# *Drosophila* Genome-wide RNAi Screens: Are They Delivering the Promise?

B. MATHEY-PREVOT AND N. PERRIMON
*Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115*

The emergence of RNA interference (RNAi) on the heels of the successful completion of the *Drosophila* genome project was seen by many as the ace in functional genomics: Its application would quickly assign a function to all genes in this organism and help delineate the complex web of interactions or networks linking them at the systemic level. A few years wiser and a number of genome-wide *Drosophila* RNAi screens later, we reflect on the state of high-throughput RNAi screens in *Drosophila* and ask whether the initial promise was fulfilled. We review the impact that this approach has had in the field of *Drosophila* research and chart out strategies to extract maximal benefit from the application of RNAi to gene discovery and pursuit of systems biology.

The completion of the *Drosophila* genome sequence in 2000 (Adams et al. 2000) has conceptually changed the approach to functional genomics as it provided the opportunity to develop genome-wide approaches to systematically explore gene functions. Newly emerged technologies were quickly put to task to extract maximal information encrypted in the raw sequence of the *Drosophila* genome. This is best illustrated with RNA interference (RNAi), which is based on the ability of double-stranded (dsRNA), small interfering RNAs (siRNAs), or small hairpin RNAs (shRNAs) to silence a target gene through the specific destruction of that gene's mRNA (for review, see Friedman and Perrimon 2004).

In the past few years, *Drosophila* has become a premier system for systematic genome-wide cell-based RNAi high-throughput screens (RNAi HTS), largely because of two major advances. First, Clemens et al. (2000) made the seminal discovery that long dsRNAs added to the medium of *Drosophila* tissue culture cells are rapidly taken up by the cells and cause efficient knockdown of their targeted mRNAs, thus opening up the application of RNAi to cell-based assays. Second, the development of cell-based assays in *Drosophila* to a high-throughput format (Armknecht et al. 2005) coupled with the production of comprehensive *Drosophila* dsRNA libraries allowed near or full genome-scale screens to systematically interrogate the function of all genes predicted from genomic sequencing (Kiger et al. 2003; Lum et al. 2003; Boutros et al. 2004; Foley and O'Farrell 2004).

In just 3 years, several large-scale RNAi screens in *Drosophila* have been published, and the results obtained from these studies allow us to reflect on the impact that this approach has had in the field of *Drosophila* research and chart out strategies to extract maximal benefit from the application of genome-scale RNAi screens. Here, we discuss whether RNAi HTS are (1) succeeding as a functional gene discovery platform, that is, whether they allow a rapid and unbiased identification of genes involved in a specific biological processes, even in the case of pleiotropic and redundant genes, and (2) allowing us to obtain a systems biology or global picture of the functions of all genes in a given process. Because of the success of *Drosophila* genetics over the years at identifying gene functions, arguably the most interesting application of RNAi HTS is to use this nascent technology to obtain a global understanding of biological processes. In particular, RNAi HTS can be used to gain insights into the structure of signaling networks. For example, as many assays can be designed to quantitatively read the activity of pathways, e.g., using transcriptional reporters or phospho-specific antibodies, the respective contribution of every gene in the genome can be measured and used to model, in combination with other data sets, the flow of information through protein networks (Sachs et al. 2005).

## RNAi HTS: THE BASICS OF THE METHODOLOGY

The RNAi HTS platform is extremely flexible and can accommodate versatile formats. Genome-scale dsRNA libraries (for a list of available *Drosophila* dsRNA libraries, see Echeverri and Perrimon 2006) can be screened in 48-(Ramet et al. 2002), 96-(Lum et al. 2003; Björklund et al. 2006), or 384-well plates in a variety of cell-based assays (Boutros et al. 2004; Agaisse et al. 2005; Baeg et al. 2005; Cherry et al. 2005; DasGupta et al. 2005; Muller et al. 2005; Nybakken et al. 2005; Philips et al. 2005; Bard et al. 2006; Gwack et al. 2006; Vig et al. 2006; Zhang et al. 2006). In addition, dsRNAs can be spotted at high density on glass slides (RNAi microarrays) and assayed in visual screens (Wheeler et al. 2004; Guertin et al. 2006), achieving even faster and cheaper means of screening large libraries (Fig. 1). Detection of phenotypes in high-throughput often relies on the use of plate reader or conventional microscopy (Armknecht et al. 2005), but it can also be based on flow cytometry (Ramet et al. 2002; Björklund et al. 2006), automated fluorometric imaging plate reader (FLIPR) (Vig et al. 2006), or high-throughput confocal microscopy (Pelkmans et al. 2005).

At the *Drosophila* RNAi Screening Center (DRSC), which we established a few years ago (see http://flyrnai/.org), screens are conducted in high-density 384-well tissue culture plates. Existing *Drosophila* cell culture

