Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Methods 40 (2006) 262–271
METHODS
www.elsevier.com/locate/ymeth

High-throughput approaches to dissecting MAPK signaling pathways

Adam Friedman, Norbert Perrimon *

Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

Accepted 13 May 2006

Abstract

With the development of genome-wide RNAi libraries, it is now possible to screen for novel components of mitogen-activated protein kinase (MAPK) pathways in cell culture. Although genetic dissection in model organisms and biochemical approaches in mammalian cells have been successful in identifying the core signaling cassettes of these pathways, high-throughput assays can yield unbiased, functional genomic insight into pathway regulation. We describe general high-throughput approaches to assaying MAPK signaling and the receptor tyrosine kinase (RTK)/extracellular signal-regulated kinase (ERK) pathway in particular using a phospho-specific antibody-based readout of pathway activity. We also provide examples of secondary validation screens and methods for managing large datasets for future *in vivo* functional characterization.
© 2006 Elsevier Inc. All rights reserved.

Keywords: High-throughput screen; RNAi; dsRNA; ERK; MAPK; RTK signaling; Data mining; dpERK; *Drosophila*

1. Introduction

Signaling through receptor tyrosine kinases (RTKs) plays highly conserved roles in metazoans, controlling fate determination, differentiation, proliferation, survival, migration, and growth [1,2]. Although RTKs activate multiple downstream pathways, one particularly well-studied and evolutionarily conserved pathway is that of the extracellular signal-regulated kinases (ERKs), members of the mitogen-activated protein kinase (MAPK) superfamily, which also include the stress-activated protein kinases (SAPKs) such as c-Jun N-terminal kinase (JNK) and p38 [3,4]. The broad importance of RTK/ERK signaling is highlighted by the well-documented pathway dysregulation in human disease, most notably cancer. Mutations in multiple RTKs have been implicated in a variety of cancers, such as the epidermal growth factor receptor family members in breast and lung cancers [5]. Thirty percent of all solid tumors have mutations in Ras or Raf, including up to 90% of pancreatic adenocarcinomas [6].

Given the recent success of targeted therapies in cancer drug discovery [7], there is considerable interest in the discovery of novel components of RTK/ERK signaling. Our current knowledge of RTK/ERK signaling is the synthesis of decades of dissection in model systems such as analysis of the yeast MAPK mating-type response, vulva formation in *Caenorhabditis elegans*, sensitized genetic screens for morphogenesis defects in *Drosophila*, and biochemical and cell biology in mammalian tissue culture. Subsequent to the characterization of the core signaling cascade culminating in ERK activation, recent research has focused on the role of feedback regulation and signaling dynamics in controlling output [8–12], placing the core cassette within a larger network of regulatory proteins.

Faithful recreation of endogenous cell circuits depends on the identification of all components of the pathways. Discovery of novel components of this pathway and those of the surrounding regulatory network requires the development of new technologies, as many genes may not have been discovered through traditional genetic techniques due to pleiotropy of genes with functions in multiple pathways, mutational bias, redundancy, and/or the requirement for visible phenotypes. We have previously shown that genome-wide RNAi screens in cell culture can identify

* Corresponding author.
E-mail address: perrimon@receptor.med.harvard.edu (N. Perrimon).

1046-2023/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2006.05.002

novel members of signaling pathways in an unbiased manner [13–15]. Building upon this work, we first outline general assays for dissection of the RTK/ERK signaling pathway. Second, we describe in detail a phospho-specific antibody method we have used to discover novel genes controlling ERK activation in a quantitative, loss-of-function assay (A.F. and N.P., submitted). These methods are generally applicable to other MAPK RNAi screens.

