Drosophila RNAi screening in a postgenomic world

Chris Bakal

Advance Access publication date 12 July 2011

Abstract

Drosophila melanogaster has a long history as a model organism with several unique features that make it an ideal research tool for the study of the relationship between genotype and phenotype. Importantly fundamental genetic principles as well as key human disease genes have been uncovered through the use of Drosophila. The contribution of the fruit fly to science and medicine continues in the postgenomic era as cell-based Drosophila RNAi screens are a cost-effective and scalable enabling technology that can be used to quantify the contribution of different genes to diverse cellular processes. Drosophila high-throughput screens can also be used as integral part of systems-level approaches to describe the architecture and dynamics of cellular networks.

Keywords: RNAi, cell-based screens; Drosophila melanogaster; genetic interactions; systems biology

INTRODUCTION

In the postgenomic era, the mapping, sequencing and cloning of genes, once considered major hurdles in genetics research, have become trivial processes. Now a key challenge in biology is to understand how genetic diversity contributes to phenotypic variation and disease. Only a quarter of human genes have well-described functions [1], and much work still needs to be done toward characterizing the genome [2]. Clearly both healthy and disease phenotypes are non-Mendelian in nature, and are due to highly complex genetic interplay between genes themselves, as well as between genes and factors such as lifespan, gender and the environment. Thus new technologies are desperately needed in order map these dynamic interactions on a systems level. In this review, we will discuss how RNAi screens using cultured Drosophila melanogaster cells will continue to be an invaluable methodology in performing modern genetics research.

Drosophila’s life as model organism started in the laboratory of William E. Castle at Harvard University in 1901 [3]. But it was shortly thereafter in T. H. Morgan’s ‘fly room’ that Drosophila truly began its role as one of the key workhorses of genetics. In the hands of Morgan and his students Alfred Sturtevant, Calvin Bridges, Edgar Altenburg and Herman Müller, Drosophila was an ideal tool to identify and map the location of genes that determine diverse phenotypic traits [4, 5]. In fact, an accurate physical map of the Drosophila genome was reached far before those of other organisms [6]. The usefulness of the fruit fly continued through the 1980s and 1990s where, for example, genome-scale in vivo genetic screens successfully identified novel genes and the relationships between them [7]. In particular, many components of highly conserved signal transduction pathways that have significant roles in human health and disease, such as the Ras-MAPK, JAK-STAT, Notch, Wnt and Hedgehog pathways were first identified in Drosophila. However, in a world where low-cost technologies can be used to map and identify genes, and any gene can be inhibited at will by RNAi in many cell types and organisms, it is remarkable that we continue to discover reasons why Drosophila is still a perfect model for use in research that aims to answer fundamental questions in biology and medicine.

Corresponding author. Chris Bakal, Dynamical Cell Systems Laboratory, Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK. Tel: +44 (0)20 7153 5080; Fax: +44 (0)20 7153 5577; E-mail: chris.bakal@icr.ac.uk

Chris Bakal is a Team Leader in the Division of Cancer Biology at the Institute of Cancer Research in London, UK. He uses quantitative functional genomic approaches in tandem with computational technologies in order to describe the signaling networks that regulate various cellular behaviors such as morphogenesis, migration, cell-cycle progression and the stress response.

© The Author 2011. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com
Drosophila’s new role in the postgenomic era began after a series of experiments in the Hamonn and Dixon laboratories where it was proven that Drosophila tissue culture cells can be conditioned to uptake long double-stranded RNAi (dsRNA) in the absence of transfection reagent—often resulting in undetectable levels of a target protein [8, 9]. Importantly, this effective knockdown with long dsRNA occurs in the absence of the inflammatory response that is initiated in mammalian cells [10]. Although at this time RNAi was already becoming an important in vivo experimental tool in many organisms including Drosophila [11], the discovery of RNAi in Drosophila tissue culture was incredibly important as it initiated the generation of genome-scale libraries of long Drosophila dsRNAs by cost-effective PCR-based methods for use in cell culture. Three years later, the Perrimon [12] and Beachy [13] groups had completed the first genome-scale RNAi screens in Drosophila cells. Less than a year later the first genome-wide screen was published [14]. Since then over 50 Drosophila RNAi screens have been completed in labs around the world [15].

The impacts of these screens have been profound in at least three ways: (i) by identifying the function of thousands of previously uncharacterized Drosophila genes and their homologs in other organisms; (ii) by identifying key disease genes and thus providing important insights into human health; and (iii) through generating novel experimental and computational methodologies that have been used in screens across organisms that have accelerated the pace of genetics research.