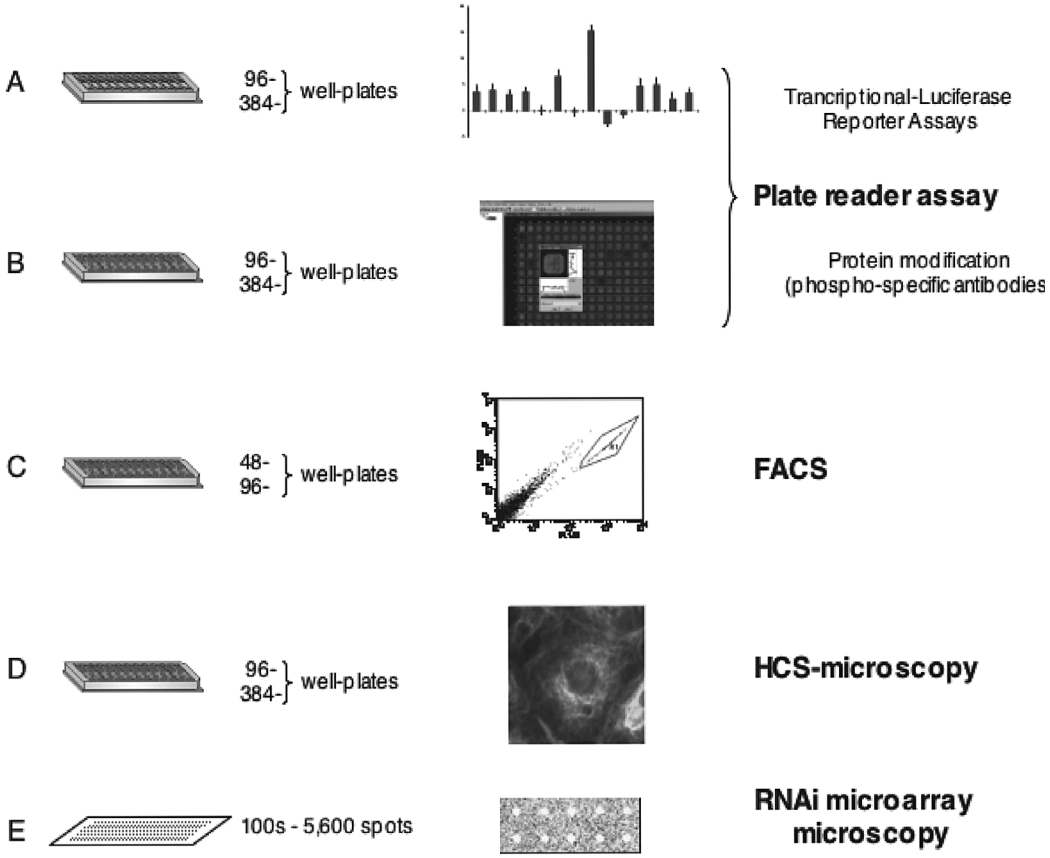


**Figure 1.** RNAi screening format and detection platforms. Listed are the most common cell-based assay formats and detection platforms used in *Drosophila* high-throughput RNAi screens. Genome-scale dsRNA libraries are typically prearrayed on 48-, 96-, or 384-well plates (*A–D*) or glass slides (*E*), and phenotypes are analyzed using various detection methods. (*A*) Conventional plate reader measurements include luminescence, fluorescence intensity, fluorescence polarization, time-resolved fluorescence, and absorbance detection modes. (*B*) The Aerius platform (LI-COR Biosciences) is a modified laser-based microscope that excites and scans in the far-red the emission from appropriately conjugated secondary antibodies bound to the primary, phospho-specific antibody. (*C*) Detection of phenotypes by flow cytometry using a fluorescence activated cell sorter (FACS). (*D*) Detection of phenotypes in high-content screening (HCS) microscopy approaches typically include high-throughput, automated wide-field, and, more recently, confocal microscopes. (*E*) dsRNAs microarrays spotted on glass slides seeded with cells are screened with automated microscopy. Screen examples carried out according to each format are given in the text.

lines (having distinct attributes that can be exploited in the various screens) are ideally suited for this approach; furthermore, the range of uses for these cells can be extended by either transient or stable transfection of DNA constructs before screening. In addition, we have developed and successfully implemented efficient protocols to conduct RNAi screens in primary embryonic cells, broadening the range of biological and developmental processes (e.g., neurite outgrowth and myofibrillogenesis) that can be investigated by this approach (J. Bai et al., in prep.).

The basic experimental design for screens carried out at the DRSC involves three major steps: (1) Gene-specific dsRNAs from our collection stored in 96-well plates are arrayed into 384-well assay plates using robotics. (2) Cells are uniformly and rapidly dispensed into the 384-well plates using a MultiDrop liquid dispenser. (3) After the appropriate incubation time, cells are subjected to individual treatments in a highly parallel fashion, fixed, or directly processed for the assay readout. The phenotypic output measured for each sample depends on the assay readout: Quantitative measurements have been acquired with a plate reader, whereas qualitative measurements have been captured by automated microscopy. The incubation period with dsRNAs varies and must be optimized for specific assays/targets. In general, we have used a 3-day incubation period in our experiments. However, many RNAi effects can be detected within a day or two of treatment. If necessary, incubation with dsRNAs can be carried out for longer (up to 1 week) without deleterious cytotoxicity.

Perhaps the most important aspect of an RNAi HTS platform is the quality of the "RNAi library" to be screened. One of the primary goals at the DRSC has been to ensure that the library of dsRNAs screened is of the highest quality. Thus, over the years, a number of "upgrades" have been made based on our experience with the reagents. Our first library, "*the DRSC 10 collection*" (Boutros et al. 2002), was based on earlier annotations from BDGP/Celera and the Sanger Center, which predicted 13,672 and 20,622 genes, respectively (Adams et al. 2000; Hild et al. 2003). A direct

comparison of the two predictions yielded a total of 21,306 nonredundant possible transcripts in the *Drosophila* genome, with 14,556 dsRNAs targeting annotations present in both the BDGP and Sanger Center sets and 6,750 dsRNAs targeting Sanger annotations not found in the BDGP set (Sanger-only dsRNAs). Subsequent independent analyses of whether the Sanger-only predictions represented real genes or were expressed revealed that only 10% were likely to be validated as genes containing introns (Yandell et al. 2005) and only 291 predictions were confirmed by expression by Stolc et al. (2004). Later releases of the BDGP genome annotation caused a number of revisions, including the prediction of new genes or new exons within a gene, as well as the reassignment of adjacent open reading frames (ORFs) into a single functional unit. These periodic revisions led us to update our library accordingly and to remap some of the older dsRNA to new functional units.