2. General considerations for high-throughput RTK/ERK assay development

Assays for genetic dissection of RTK/ERK signaling depend on knowledge of the core signaling cassette to measure activities of proteins at particular points during signal propagation. A summary of potential assays for pathway measurement is presented in Fig. 1. Approaches for elucidating signaling pathways in general by RNAi screening in mammalian cells has been recently reviewed [16]. Theoretically, more proximal assays measuring pathway activity near the activated RTK would identify fewer regulators than more distal assays and may miss components of input branches from other receptors. However, distal pathway readouts (e.g., transcriptional reporters or morphological outputs) may integrate more pathways than is desirable, only some of which may be known, challenging outcome attribution to the upstream MAPK pathway. Analysis of signal transduction cascades has traditionally relied on the use of transcriptional reporters, e.g., [13–15]. Although many effectors of RTK/ERK signaling have been identified, well-known transcriptional targets of this pathway include members of the ETS-domain-containing class of transcription factors [17]. A multimerized ETS-binding site-based transcriptional reporter has been successfully used in *Drosophila* for readout of this pathway [18,19] and may be a viable HTS RNAi assay. However, transcriptional reporters that rely on such artificial, multimerized motifs may not faithfully report output in all cell types or may integrate additional pathways during transcriptional activation; in addition, the quantitative relationship between endogenous pathway activity and output of these reporters is unclear. More generally, microarray-based functional genomic analysis, based on responses to gain or loss-of-function of specific pathway components, can yield downstream effectors. For RTK/ERK signaling in *Drosophila*, this approach has been successfully applied to measure changes in transcription *in vivo* in response to Ras^{V12} expression in hemocytes [20] and gain- and loss-of-function alleles of RTK/ERK pathway members in the *Drosophila* ovary [21].

More proximal assays may rely on microscopy ("high-content screening," HCS) or plate-reader (PR) assays of specific signaling events. Although PR assays are usually faster than HCS and simplify analysis by reducing pathway activity to a single number representing a population average, HCS can capture much more information about a pathway [22]. HCS assays can be multiplexed to follow, for example, both protein localization and abundance, and can capture many parameters of pathway modulation such as cell-to-cell variability in signal transduction [23]. On the other hand, HCS automated image analysis, feature

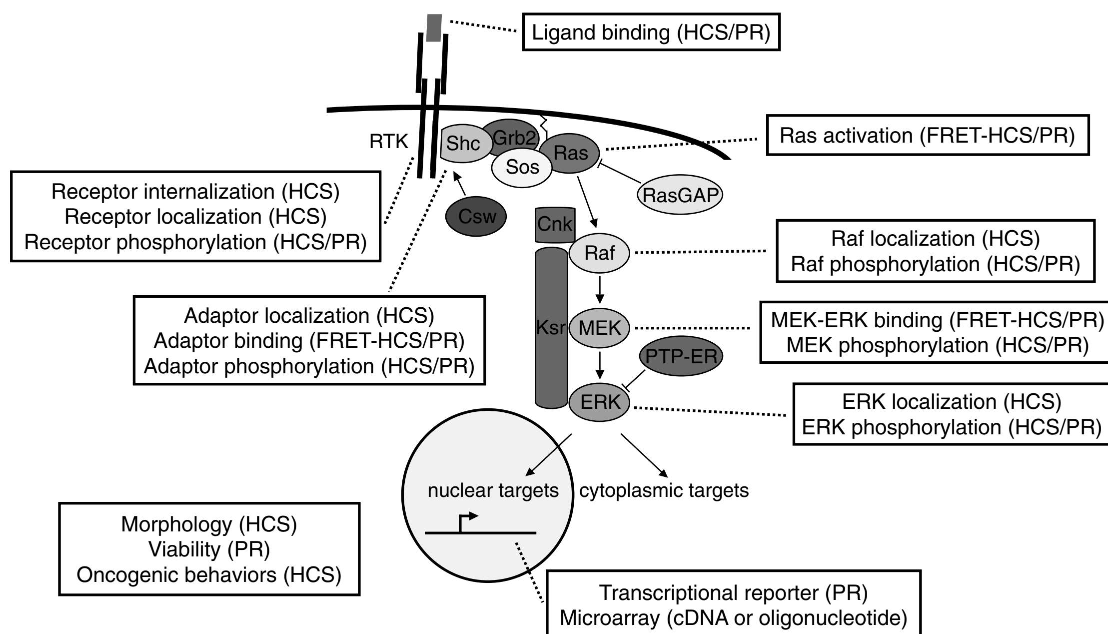


Fig. 1. Potential assays for high-throughput analysis of RTK/ERK signaling. We show for reference the canonical RTK/ERK signaling cascade in *Drosophila*, from ligand binding and RTK activation, to ERK phosphorylation (see review by [2]). For most proteins in the cascade, reagents have been developed that can be used in HTS assays in cell culture. Many assays, such as protein phosphorylation, may be assayed by either PR or HCS. In addition, pathway outputs, such as transcriptional activation or phenotypic changes, could be used as screen assays. FRET, fluorescence resonance energy transfer; HCS, high-content microscopy-based screening; PR, plate-reader.