Drosophila RNAi screens: a cost-effective, high throughput and scalable technology

Virtually any laboratory can generate a Drosophila RNAi screening library with common laboratory reagents at a very low cost when compared to the purchase of RNAi for use in mammalian cells. A number of excellent protocols describing the design and generation of Drosophila RNAi have been published extensively elsewhere [16]. Briefly, this involves the PCR-mediated amplification of small DNA fragments (or ‘amplicons’) of ~150–750 bp fragments from Drosophila genomic DNA using primers containing a T7 priming sequence. T7-mediated In Vitro Transcription (IVT) is then used to generate dsRNA from DNA amplicons. These methods facilitate the generation of multiple independent dsRNAs targeting a single gene. For example, in a recent screen of kinase and phosphatase–encoding genes, we used as many as nine individual dsRNAs to inhibit a single gene [17]. The ease and low cost of Drosophila RNAi production is particularly advantageous over mammalian systems when large amounts of very specific RNAi reagents are required during rescreening and validation of primary RNAi hits. Furthermore, as many Drosophila cells uptake dsRNA in the absence of costly lipid-based transfection reagents, the cost of performing the actual screen itself is considerably less than that of executing similar screens in mammalian cells. As long dsRNAs introduced into Drosophila cells are subsequently processed by the Dicer complex into multiple different 21 bp fragments, this also increases the chance that knockdown will be effective compared to using only a single 21 bp RNAi.

Notably, the fly genome is most useful attribute of Drosophila RNAi screening is that the fly genome is highly similar to that of higher metazoans (~50% of Drosophila protein sequences have fly homologs) [6]. In fact, many studies have shown that hits identified in Drosophila screens can be directly translated to inference of gene function in humans, mice or zebrafish [18, 19]. Moreover the Drosophila genome is ~3-fold less redundant than the human or mouse genome. Thus screens in Drosophila are more likely than those performed in mammalian cells to lead to provide insight as to that gene’s biological role. One critical facet of the similarity between the two organisms is that ~75% disease causing genes in humans have Drosophila homologs [20]. Meaning not only is a genetic screen in Drosophila more likely to identify genes all genes related to a particular process, but it is more likely the genes identified in such screens will have relevance to human health and even serve as potential drug targets. An excellent example of the power of Drosophila cell-based RNAi screens to isolate disease genes was the simultaneous identification by several groups of the Onai gene that encodes an essential component of the Calcium release-activated Ca2+ (CRAC) channel complex [21–24].

There are additional available resources in that make Drosophila cell culture a powerful system for high-throughput genetics. The first is the growing number of Drosophila RNAi-expressing lines [25–28]. Thus hits generated in cell-culture screens can be very quickly validated and characterized in vivo. Moreover, data generated as part of the Drosophila MODEncode (model organism

Drosophila genome database). The second is the availability of the Drosophila genome database. The third is the availability of the Drosophila genome database. The third is the availability of the Drosophila genome database. The third is the availability of the Drosophila genome database.
ENCyclopedia of DNA Elements) project provides a fantastically rich resource with which to further analyze hits [29, 30]. In particular, the modENCODE project is characterizing the same cell lines that are frequently used in RNAi screens, which is the ideal situation for computational integration of orthogonal data sets.

Some issues to consider: false positives and off-target effects

Although generating long dsRNA by PCR/IVT is a cost-effective means of generating highly effective RNAi reagents, a different type of price comes with the use of these reagents—a higher probability of off-target effects (OTEs). OTEs occur when the levels of mRNAs/proteins that are not being intentionally targeted by a specific RNAi are reduced. Thus OTEs increase the rate of false positives, an important issue in any type of screen. While OTEs due to a dsRNA-initiated interferon response do not occur in Drosophila tissue culture, careful studies have shown that dsRNAs with regions of homology \( \geq 19 \) bp can result in an RNAi-mediated gene inhibition [31]. OTEs are a particular problem when dsRNAs target gene regions that include CAN (where \( N \) represents any nucleotide base) repeats that are present in many Drosophila genes [32]. The prevalence of OTEs was not initially appreciated in early RNAi screens, but the field has placed considerable effort into developing methods to minimize their impact [33]. Techniques to reduce OTEs and false positives include: computational approaches to design dsRNAs with very limited or no regions of homology to other genes; having multiple independent dsRNAs score as significant enhancers or suppressors in a screen before a gene is considered a hit [17]; and using the overexpression of homologous genes from related Drosophila species to rescue deficiencies caused by RNAi [34].