In addition to the upgrade due to changes in genes' annotations, the DRSC library has been updated to address the issue of off-target effects (OTEs) associated with the use of long dsRNAs. This issue emerged as an unanticipated complication in the analysis of large-scale *Drosophila* screens, likely causing the inclusion of a number of false positives among hits reported in early screens (M. Kulkarni et al. 2006). Although this issue is familiar to investigators using siRNAs in mammalian systems (where it had been recognized early on; Jackson et al. 2003), it took the *Drosophila* community by surprise, as it was widely believed that OTE was unlikely to take place in *Drosophila* because of the use of long dsRNAs. To many, the processing of long dsRNAs by Dicer into many short 21- to 23-nucleotide triggers (Hammond et al. 2000) meant that any OTE potentially associated with a particular siRNA trigger from the pool would be diluted by the specific effects of the "good" siRNAs present in excess in that pool. Although that may be true in some cases, a retrospective analysis done by the DRSC challenged this assumption. We performed a statistical analysis of the results from more than 30 DRSC genome-wide screens and asked how the various dsRNAs behaved across these screens. In particular, we examined whether dsRNA predicted to have regions of perfect homology with genes other than the intended target (using a simple string search of all possible siRNAs generated from a dsRNA against all gene sequences in *Drosophila*) led to a greater probability (than by chance alone) of causing a phenotype. In other words, there was a clear correlation between the presence of predicted off-targets in dsRNAs and their likelihood to cause a phenotype in a cell-based assay (M. Kulkarni et al. 2006). Importantly, the homology length at or above which it became problematic was 19 nucleotides rather than the initially predicted cutoff of 21 nucleotides, thus increasing the number of potentially "problematic" dsRNA reagents in our collection.

As a result of this analysis, we assembled a new dsRNA collection (the "*DRSC 2.0 collection*") to eliminate any dsRNAs predicted to have potential OTE. This was achieved by keeping all original dsRNAs from DRSC 1.0 that lacked predicted off-targets and generating 7,692 new, independently synthesized dsRNAs to replace DRSC 1.0 dsRNAs predicted to have 1 or more off-targets. In addition, Sanger-only dsRNAs that were not detected by Stolc et al. (2004) are not represented in the DRSC 2.0 collection. As we are still learning the rules for what might constitute an offending sequence associated with OTEs, we decided to provide screeners with the ability to quickly validate the effects of dsRNAs identified in a primary screen with a second or third independent dsRNA, even if the original dsRNA had no 19-nucleotide perfect homology with any non-target genes. With this objective in mind, we are currently assembling the "*DRSC 3.0 collection*," which consists of new dsRNAs that are distinct from any dsRNA present in DRSC 2.0. These dsRNAs are devoid of predicted perfect homologies with non-target genes and correspond to every hit identified in a completed DRSC screen.

## RNAi AS A FUNCTIONAL GENE DISCOVERY PLATFORM

In addition to the quality of the RNAi library, the success of RNAi HTS depends on the robustness of the cell-based assay and its applicability to high-throughput screening. Many considerations, such as signal-to-noise issues, normalization methods, choice of cell type, and specificity of the readout, should be taken into account when designing an assay. This thorough assessment is probably the most important step of RNAi HTS (for more details, see review by Echeverri and Perrimon 2006). In the context of gene discovery, a screen that leads to hundreds of positives may be considered less successful than a screen that identifies a smaller number of candidates. As such, the design of an assay should be aimed as much as possible at capturing specific features inherent to the process under study to limit the number of positives. For example, a screen for cell viability is expected to lead to hundreds of hits, whereas a screen for subcellular localization of a protein may only lead to the identification of a few hits. Altogether, assay development today is probably the most important step of the RNAi HTS area and where many sophisticated innovations will occur. Below, we describe a few screens that have been done today as a means to document the currently available technologies (Fig. 1).

### Transcriptional Reporter Screens

Many assays are based on transcriptional reporters whose overall chemiluminescence or fluorescence output is rapidly measured using a plate reader (see, e.g., Boutros et al. 2004; Baeg et al. 2005; DasGupta et al. 2005; Nybakken et al. 2005; Bard et al. 2006). The generation of numerical readouts for each condition or well tested makes it possible to normalize the data and subject it to various statistical analyses. This approach was used to investigate, for example, the evolutionarily conserved Wnt/Wg signaling pathway, which regulates many aspects of metazoan development. A cell-based assay based on the "TOP-Flash" (Tcf Optimal Promoter) reporter construct was developed and optimized for high-throughput conditions in 384-well plates (DasGupta et al. 2005). The TOP-Flash construct consists of multimerized Tcf-binding sites cloned upstream of a cDNA encoding firefly luciferase.

Transfection of this construct, a Renilla luciferase normalization vector, and dsRNA into *Drosophila* cells in the presence or absence of Wg serves as the basis of the assay. A normalized readout of luciferase expression is measured under every experimental condition, with its value being directly proportional to the extent of pathway activation. The screen was performed in duplicate, and 238 hits were identified that either reduced Wg pathway activity by more than 1.5 SD or increased reporter activity by more than 3 SDs. Importantly, more than 16 of the known regulators of the Wg pathway scored in this assay, including Armadillo, Pangolin, Legless, Pygopus, Axin, CK1α, Frizzled, and Arrow. The positive and negative regulators were then systematically ordered in the pathway by several epistasis experiments to ascertain at which step in the signal transduction cascade the candidate genes have a potential function. The hits comprise many genes assignable to certain molecular complexes or biological functions, and include (1) HMG/homeodomain transcription factors, (2) kinases and phosphatases, (3) proteosomal components and ubiquitin ligases, (4) G-protein family, and (5) membrane-associated proteins. Of specific interest are some of the kinases and phosphatases, such as Cdc2 and String (Cdc25). Both have been shown previously to genetically interact with Armadillo, thus implicating them as having some role in the Wg pathway. However, their mechanisms of action in the regulation of the Wg pathway are unknown.