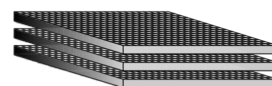
extraction, and data storage can be challenging and is undergoing rapid development.

Here, we describe specifically an assay based on activation of ERK using phospho-specific antibodies. Many other activation-state-specific reagents have been developed, such as antibodies to phosphorylated RTKs themselves, activating and inhibitory sites on Raf, and phosphorylated MEK. Staining intensity could be analyzed either using a plate-reader, which reports a single intensity measurement in each well, or HCS, which captures the immunofluorescence image. More exotic assays, based on fluorescence resonance energy transfer (FRET) detecting recruitment of adaptor proteins to receptors [24] or Ras activation [25] are also feasible. Other HCS assays such as nuclear–cytoplasmic shuttling of ERK [26], receptor internalization [27], or Ras localization [28] can be combined with phospho-specific antibody reagents to yield multiple variables that may give further insight into the function of the given gene. Finally, downstream phenotypic outputs of MAPK pathway activation could also identify novel genes. For example, a previous HCS screen focusing on cell morphology identified RTK/ERK signaling components including Ras, Raf, and Ksr [22]; other genes isolated in this screen with similar patterns of morphological changes may also be components of this pathway. Oncogene-mediated growth arrest or transformation phenotypes have also been used as assays in RNAi HTS, uncovering components of MAPK pathways [29,30].

3. A high-throughput assay for RTK/ERK signaling in *Drosophila* cells

Given the pivotal role of ERKs as RTK effectors and the commercial availability of well-characterized antibodies directed towards highly conserved, dually phosphorylated, active ERK1/2 (dpERK), we developed a high-throughput assay for analyzing dpERK levels in *Drosophila* cells (A.F. and N.P., submitted). ERK activation is a common feature of RTK activation for both mammalian receptors and their *Drosophila* orthologs. In *Drosophila* cells, elevation of dpERK following ligand stimulus has been observed following insulin activation of insulin receptor (InR) [31], epidermal growth factor (EGF) ligand Spitz activation of *Drosophila* EGFR [32], and Pvf ligand activation of the PDGF/VEGF homolog receptor (PVR) [33].

Our assay was adapted from traditional immunohistochemical techniques (outlined in Fig. 2). More general procedures for RNAi HTS, particularly using transcriptional reporters, have been previously published [34]. Advantages of RNAi HTS for MAPK screening in *Drosophila* include (1) the broad conservation of the RTK/ERK signaling cassette between *Drosophila* and mammals, (2) the advantages of gene discovery in *Drosophila* due to reduced gene redundancy, and (3) the ease of genetic manipulation in both *Drosophila* cell culture through RNAi technology and



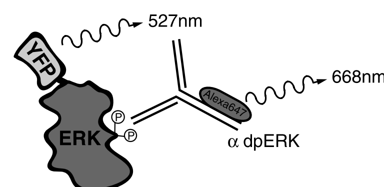


Fig. 2. Outline of phospho-specific antibody HTS. Flowchart for a phospho-specific antibody-based RNAi screening procedure detecting dually phosphorylated, active ERK, normalized to its YFP tag, as described in the text.

in vivo through mutational and transgenic techniques. All of the principles we describe can easily be applied to mammalian RNAi screening platforms.

3.1. Assay optimization

We chose cell number, length of soaking in serum-free medium, length of incubation, and reagent volumes to maximize signal:noise and cell viability. Although there is baseline ERK activation in S2 cells [33], increasing signal:noise by stimulating the pathway improves the dynamic

range and thus the sensitivity to perturbations such as by dsRNAs targeting genes involved in pathway regulation. As S2 cells endogenously express the *Drosophila* insulin receptor and can be easily activated by commercially available insulin [31], we used this ligand to stimulate the pathway. As dpERK levels peak at 10 min following insulin stimulation, we measured pathway activity at this time point as well as at baseline. Dynamic range and sensitivity of the assay were optimized using dsRNAs targeting known components of the RTK/ERK pathway such as InR itself, Ras, Raf, PTP-ER, and MEK. In general, this phospho-specific antibody approach led to a dynamic range of 30–150% normalized dpERK compared to control dsRNAs; this range is much narrower than enzyme-based transcriptional reporters but may more faithfully report endogenous signal intensity.