A very important point is that if a dsRNA has a small region of sequence homology to other genes this does not necessarily mean that it will be effective at knocking down these unintended targets. Previous studies have shown that 50 of 135 computationally predicted off-targets in a 19 bp dsRNA targeting the \( PP2A-B' \) gene are not affected by this RNAi [31]. Thus \emph{in silico} predictions are usually very conservative overestimates of OTEs [35]. Conversely, it is very clear that not all dsRNAs are effective at knockdown of their intended targets – resulting in false negatives. For example, we have observed in RNAi screening for components of the Jun N-terminal Kinase (JNK) signaling network that there are cases where three or four dsRNAs targeting a single gene score above/below a pre-determined threshold (e.g. MLK, msn), which strongly suggests these genes are indeed JNK regulators, but not all dsRNAs targeting these genes are similarly effective [17]. The fact that some dsRNAs do not work could be due to the fact that different splice variants are expressed in some cells, and/or that Dicer-mediated processing of particular dsRNA does not lead to a single effective siRNA. As neither on-target nor OTEs can be completely predicted, there still remains much to be learned regarding the ‘rules’ of how to design an RNAi library.

One additional problematic issue with performing RNAi screens in Drosophila is that compared to mammalian systems there is a limited number of cell types that can be screened, which limits the diversity of cellular behaviors and processes which can be investigated in a high-throughput manner. Approximately five cell types have been well used in Drosophila screens; S2, S2R+, Kc167, Clone-8 and BG (BG-2 and BG-3) cells. S2, S2R+ and Kc167 cells are all embryonic lines although based on expression profiling are quite distinct from each other [36, 37]. Clone-8s are derived from L3 wing discs, and have some epithelial properties. But unlike classical epithelial cells, Clone-8 cells are not well-polarized and do not form cell–cell junctions in culture. BG cells are neuronal in origin, but have not reported to be capable of executing differentiation-type processes such as neurite extension [38]. Unfortunately S2, Clone-8 and Kc167 are also round small cells which and are poorly suited for image-based screens that require high-resolution of subcellular structures.

Further complicating the matter is that recent profiling studies of common Drosophila cell lines have shown that many ligands which activate key signaling pathways such as Notch, Wnt and even insulin signal transduction pathways are not expressed at all, or are expressed at very low levels [36]. Thus in order to screen for some particular phenotypes, Drosophila cells must be engineered to exhibit particular behaviors by administration of exogenous ligands or through overexpression of different constructs. However, just as Drosophila cells are excellent for genetic screening, they are also well suited for synthetic biology as demonstrated by recent studies that have successfully induced polarity in S2 cells [39].
The development of increasingly complex phenotypic readouts

To date a number of diverse, and sometimes ingenious, phenotypic readouts have been used in the context of *Drosophila* RNAi screens to comprehensively identify genes that contribute to different cellular behaviors. Many screens for components of signal transduction networks have been performed using reporters that incorporate multiple binding transcription factor binding sites upstream of luciferase genes [18, 35, 40–46], phosphospecific antibodies [19, 47, 48], or live FRET-based reporters [17]. Because they can be infected by various pathogens, *Drosophila* cells are also excellent models for studies of the immune response [49–52]. Additionally, some highly novel screens that have been completed include screens for light-mediated CRY degradation [53] or protein secretion [54]. While these screens have been effective at identifying new genes, one issue with these and many other genetic screens are that the phenotypic output is binary in nature and that hits can only be characterized as enhancers or suppressors of a particular pathway. Finally on the basis of binary screening data alone one cannot typically infer connections between genes.