### Antibody-based Screens

Another powerful application of antibody-based screens involves the use of phospho-specific antibodies. Provided the specificity of the phospho-antibodies is good, such screens are highly quantitative and can be performed using either a plate reader to measure overall levels of fluorescence emitted by the fluorescently coupled secondary antibody or with the Aerius platform (LI-COR Biosciences), a modified laser-based microscope that excites and scans in the far-red the emission from appropriately conjugated secondary antibodies bound to the primary, phospho-specific antibody (Fig. 1). For instance, the cellular network responsible for mitogen-activated protein kinase (MAPK) activation was investigated using a fluorescently conjugated antibody (cell signaling) that recognizes the diphosphorylated (activated) form of the single *Drosophila* ERK, Rolled. ERK activity was monitored by dpERK staining, at baseline (resting) and under stimulation by Insulin, in an SL2-derived cell line that was engineered to express yellow fluorescent protein (YFP)-tagged Rolled, as a means to normalize for total ERK protein levels using YFP fluorescence. Importantly, the kinetics of MAPK activation and the effects of known component knockdown were found to be identical in wild-type and Rolled-YFP-expressing cells; 1168 unique dsRNAs were found in the primary screen to significantly affect the level of ERK phosphorylation. Although this unbiased list was not fully validated, it included the entire *Drosophila* core pathway, and more than 60% of the candidates had identifiable human orthologs. Various criteria ranging from GO annotation consideration and evolutionary conservation were applied to filter the initial list down to 362 candidates, which were tested in secondary screens in different cell lines and under various ligand stimulations. Of those, 331 genes were validated, and greater than 85% of those could be confirmed with the use of a second or third independent dsRNA (A. Friedman and N. Perrimon, in prep.). In addition to identifying new regulators (see Table 1), data from this quantitative, unbiased approach can be integrated with other genomic and proteomic approaches (as outlined in Fig. 2) to provide a blueprint of the complex regulatory network leading to MAPK activation.

**Table 1.** *Drosophila* RNAi Genome-scale Screens and Gene Discovery: Selected Examples

| Screen | Gene/putative function | References |
|---|---|---|
| Wg signaling | Evi/Wg secretion, positive regulator | Bartscherer et al. (2006) |
| Store-operated $Ca^{2+}$ entry | Orai1/CRACM1/Olf186-F; modulator of CRAC-mediated current | Feske et al. (2006); Vig et al. (2006); Zhang et al. (2006) |
| JAK/STAT signaling | PTP61F/protein tyrosine phosphatase, negative regulator | Baeg et al. (2005); Muller et al. (2005) |
| Host factors involved in *Mycobacteria* infection | CD36/class-B scavenger receptor required for uptake of mycobacteria | Philips et al. (2005) |
| Hh signaling | Ihog/type I membrane protein binds active Hh protein and mediates its signaling | Yao et al. (2006) |
| Hh signaling | PP2A/multimeric protein phosphatase 2A, negative regulator | Nybakken et al. (2005) |
| EGFR signaling | PLC-γ/required for ER retention of cleaved Spitz during fly eye development | Schlesinger et al. (2004) |
| Cytokinesis inhibitors | Borr/protein involved in Aurora-B kinase pathway | Eggert et al. (2004) |
| MAPK pathway | dGCKIII/member of Ste20 kinase family, positive regulator | A. Friedman and N. Perrimon (in prep.) |
| Myofibrillogenesis | Sals/actin-binding protein regulating proper sarcomere length | J. Bai and N. Perrimon (in prep.) |

This list is not meant to be exhaustive but rather highlights selective examples of genes found to participate in particular signaling pathways or biological processes investigated with unbiased RNAi screens. In addition, all screens (cited throughout the text) have identified major molecular machines including the ribosomal complex, the protein degradation machinery (proteasome, ubiquitination), the vesicular and nuclear transport machinery, and the RNA processing machinery.

