., Irizarri, R. & Dudoit, S. *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Springer, New York, 2005).
39. Sims, D., Bursteinas, B., Gao, Q., Zvelebil, M. & Baum, B. FLIGHT: database and tools for the integration and cross-correlation of large-scale RNAi phenotypic datasets. *Nucleic Acids Res.* **34**, D479–D483 (2006).
40. Flockhart, I. *et al.* FlyRNAi: the *Drosophila* RNAi screening center database. *Nucleic Acids Res.* **34**, D489–D494 (2006).
41. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
42. Carpenter, A.E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
43. Bera, A.K. & Jarque, C.M. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.* **6**, 255–259 (1980).
44. Friedman, A. & Perrimon, N. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. *Nature* **444**, 230–234 (2006).
45. Wheeler, D.L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **35**, D5–D12 (2007).
46. O'Brien, K.P., Remm, M. & Sonnhammer, E.L. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* **33**, D476–D480 (2005).
47. Hirsh, A.E. & Fraser, H.B. Protein dispensability and rate of evolution. *Nature* **411**, 1046–1049 (2001).
48. Jordan, I.K., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* **12**, 962–968 (2002).
49. Crosby, M.A., Goodman, J.L., Strelets, V.B., Zhang, P. & Gelbart, W.M. FlyBase: genomes by the dozen. *Nucleic Acids Res.* **35**, D486–D491 (2007).
50. Grumbling, G. & Strelets, V. FlyBase: anatomical data, images and queries. *Nucleic Acids Res.* **34**, D484–488 (2006).
51. Pacifico, S. *et al.* A database and tool, IM Browser, for exploring and integrating emerging gene and protein interaction data for *Drosophila*. *BMC Bioinformatics* **7**, 195 (2006).
52. DasGupta, R., Kaykas, A., Moon, R.T. & Perrimon, N. Functional genomic analysis of the Wnt-wingless signaling pathway. *Science* **308**, 826–833 (2005).
53. Schneider, I. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* **27**, 353–365 (1972).
54. Yanagawa, S., Lee, J.S. & Ishimoto, A. Identification and characterization of a novel line of *Drosophila* Schneider S2 cells that respond to wingless signaling. *J. Biol. Chem.* **273**, 32353–32359 (1998).
55. Echalié, G. & Ohanessian, A. [Isolation, in tissue culture, of *Drosophila melanogaster* cell lines]. *C. R. Acad. Sci. Hebd. Seances. Acad. Sci. D* **268**, 1771–1773 (1969).
56. Haars, R., Zentgraf, H., Gateff, E. & Bautz, F.A. Evidence for endogenous reovirus-like particles in a tissue culture cell line from *Drosophila melanogaster*. *Virology* **101**, 124–130 (1980).
57. Ui, K. *et al.* Newly established cell lines from *Drosophila* larval CNS express neural specific characteristics. *In Vitro Cell. Dev. Biol. Anim.* **30A**, 209–216 (1994).
58. Ui, K., Ueda, R. & Miyake, T. Cell lines from imaginal discs of *Drosophila melanogaster*. *In Vitro Cell. Dev. Biol.* **23**, 707–711 (1987).