Image-based screening has been particularly powerful at revealing novel genetic insights into diverse cellular process, and will continue to do as there is a practically limitless number of assays that can be developed using diverse labeling reagents and fluorescent reporters. Because nearly 100% of *Drosophila* cells in culture will uptake dsRNA, low penetrance of analyzable phenotypes due to low transfection efficiency is not a complicating factor as in many mammalian screens. High efficiency of dsRNA uptake facilitates the quantification of phenotypic homogeneity and heterogeneity following gene knockdown, whereas in mammalian screens the effects of an RNAi on a population must be determined over the background of untransfected (and thus wild-type) cells. Using image-based approaches multiple features can be quantified simultaneously following the gene knockdown to generate ‘phenotypic signatures’ which quantitatively describe the contribution of a gene to a phenotype such as cell shape [55, 56]. Multidimensional signatures can also be generated that describe cellular structures such as mitotic spindles [57]. The high-dimensional aspect of these signatures can be exploited to determine functional relationships between genes (Figure 1). Increasing the number of independent uncorrelated features that describe different phenotypes increases the specificity of the categories into which genes can be sorted. For example, by hierarchical clustering of multidimensional descriptors of cell shape following systematic RNAi, or Quantitative Morphological Signatures (QMSes), we were able to group genes into ‘local networks’ of genes that regulate diverse aspects of morphogenesis such as adhesion and protrusion [55]. Genes that belong to a particular local network not only share the same function, but are predicted to encode proteins with similar subcellular colocalisation and which may physically interact. Further advances in computational and statistical methods combined with improved image-recognition techniques are likely to provide researchers with the ability to go from image data to network models.

Despite the limited number of cell lines with which to perform RNAi screens, it would seem that *Drosophila*, as always, is not finished providing novel and powerful genetic tools. Recent studies have demonstrated that primary cell cultures of muscle and neuronal origin can be used in high-throughput screens [58–61]. These cultures take up long dsRNA (without transfection reagent) and RNAi is as efficient as in established lines. Primary cells also exhibit many of the same behaviors in a 384-well dish as they do in vivo. For example, muscles primary cultures form highly well-organized sarcomeric structures, and stunningly, even rhythmically contract [58, 60]. Bai et al. [60] were able to use primary muscle cultures in order to identify sarcomere length short (SALS), a novel regulator of sarcomere biogenesis. Primary neurons also form long neuritic processes—which allowed Sepp et al. [61] to determine isolate genes that regulate neuronal morphogenesis. Using appropriate tissue culture conditions it is likely other types of lines could be generated from different genetic backgrounds.

*Drosophila* RNAi as a method for describing genetic interactions

It is now clear that most complex phenotypes are due to the interactions of multiple genes. In fact, it was work by Altenburg and Müller [62] on determining the genetic interaction underlying the ‘truncate’ mutation in *Drosophila* that not only first revealed the possible complexity of the genotype–relationship, but also provided the technical method of linkage analysis that is still in use today.
**Figure 1:** From Images to Networks. Cell segments are generated using automated image recognition software that recognizes cellular boundaries. From these cell segments, features that describe various aspects of morphology are measured. For example, roundness, size, protrusion number, as well as hundreds of other features can be quantified. Using statistical methods such as hierarchical clustering, the correlation between two or more signatures can be described. When gene inhibition by RNAi results in two or more highly correlated signatures, these genes are very likely to encode components of the same signaling pathway or complex. Typically, statistical methods must be implemented in order to reduce the dimension of morphological signatures prior to the use of clustering or other approaches.
However, while hundreds of genome–wide associ-ates studies (GWAS) have revealed the identity of single alleles that quantitatively contribute to diverse phenotypes, performing the experiments required to identify and/or validate multigenic interactions, is still problematic in human cells. Research that has had profound impact on our understanding of genetic networks has involved screens for synthetic lethality or other phenotypes in \textit{Saccharomyces cerevisiae}. In these screens, viability is scored following the simultaneous deletion or inhibition of two genes in extremely high throughput (for essential genes hypomorphic alleles are often used) \cite{63–67}. When deletion of two genes results in a phenotype that is not predicted based on the phenotypes following mutation of either single gene, this provides insight into multigenic effects. Yeast genetic interactions have been mapped onto to high–quality protein–protein interaction data sets to understand the physical basis for epistasis \cite{68–71}. However, there is obviously limited relevance of yeast synthetic interactions to multigenic interactions underpinning human phenotypes and disease. In addition, the nature of yeast genetic interaction studies are such that systematically determining the phenotype following the deletion/inhibition of more than two genes in high throughput is technically difficult.