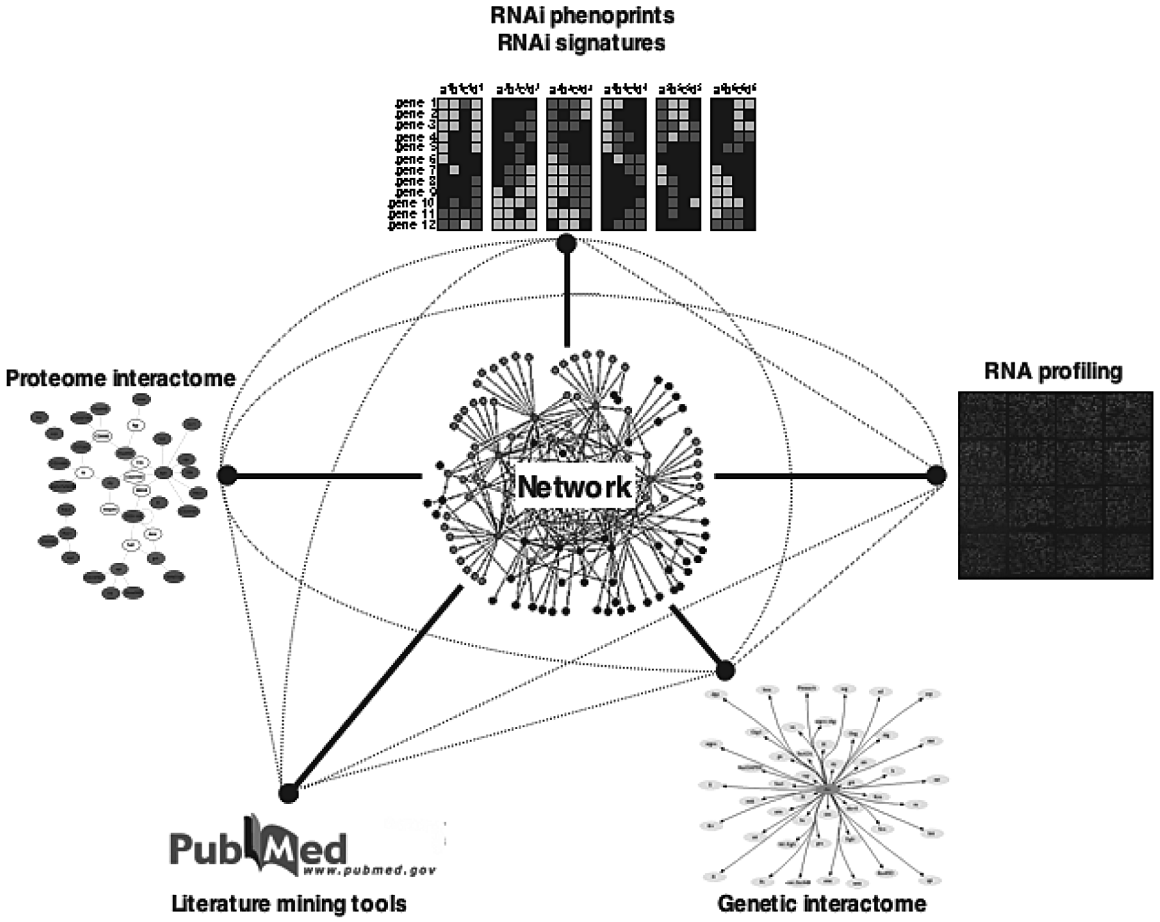


**Figure 2.** Network building through the integration of RNAi signatures with other data sets. RNAi HTS in *Drosophila* cells can be used to assign phenotypic signatures (referred to as phenoprints) to every *Drosophila* gene. Phenoprints can then be used to cluster genes that are functionally related, guiding functional genomics efforts to assign a biological function to uncharacterized or unknown genes based on where they cluster in RNAi screens. The RNAi signatures/phenoprints panel illustrates how 12 genes (1–12) can be functionally clustered based on the comparison of distinct phenotypes ($a^n$–$d^n$) scored in hypothetical screens (1–6). Furthermore, global correlations between phenoprints, transcriptional profiling (RNAi profiling), interactome data sets (proteome and genetic interactomes), and published literature (literature-mining tools) can be used to derive network graphs critical to data mining.

### FACS-based Screens

Several large-scale RNAi screens have relied on the use of flow cytometry to follow phenotypes (Ramet et al. 2002; Björklund et al. 2006; Ulvila et al. 2006). Coupled with high-resolution imaging microscopy, this approach can be very powerful and allows multiple parameters to be analyzed simultaneously. Using this strategy, Björklund et al. (2006) set out to identify pathways regulating cell size and cell cycle progression in *Drosophila* S2 cells. RNAi-treated cells were simultaneously analyzed by fluorescence-activated cell sorter (FACS) for six distinct phenotypes ($G_1$ arrest, $G_2$/M transition, cell death, cytokinesis, and cell size in $G_1$ or $G_2$). This multiparameter analysis identified 488 candidate genes, whose gene products include cell cycle regulators, members of the ubiquitin pathway, components of vesicular and nuclear transport, and mediators of four ligand-induced signaling pathways (Wnt/Wg, p38βMAPK, FRAP/TOR, and JAK/STAT) (Björklund et al. 2006). Although profiling DNA content by FACS analysis has proved to be valuable and informative, one slight limitation of this approach is that many *Drosophila* cell lines do not have a normal mode of chromosomes and tend to be polyploid.

### Microscopy-based Screens

Arguably, the most informative cell-based assays are microscopy-based as they provide a wealth of information since specific information can be scored. Image-based screens can use not only an antibody staining, but also cellular compartments or structures (e.g., Golgi, mitochondria, nuclei, and actin filaments) that can be selectively labeled with either fluorescently labeled dyes or GFPs (green fluorescent protein) tagged with the appropriate localization tag (Kiger et al. 2003; Gwack et al. 2006). A powerful extension of this approach is exemplified in host/pathogens screens where GFP is constitutively expressed when the pathogen localizes to the endosomal compartment of a cell. This approach can be used to identify host factors highjacked by *Mycobacterium* for survival within macrophages as well as host factors involved

in pathogen killing. Philips et al. (2005) developed an ex vivo assay of infection using S2 cells and GFP-expressing mycobacteria and showed that the nonpathogenic species, *M. smegmatis*, is killed by S2 cells, whereas the pathogenic species, *M. fortuitum*, is able to grow within S2 cells. During infection of S2 cells, expression of the macrophage-activated promoters *map24* and *map49* is induced, revealing that the intracellular milieu within S2 cells is similar to that within mammalian macrophages. This cell-based assay was used in a genome-wide RNAi screen to identify host factors required for entry or growth of *M. fortuitum* within S2 cells; 30 dsRNAs were reported to allow *M. fortuitum* to grow better within S2 cells, whereas 86 dsRNAs were found to disrupt infection by *M. fortuitum*. Most of the dsRNAs that disrupt infection target genes predicted to have a role in vesicular trafficking, the actin cytoskeleton, ubiquitin or proteolysis, or fatty acid metabolism. Many of these factors also disrupt phagocytosis of other pathogens, such as *Escherichia coli*. However, some appear to be uniquely required for mycobacterial infection. In particular, a member of the CD36 family of scavenger receptors was found to be required for uptake of two mycobacterial species and *Listeria monocytogenes*, but dispensable for uptake of *E. coli* and *Staphylococcus aureus*, suggesting a role in pattern recognition of distinct bacterial species (Table 1).