3.2. Cell seeding

We chose the S2 sub-line S2R+ because it is adherent and thus has less cell loss during plate manipulations. *Drosophila* S2R+ cells (4×10^4) were seeded in 20 μL serum-free medium in black, clear-bottom 384-well plates (Corning) with the DRSC dsRNA collection suspended in 5 μL water per well and spun for 1′ at 1200 rpm. After 45 min, 20 μL of medium with 20% serum (2×) was added and the plates spun again. Plates were incubated for four days at 25 °C. We placed the plates in humidified secondary containers in humidified incubators to reduce evaporation, which can lead to "edge effects" and differential growth by well location.

3.3. Detection

Following incubation, the plates were stimulated by adding 10 μL phosphate-buffered saline (PBS) containing insulin (Sigma), to a 25 μg/mL final concentration. After 10 min, the stimulated and unstimulated cells were fixed in formaldehyde at 4% final concentration in 10 μL PBS + 0.1% Triton (PBST). As primary and secondary screening often requires parallel processing of many plates at a time, we manipulated the plates such that they were all stimulated and fixed in the same order to ensure uniform activation time. Following a 10 min fixation, the cells were spun for 1′ at 1200 rpm and washed twice with 50–80 μL PBST for a total of at least 30 min of washing. After each wash step and all subsequent steps the plates were spun for 1′ at 1200 rpm.

After the second wash step, the cells were incubated in 10 μL PBST containing 3% bovine serum albumin (BSA) and 750 ng/mL dpERK antibody (Cell Signaling Technology). In order to reduce the number of manipulations required, we fluorescently conjugated the purified mouse monoclonal antibodies using the Alexa Fluor 647 Protein Labeling Kit (Molecular Probes). Following overnight incubation at 4 °C, the cells were again washed twice with 50–80μL PBST and staining intensity read on a fluorescent plate-reader (Analyst GT by Molecular Devices) in 30 μL PBST.

Many dsRNAs in the DRSC collection reduce cell viability [35]. A normalization vector is therefore required to account for general effects on cell number rather than regulatory changes to the relative phosphorylation status of ERK. The cell line we used for primary screening stably expressed *Drosophila* ERK tagged with YFP; this overexpression increased signal:noise of the assay in general and provided a rapid internal normalization for total ERK. For secondary screening, however, we used wild-type cells and normalized with total ERK antibodies. For detection, we measured YFP fluorescence using the following filters (Chroma): excitation 500/20 nm, emission 535/30 nm, and a 445/520 nm double dichroic. For Alexa 647 detection, we used excitation 615/40 nm, emission 650LP, and the 445/520 nm double dichroic, which contains a harmonic window in the far red range. Excitation time was 100 ms/well. Although the YFP channel of the fixed cells provides an approximation of cell number, we also measured the live YFP channel during stimulation time in order to more accurately measure this value before the many plate manipulations later, which may alter true effects on viability.

3.4. Data analysis

Correcting and normalizing the large number of data points from HTS is essential to extracting valid, unbiased genomic information, and isolating the true regulators [36]. Raw fluorescence values for the Alexa 647/dpERK channel and YFP/ERK normalization channel were background subtracted using an average of readings from three blank 384-well plates and the dpERK/ERK ratio was calculated for each well. Since there is considerable geographic variability in screening plates due to plate manufacturing defects and/or differential growth of cells, we corrected for this by dividing the dpERK/ERK ratio by the product of the median of each well's row and column (moving median). To account for plate-to-plate variability, we converted these normalized, corrected ratios to "*Z*-scores," the plate average subtracted from the well value divided by the standard deviation of the plate. DRSC screens are typically performed in duplicate, leading to two such *Z*-scores that are averaged. Occasionally assay noise leads to outliers in one of the two readings. To filter these outliers, we converted the individual *Z*-scores to their rank within the plate, summed the two ranks, and re-selected the top and bottom 1000 wells. These 2000 wells were then re-converted to the *Z*-scores and the two replicates averaged. We chose a *Z*-score cutoff of ±1.5, but this is usually chosen by the screener based on location of the known regulators in the list, the distribution of all amplicons, and the desired number of regulators for follow-up.