\textit{Drosophila} RNAi in cell culture is an ideal system in order to bridge the gap between high–throughput studies of synthetic lethality in yeast and understanding of complex phenotypes in human as combinatorial RNAi experiments are more feasible and less expensive than in mammalian cells. Owing to the high homology between fly and human genomes, synthetic screens in \textit{Drosophila} cells are likely more relevant to disease phenotypes than those performed in yeast. We have already demonstrated that double–RNAi screens in \textit{Drosophila} cell culture are feasible in a medium–throughput study of 17 724 RNAi interactions \cite{17}. Recently, the Boutros laboratory has performed an even larger screen of interactions between MAPK signaling components \cite{72}. Some, but likely not all, of the intuition and methods that has been applied to the interpretation of yeast synthetic lethal screens may be transferable to studies of \textit{Drosophila} genetic interactions. For example, in yeast negative or aggravating interactions often occur between genes that function in redundant or compensatory pathways (‘between pathway’ interactions), whereas positive or alleviating interactions are enriched for function in the same pathway or complex \cite{73–79}, but whether phenotypes resulting from double RNAsi of two genes in \textit{Drosophila} can be interpreted in analogous fashion is unclear. In a screen for JNK components, we identified many more genetic interactions in backgrounds where canonical JNK regulators such as ‘slipper’ (86 interactions) or ‘puckered’ (72 interactions) were inhibited by RNAi than in genes that likely are not part of the core JNK pathway (e.g. 7 interactions were detected in the ‘hippo’ deficient background). Moreover, many predicted kinase–substrate interactions overlap with genetic interactions (both alleviating and aggravating) \cite{17}. This data suggests that epistasis following double RNAi more often indicates within pathway/complex interactions, mirroring the results from classic genetic interaction screens using sensitized background in \textit{Drosophila} which tend to identify proximal components of a regulatory pathway rather than genes acting in a parallel fashion \cite{80}. But this model remains to be formally proven. In any case, as more combinatorial screens are performed, it is essential that computational methods be developed in order to properly interpret the results of these data sets. Towards developing tools for analysis of high–throughput combinatorial screens we recently developed a novel method to infer both within– and between–pathway interactions in the Rho–signaling network \cite{81}.

\textbf{Super–complex phenotypes following RNAi}

As the phenotypic readouts that can be implemented in the context of RNAi screens become ever more complex, this should provide ever more detailed insight into gene function. Image–based readouts, such as those which describe hundreds of parameters of cell shape are on type of complex signature which can be used to both describe the specific contribution of individual genes, but also to model relationships between genes. Other types phenotypic readouts that would provide very high–dimensional data include transcriptional profiles, metabolomic data, phosphoproteomic analyses \cite{82–84} or possibly even RNA sequencing data. Although, obviously all these approaches currently are not amenable to high throughput either due to cost (e.g. transcriptional profiling, mRNA sequencing) and/or the large volume of cells and media, that must be used for each sample (e.g. metabolomics,
phosphoproteomics). We have recently explored the use of transcriptional profiling following medium-throughput RNAi (approximately 150 samples) to model the Rho-regulatory network [85]. Screens using expression profiling in particular are likely going to become more common as the cost of microarrays continues to decrease [86, 87].

CONCLUSION
Since the days of Morgan’s fly room, Drosophila has been at the leading edge of genetics research and continues to do so over a century later. While RNAi screening technologies in mammalian cell lines have improved to the point where genome-wide screens are extremely feasible, the cost-effectiveness of Drosophila screens alone still makes the use of Drosophila an excellent research tool. Moreover, it is likely that increasingly complex combinatorial screens in order to probe genetic interactions or the screening of primary cell types will be more feasible in Drosophila cells over the near future. The fact that a number of rich orthogonal data sets are being generated in Drosophila lines has improved to the point where genome-wide screens are extremely feasible, the near future. The fact that a number of rich orthogonal data sets are being generated in Drosophila lines in parallel also means that the data generated by Drosophila RNAi screens will be extremely useful for systems-level integrative analysis.

Key Points
- Drosophila cell-based RNAi screens are a cost-effective way to quantify the contribution of genes to diverse phenotypes in a very high-throughput manner. The low cost of dsRNA production facilitates the generation of libraries that contain many independent dsRNAs targeting the same gene which reduces the occurrence of both false negatives and false positives.
- Image-based genetic screens that can describe phenotypes such as cell shape in a highly multiparametric fashion can be used to not only identify the functions of single genes, but also determine relationships between genes and thus model cellular networks.
- Because combinatorial screening is more technically feasible and cost-effective using Drosophila versus mammalian cell lines, but Drosophila genetic interactions are more likely to be recapitulated in mice or human cells than yeast genetic interactions, Drosophila is likely going to be an important model organism in the future to describe genetic interactions on a systems level.

FUNDING
Wellcome Trust Research Career Development Fellowship (to C.B.); Biotechnology and Biological Sciences Research Council (BB/I002510/1 to C.B.).

References