### RNAi Microarrays

In this innovative approach, 2–3 nl of dsRNA from large dsRNA collections is arrayed at high density (upward of 5000 spots per slide) on coated glass slides. Cells are seeded onto the slide and incubated for a few days before RNAi phenotypes are scored by microscopy over the clusters of cells that have landed on each dsRNA spot. In this manner, thousands of dsRNAs can be quickly and economically screened (Wheeler et al. 2005). To demonstrate its feasibility, Wheeler et al. (2004) printed arrays with 384 preselected dsRNAs (at a density suitable to array 5600 dsRNAs on a single slide) and screened for regulation of cell number and viability; 44 dsRNAs in this set showed a reproducible phenotype. In addition to confirming the function of cell cycle regulators and apoptotic modulators, targeted by dsRNAs present in the set, these authors identified two surface receptors (InR and Pvr), as well as a number of kinases and phosphatases that affected cell number and viability, underscoring the promise of this approach for systematic synthetic lethal screens using combinations of dsRNA treatments (Wheeler et al. 2004).

## ARE GENOME-WIDE RNAi SCREENS DELIVERING THE PROMISE?

### RNAi HTS as a Functional Gene Discovery Platform

Clearly, as exemplified by the screens described above and others, RNAi HTS are being highly successful at gene discovery (Table 1). The success of the approach relies on an overall low rate of false negatives. Indeed, as shown in Table 2, most expected components that should score in a screen are identified in top hits. Thus, the approach is very robust at generating an enriched list of genes that are likely to be specific for a process. Follow-up experiments are then required to further validate the genes identified.

In the upcoming years, many exciting advances will take place, both in the sophistication of cell-based assay designs and in the detection of phenotypes. Assays that are more biologically relevant, such as those that use primary cells, will become favored. Either antibody-, FISH-, or Luminex-based assays relying on multiplexed endogenous readouts (Levsky et al. 2002; Pelech 2004; Sachs et al. 2005) will be preferred to the current luciferase transcriptional reporter-based assays since they capture richer information. Similarly, high-content microscopy screens that extract and quantify multiple features from each image and are carried out in different established or primary cell cultures will greatly expand our ability to probe for complex cell biological processes. In addition, improvements in image acquisition together with the development of novel molecular probes will alleviate our current limitation in addressing questions of spatial and temporal regulation of signaling pathways and cytoskeletal organization. Parallel RNAi and small-molecule screens comparing small-compound- and dsRNA-induced phenotypes will offer a powerful venue for drug target discovery (Eggert et al. 2004). Combinatorial RNAi experiments where collections of dsRNAs are screened for their ability to suppress or enhance the phenotype caused by another dsRNA or small molecules will become common, and we anticipate the need for such screens to grow up exponentially. A prevalent illustration for this kind of application will be the search for synthetic lethal phenotypes, an approach that will necessitate a very large number of experiments. Although these screens can be conducted in the 384-well plate format, the miniaturization and economy of cells and reagents intrinsic to RNAi cell microarrays (Wheeler et al. 2005) offer an ideal solution to this challenge. Other potential uses of RNAi microarrays include suppression (or enhancement) of small-compound-induced phenotype(s) as recently exploited by Guertin et al. (2006), who com-

**Table 2.** Low False-Negative Rate in Genome-wide RNAi Screens

| | | Canonical genes | | |
|---|---|---|---|---|
| Pathway | Reference | expected | found | Scoring efficiency (%) |
| Wg | DasGupta et al. (2005) | 17 | 15 | 88 |
| Hh | Nybakken et al. (2005) | 15 | 12 | 80 |
| JAK/STAT | Baeg et al. (2005) | 6 | 5 | 83 |

Data reported in three published screens investigating the Wnt/Wg, Hh, and JAK/STAT pathways were used to estimate the rate of false-negative associated with genome-wide screens by tallying the number of core component genes identified in each screen and comparing it to the expected number of genes known to belong to the pathway of interest. Note that for some screens, certain core components could not be evaluated because of the design of the assay.

bined the use of the small-molecule inhibitor rapamycin and RNAi to identify TOR-regulated genes that control growth and division.

### RNAi HTS as a Tool for a "Systems Biology" Approach

One of the most interesting promises of RNAi HTS is that it potentially provides a means to identify all of the parts of a network and thus could be used as the first step in a systems biology approach to understand the contribution of the genome to a biological process. Indeed, the data emerging from RNAi HTS could be used to integrate data sets generated from other "Omic" approaches (Fig. 2). The advantage of using data sets from RNAi HTS for such a purpose would be that it provides, unlike other Omic methods, direct functionality. Cell-based high-content screens (HCS) that rely on RNAi-induced cellular phenotypes are particularly well suited for this approach because they generate data sets that are rich in information. Each feature scored in such an assay is assessed independently, according to a controlled vocabulary.