3.5. Interpretation

Like traditional forward genetic screens, RNAi HTS can isolate novel regulators of particular signaling pathways. Genes in a primary screen hit list can be validated in secondary screens, and investigators can follow-up on a few

chosen novel regulators with more detailed functional analyses. However, unbiased, genome-wide RNAi screens also provide a glimpse into the systems-level regulation of the particular pathway—how the cell as a whole modulates the precise level of pathway output and develops robustness to perturbation through loss-of-function. These types of analyses are only more recently becoming possible through standardization of statistical techniques and availability of many genome-scale datasets. We have used the functional genomic information from an RTK/ERK signaling screen to observe particular patterns in the dataset, usually enrichment for particular classes. For example, we compared the prevalence of genes conserved in humans or particular gene ontology (GO) functional groups in the original screening collection (essentially the *Drosophila* genome) and in our final primary screen hit list. We determined conservation patterns from the Homologene [37] and InParanoid [38,39] genome comparisons. Statistical techniques to gauge the significance of such patterns, such as tests using the hypergeometric distribution, are common and have been incorporated into some software packages, e.g., [40].

More broadly, RNAi HTS datasets can be combined, for example, with functional expression analyses using microarrays or high-throughput protein–protein interaction (PPI) maps using statistical techniques to model the signaling regulatory network on a larger scale. Such models can help place the individual regulators isolated from a RNAi screen at particular points within the network, uncovering the hierarchical relationships between proteins. Combining this data on a large scale has been successful for *Caenorhabditis elegans* [41] and *Saccharomyces cerevisiae* [42] genomic information and will likely be applied to RNAi HTS in *Drosophila* and mammals once similar datasets are generated.

4. Secondary screening

In most cases, only a s mall number of replicates can be provided per dsRNA for genome-wide screens due to the large number of genes to analyze. In order to validate that the isolated genes are indeed true positives and not due to assay noise, secondary screening is required. Secondary screening of a smaller number of genes also provides the opportunity to begin functional dissection of each gene by using different assays or epistasis tests.

4.1. Gene choice

Superficially, the genes further screened can be the top 200 or so genes isolated. However, it is clear from many HTS performed thus far that the strongest regulators, in addition to known components of the pathway, may also be components of large molecular complexes such as the proteasome or ribosome. These proteins may play a direct role in modulating pathway activity by interacting with the core signaling cassette, or their activities (e.g., protein synthesis) may be non-specifically permissive for signal transduction in general. Components of a single complex likely function similarly in the pathway. Thus using only strength of primary screen hit as the selection criteria may unnecessarily further validate and classify genes that all play very similar and possibly indirect roles in the pathway.

In our primary screen for RTK/ERK regulators, based on our Z-score threshold, we eventually chose ~1700 genes as "hits." This is an unreasonable number to attempt secondary validation and we therefore hand-selected ~375. In selecting a secondary validation set, we chose many of the genes with the strongest effects on the pathway, but, only used representative members of large complexes; for example, we selected a member of the large and small ribosomal complexes, regulatory and catalytic proteasome particles, and a handful of other translational/mRNA processing genes. We also biased towards genes that were conserved in mammals and annotation categories known to be important in signal transduction cascades, e.g., kinases and phosphatases. We resynthesized these cherry picked dsRNAs from the DRSC collection, as described [43].

4.2. Reducing false positives

Although little is understood of their exact mechanism, sequence-specific off-target effects of siRNAs in mammalian cells or dsRNAs in *Drosophila* cells are a known source of false positives in RNAi screens. Consequently, any secondary validation of HTS for MAPK pathways likely should include an estimation of this rate. A reasonable approach to this is to synthesize additional, non-overlapping dsRNAs targeting the same gene, assuming these dsRNAs do not have overlapping off-target effects (for a discussion of OTE, see [44]). Tools for designing and synthesizing new dsRNAs are available online (e.g., [43]).

4.3. Secondary assay format

Unlike in primary screening, secondary screens provide the opportunity to more robustly determine statistical significance of the genes tested by increasing the number of replicates and reducing assay noise. For our secondary screens, we used 5–7 replicates (250 ng/well of each dsRNA) per assay and distributed these replicates randomly in the 384-well screening plates, along with 14–30 controls (a negative control dsRNA such as luciferase). We then ran each assay 2–3 times, providing between 10–21 samples and up to 90 controls per data point. Using this data, we calculated the average reduction or elevation in normalized dpERK and p values for each gene. Because of the large number of samples and conditions tested (see below), these p values are corrected for the multiple hypotheses using the False Discovery Rate often implemented in microarray analysis [45,46]. A Bonferroni correction can also be applied but may be too conservative for secondary screens when such a high number of true positives are expected. A third approach is to empirically determine the significance threshold based on permutation of the data points.