The compilation of these features defines a phenotypic profile or "phenoprint," which is specifically associated with each gene knockdown (Piano et al. 2002). Using this approach, Piano et al. (2002) characterized early embryonic defects in *Caenorhabditis elegans* for 161 genes. Using time-lapse microscopy to systematically describe the defects for each gene in terms of 47 RNAi-associated phenotypes, these authors then clustered the genes into functionally related groups, an approach that can prove to be extremely powerful to functionally annotate unknown genes. However, for it to be useful in building or ordering large biological networks, RNAi HTS must generate an RNAi signature of high confidence, meaning that the rates of false positives in the RNAi HTS screens are low. Estimating the rates of false positives in RNAi screens is possibly one of the most difficult issues right now with the methodology as false positives can have originate from many sources (Echeverri and Perrimon 2006). Indeed, comparisons between related RNAi screens already performed by different groups, sometimes using different or similar RNAi libraries, are revealing poor overlap between data sets (Björklund et al. 2006; M. Kulkarni et al. 2006). The origin of the discrepancy between the studies is complex, as false positives can originate from stochastic, biological, and off-target noise. Stochastic noise refers to any experimental variation caused by random instrumentation malfunctions, plate-manufacturing defects, effectiveness of the reagents used (e.g., knockdown efficiency), as well as human error. Biological noise entails the unpredictable contribution of ill-defined biological variables to a phenotype or readout of interest that cannot easily be controlled. For instance, the state or health of a culture, passage number, or adaptation over time to certain conditions may alter how cells respond to RNAi in general or to a treatment in particular. Both stochastic and biological noise can be quantified and dealt with by performing multiple replicas of the assay and subjecting the results to strict statistical treatments, conditions that are rarely practical when performing a primary screen. In contrast, the contribution of OTE to the rate of false positives cannot easily be accounted for unless one has prior knowledge of OTE rules, which we do not have at present. For instance, setting the limit at 19 nucleotides for perfect homology with unintended targets as definition for OTE may not be sufficient, since shorter-length homologies have been reported to lead to OTE, at least when using single siRNAs (Birmingham et al. 2006). Consequently, using multiple distinct dsRNAs targeting the same gene should help eliminate or minimize the rate of OTE, as non-overlapping dsRNAs are unlikely to have overlapping off-targets. Using two or three dsRNAs targeting the same gene is therefore highly recommended, particularly if one desires to assign a particular function to a gene based on RNAi data alone. Validation of hits in secondary screens provides an ideal opportunity to supply statistical robustness to account for stochastic and biological noise and specificity to minimize OTE. Ideally, for each gene tested in a secondary screen, one would want to array multiple copies (5–7) of two or three independent dsRNAs interspersed with mixed positive and negative control dsRNAs (13–30 total) in a screening plate. Assaying each plate in duplicate ensures that numerous data points are obtained for each dsRNA, allowing statistical significance in the measured output and providing an accurate determination of the validation rate for each gene tested.

In conclusion, RNAi HTS technology, although a successful approach for gene discovery, is not yet fully mature as a tool for system biology. Only when the rate of false positives is better understood and the quality of the data consistently reliable will RNAi signature data sets become a major player in system biology. This should be achievable in the next few years by more carefully controlled experiments and better understood reagents.

## ACKNOWLEDGMENTS

We thank Jianwu Bai, Matt Booker, Adam Friedman, Lutz Kockel, Meghana Kulkarni, and Katharine Sepp for contributing ideas, material, and unpublished experimental data included in this manuscript. Work at the DRSC is supported by a grant R01 GM067761 from the NIGMS. N.P. is an Investigator of the Howard Hughes Medical Institute.

## REFERENCES

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185.

Agaisse H., Burrack L.S., Philips J., Rubin E.J., Perrimon N., and Higgins D.E. 2005. Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* **309:** 1248.

Armknecht S., Boutros M., Kiger A., Nybakken K., Mathey-Prevot B., and Perrimon N. 2005. High-throughput RNA interference screens in *Drosophila* tissue culture cells. *Methods Enzymol.* **392:** 55.

Baeg G.H., Zhou R., and Perrimon N. 2005. Genome-wide RNAi analysis of JAK/STAT signaling components in *Drosophila*. *Genes Dev.* **29:** 1861.

Bard F., Casano L., Mallabiabarrena A., Wallace E., Saito K.,

Kitayama H., Guizzunti G., Hu Y., Wendler F., Dasgupta R., et al. 2006. Functional genomics reveals genes involved in protein secretion and Golgi organization. *Nature* **439:** 604.

Bartscherer K., Pelte N., Ingelfinger D., and Boutros M. 2006. Secretion of Wnt ligands requires Evi, a conserved transmembrane protein. *Cell* **125:** 523.

Birmingham A., Anderson E.M., Reynolds A., Ilsley-Tyree D., Leake D., Fedorov Y., Baskerville S., Maksimova E., Robinson K., Karpilow J., et al. 2006. 3′UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3:** 199.

Björklund M., Taipale M., Varjosalo M., Saharinen J., Lahdenpera J., and Taipale J. 2006. Identification of pathways regulating cell size and cell-cycle progression by RNAi. *Nature* **439:** 1009.

Boutros M., Agaisse H., and Perrimon N. 2002. Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev. Cell* **3:** 711.

Boutros M., Kiger A.A., Armknecht S., Kerr K., Hild M., Koch B., Haas S.A., Paro R., and Perrimon N.; Heidelberg Fly Array Consortium. 2004. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303:** 832.

Cherry S., Doukas T., Armknecht S., Whelan S., Wang H., Sarnow P., and Perrimon N. 2005. Genome-wide RNAi screen reveals a specific sensitivity of IRES-containing RNA viruses to host translation inhibition. *Genes Dev.* **19:** 445.

Clemens J.C., Worby C.A., Simonson-Leff N., Muda M., Maehama T., Hemmings B.A., and Dixon J.E. 2000. Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl. Acad. Sci.* **97:** 6499.

DasGupta R., Kaykas A., Moon R.T., and Perrimon N. 2005. Functional genomic analysis of the Wnt-wingless signaling pathway. *Science* **308:** 826.

Echeverri C.J. and Perrimon N. 2006. High-throughput RNAi screening in cultured cells: A user's guide. *Nat. Rev. Genet.* **7:** 373.

Eggert U.S., Kiger A.A., Richter C., Perlman Z.E., Perrimon N., Mitchison T.J., and Field C.M. 2004. Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol.* **2:** e379.