In our specific assay, we adjusted our assays to use wild-type cells rather than ERK-YFP-expressing cells. In order to provide a normalization vector, we used unconjugated mouse monoclonal dpERK antibodies (1:1000, Cell Signaling Technology) and rabbit polyclonal total ERK antibodies (1:1000, Cell Signaling Technology) in the 10 μL PBST + 3% BSA incubation reaction at 4 °C overnight. After washing twice in 50–80 μL PBST, we incubated the secondary screening plates for 1–2 h at room temperature in 10 μL PBST + 3% BSA and the following secondary antibodies: goat anti-rabbit Alexa 488 conjugated (1:1000, Molecular Probes) and goat anti-mouse Alexa 647 conjugated (1:1000, Molecular Probes), along with DAPI to stain cell nuclei (10 μg/mL, Molecular Probes). Following two washes in PBST (50–80 μL), we read the three fluorescent channels on a plate-reader in 30 μL PBST. After every reagent exchange, we spun the plates for 1′ at 1200 rpm. For detection, we again used a 100 ms read time and the following filter sets: for DAPI detection, excitation 360/40 nm, emission 460/50 nm, and a 400 nm dichroic; for Alexa 488, excitation 485/20 nm, emission 530/25 nm, and a 505 nm dichroic; for Alexa 647 detection, excitation 615/40 nm, emission 650LP, and the 445/520 nm dichroic. Similar background subtraction and median corrections were performed and normalized, corrected dpERK/ERK values were used for the final analysis. Normalization to DAPI gave similar results.

4.4. Secondary screen assays

Repeating the screening assay can validate the initial primary screening result with greater statistical confidence. However, because fewer genes and therefore plates are processed in secondary screens, this format is useful to begin a functional classification of genes by conducting different assays. For example, in our assay for RTK/ERK regulators, we have tested different cell lines and multiple ligands stimulating different RTKs. Assays measuring other MAPK cascades could similarly use different cell lines and stimuli, such as different stress sources for p38/JNK, to classify genes as condition-specific or non-specific.

Moreover, any of the assays we suggested above as a primary screen assay could theoretically be used in secondary screens to begin placing genes within the signaling cascade according to function. For example, an ERK transcriptional reporter primary screen could be followed up by secondary screens measuring Ras activation, ERK phosphorylation, and ERK nuclear–cytoplasmic shuttling in a HCS microscopy-based format.

Another approach to functional classification of genes from HTS is to perform epistasis analysis, which has been reviewed in the context of mammalian RNAi screens [16]. We have used this approach previously to order genes within the Wg/Wnt pathway [15]. As in traditional genetic epistasis analyses in model organisms, known components of the pathway with opposing effects on output can be over-expressed or knocked down by dsRNAs during secondary screening. In our dpERK screen, we mixed cells immediately before plating with 250 ng/well of either Ras1 or Gap1 dsRNA, the latter being the Ras GTPase activating protein, a negative regulator of the pathway. Thus, the effects of each gene in combination with either dsRNA are compared to controls also incubated with either dsRNA (Fig. 3A). In principle, positive regulators whose effects are eliminated with Gap1 dsRNA possibly function upstream of Gap1 activity in the pathway and negative regulators whose effects are eliminated with Ras1 dsRNA possibly function upstream of Ras in RTK/ERK signaling (Fig. 3B). However, in contrast to traditional genetic epistasis techniques, RNAi epistasis analyses may be complicated by disparities in the timing of gene knockdown by the two dsRNAs and the hypomorphic nature of RNAi. Thus, the strength of effect of each dsRNA and target stability may be just as important as the hierarchy within the pathway in determining the degree of epistasis. Nevertheless, this approach with multiple members of MAPK cascades may begin to order novel genes within the pathway.

Lastly, this type of analysis can be applied to a primary screen to search for suppressors or enhancers of a particular gene phenotype. We have used this previously, for example, to identify modifiers of a *Pten*-mediated morphological phenotype [22]. This kind of screen may potentially

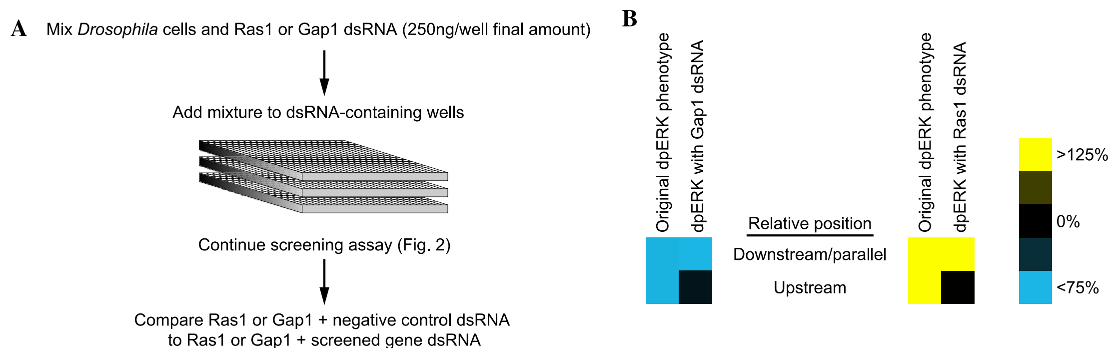


Fig. 3. Ordering genes within the RTK/ERK pathway by RNAi epistasis. (A) Genes validated in secondary screens can be re-screened in the presence of Ras1 or Gap1 dsRNA, and the normalized dpERK values compared to controls also treated with the given dsRNA. (B) Positive regulators which are suppressed by Gap1 dsRNA (left) or negative regulators which are suppressed by Ras1 dsRNA (right) may function upstream of Ras activation in the pathway. Scale is given as percent of negative control (also treated with Ras1 or Gap1 dsRNA) normalized dpERK.

be more sensitive in identifying new components of a given pathway, analogous to traditional genetic screens in sensitized backgrounds originally used to identify components of the Ras pathway (e.g., [47]).

5. Data mining techniques for in vivo validation

HT screening, both primary and secondary, is a powerful approach to identifying and classifying new components of signaling pathways. However, the *in vivo* relevance of the given genes can currently only be tested using "low-throughput" traditional approaches in model organisms. For our *Drosophila* cell-based screens, we tested novel regulators in the context of RTK/ERK signaling-dependent development. *Drosophila* is a well-characterized system with many tools available for genetic dissection of this pathway, as well as other MAPK pathways (e.g., JNK, reviewed in [48]). Genes that validate as physiologically relevant are potentially conserved and of greater importance in regulating the RTK/ERK pathway.

Even in secondary screens, hundreds of genes are analyzed, and testing for physiological relevance *in vivo* can only feasibly be performed in any depth on a few genes. Choosing which genes to analyze from the hundreds of validated regulators is a significant challenge following HT screens. While the strongest novel regulators may seem to be the most likely to test first, we have observed, anecdotally, that these have not necessarily been the strongest *in vivo* regulators. Our approach, outlined below, has been to use additional data sources to identify which genes are the most likely to regulate the core signaling cassette. As we describe above for primary screen data sets, combination of functional genomic and proteomic information can be performed in a more systematic and statistically robust manner given the availability of such resources and analytical tools. Our methods are based on manual curation of public *Drosophila* databases; similar approaches could be applied to mammalian datasets.

5.1. Genetic interactions

Because of the extensive use of *Drosophila* to study RTK/ERK signaling, there are many alleles of pathway components available and a wealth of genetic interactions tested using pathway components. We hypothesized that novel regulators isolated from our screen may have direct or indirect genetic interactions previously tested, given availability of known alleles. Using an online genetic interactions tool [49] that references interactions curated from FlyBase [50], the genetic interactions between known pathway components and all other genes were downloaded. Next, genetic interactions between regulators validated in our secondary screens and all other genes were collected. By cross-referencing these lists, two sets of interactions were generated: (1) previously described direct genetic interactions between known components of the pathway and validated regulators from our screen; and (2) "Transitive" genetic interactions, such as those where a known pathway component and a novel regulator both interact with another gene, "Gene X," implying an potential indirect genetic interaction between the known pathway component and the novel regulator. This list of binary interactions was formatted and visualized using the freely available GraphViz software. These interactions are shown in Fig. 4A, indicating the position of known pathway components and screen hits. In general, however, this analysis was not particularly useful: first, only known genes with alleles can be examined, excluding all truly uncharacterized genes. Second, for those genes with alleles, genetic interactions tested tended to be highly biased towards one or two known pathway components frequently used for such interactions, as evidenced from the large hubs in the genetic interaction network. In addition, automatic curations can be contaminated with failed (negative) genetic interactions as well as true interactions, complicating the analysis.