Feske S., Gwack Y., Prakriya M., Srikanth S., Puppel S.H., Tanasa B., Hogan P.G., Lewis R.S., Daly M., and Rao A. 2006. A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441:** 179.

Foley E. and O'Farrell P.H. 2004. Functional dissection of an innate immune response by a genome-wide RNAi screen. *PLoS Biol.* **2:** E203.

Friedman A. and Perrimon N. 2004. Genome-wide high-throughput screens in functional genomics. *Curr. Opin. Genet. Dev.* **14:** 470.

Guertin D.A., Guntur K.V., Bell G.W., Thoreen C.C., and Sabatini D.M. 2006. Functional genomics identifies TOR-regulated genes that control growth and division. *Curr. Biol.* **16:** 958.

Gwack Y., Sharma S., Nardone J., Tanasa B., Iuga A., Srikanth S., Okamura H., Bolton D., Feske S., Hogan P.G., and Rao A. 2006. A genome-wide *Drosophila* RNAi screen identifies DYRK-family kinases as regulators of NFAT. *Nature* **441:** 646.

Hammond S.M., Bernstein E., Beach D., and Hannon G.J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404:** 293.

Hild M., Beckmann B., Haas S.A., Koch B., Solovyev V., Busold C., Fellenberg K., Boutros M., Vingron M., Sauer F., et al. 2003. An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5:** R3.

Jackson A.L., Bartz S.R., Schelter J., Kobayashi S.V., Burchard J., Mao M., Li B., Cavet G., and Linsley P.S. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21:** 635.

Kiger A., Baum B., Jones S., Jones M., Coulson A., Echeverri C., and Perrimon N. 2003. A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2:** 27.

Kulkarni M.M., Booker M., Silver S.I., Friedman A., Hong P., Perrimon N., and Mathey-Prevot B. 2006. Evidence of off-target effects associated with long dsRNAs in *Drosophila* cell-based assays. *Nat. Methods* **3:** 833.

Levsky J.M., Shenoy S.M., Pezo R.C., and Singer R.H. 2002. Single-cell gene expression profiling. *Science* **297:** 836.

Lum L., Yao S., Mozer B., Rovescalli A., Von Kessler D., Nirenberg M., and Beachy P.A. 2003. Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299:** 2039.

Muller P., Kuttenkeuler D., Gesellchen V., Zeidler M.P., and Boutros M. 2005. Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature* **436:** 871.

Nybakken K., Vokes S.A., Lin T.Y., McMahon A.P., and Perrimon N. 2005. A genome-wide RNA interference screen in *Drosophila melanogaster* cells for new components of the Hh signaling pathway. *Nat. Genet.* **37:** 1323.

Pelech S. 2004. Tracking cell signaling protein expression and phosphorylation by innovative proteomic solutions. *Curr. Pharm. Biotechnol.* **5:** 69.

Pelkmans L., Fava E., Grabner H., Hannus M., Habermann B., Krausz E., and Zerial M. 2005. Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* **436:** 78.

Philips J.A., Rubin E.J., and Perrimon N. 2005. *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. *Science* **309:** 1251.

Piano F., Schetter A.J., Morton D.G., Gunsalus K.C., Reinke V., Kim S.K., and Kemphues K.J. 2002. Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* **12:** 1959.

Ramet M., Manfruelli P., Pearson A., Mathey-Prevot B., and Ezekowitz R.A. 2002. Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* **416:** 644.

Sachs K., Perez O., Pe'er D., Lauffenburger D.A., and Nolan G.P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308:** 523.

Schlesinger A., Kiger A., Perrimon N., and Shilo B.Z. 2004. Small wing PLCgamma is required for ER retention of cleaved Spitz during eye development in *Drosophila*. *Dev. Cell* **7:** 535.

Stolc V., Gauhar Z., Mason C., Halasz G., van Batenburg M.F., Rifkin S.A., Hua S., Herreman T., Tongprasit W., Barbano P.E., et al. 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306:** 655.

Ulvila J., Parikka M., Kleino A., Sormunen R., Ezekowitz R.A., Kocks C., and Ramet M. 2006. Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in *Drosophila* S2 cells. *J. Biol. Chem.* **281:** 14370.

Vig M., Peinelt C., Beck A., Koomoa D.L., Rabah D., Koblan-Huberson M., Kraft S., Turner H., Fleig A., Penner R., and Kinet J.P. 2006. CRACM1 is a plasma membrane protein essential for store-operated Ca2+ entry. *Science* **312:** 1220.

Wheeler D.B., Carpenter A.E., and Sabatini D.M. 2005. Cell microarrays and RNA interference chip away at gene function. *Nat. Genet.* (suppl.) **37:** S25.

Wheeler D.B., Bailey S.N., Guertin D.A., Carpenter A.E., Higgins C.O., and Sabatini D.M. 2004. RNAi living-cell microarrays for loss-of-function screens in *Drosophila melanogaster* cells. *Nat. Methods* **1:** 127.

Yandell M., Bailey A.M., Misra S., Shu S., Wiel C., Evans-Holm M., Celniker S.E., and Rubin G.M. 2005. A computational and experimental approach to validating annotations and gene predictions in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci.* **102:** 1566.

Yao S., Lum L., and Beachy P. 2006. The ihog cell-surface proteins bind Hedgehog and mediate pathway activation. *Cell* **125:** 343.

Zhang S.L., Yeromin A.V., Zhang X.H., Yu Y., Safrina O., Penna A., Roos J., Stauderman K.A., and Cahalan M.D. 2006. Genome-wide RNAi screen of Ca2+ influx identifies genes that regulate Ca2+ release-activated Ca2+ channel activity. *Proc. Natl. Acad. Sci.* **103:** 9357.