5.2. Protein–protein interactions

Unbiased datasets could provide more useful information than the biased genetic datasets. At least two large-scale protein–protein interaction (PPI) datasets have been generated for the *Drosophila* proteome [51,52]. Hits isolated from HTS which interact with known components of MAPK pathways may be *bona fide* novel components of the cascades and provide testable hypotheses for future functional characterization. As with the genetic interactions, we downloaded PPI between known pathway components and all other proteins using the Biomolecular Interaction Network Database (BIND) [52] and the FlyNet server [53], which also provides interologs, or predicted interactions, based on *S. cerevisiae* and *C. elegans* screens, which tend to be richer than the *Drosophila* databases. We also downloaded interactions between the regulators validated in our secondary screens and all other proteins using these databases. We assembled a RTK/ERK signaling protein interaction network by combining these two lists and mapping interactions between known pathway components and novel regulators or between known components and a single intervening protein, which also interacts with novel regulators. We visualized this network using GraphViz, as before (Fig. 4B).

These interactions provided insight into the potential functions of novel regulators. Although large-scale PPI maps contain many false positives, we were able to validate some of these interactions by co-immunoprecipitation (A. F. and N. P., unpublished). However, these maps currently also have a very high false negative rate, as evidenced by the low overlap between the two large-scale datasets and the lack of interactions between known components of the pathway extensively validated to bind. As additional interactions are added, converging RNAi and PPI datasets will become more useful for hypothesis

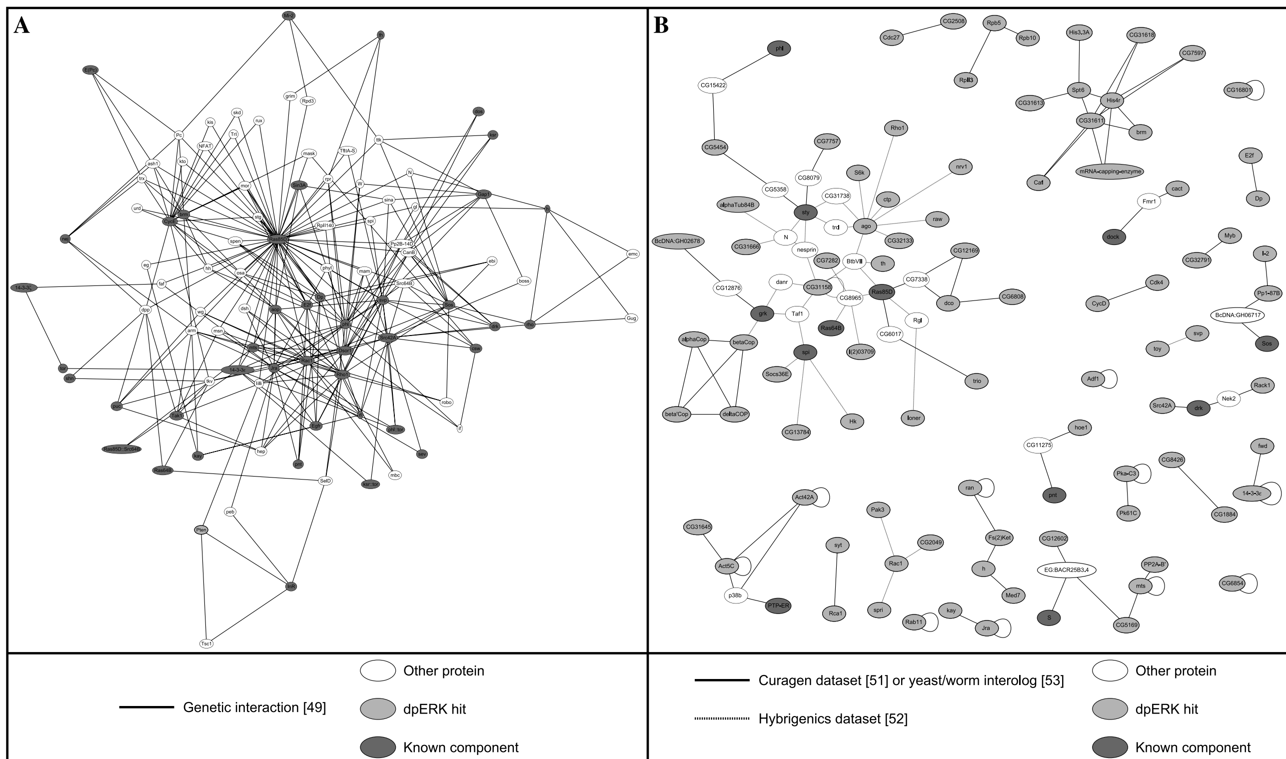


Fig. 4. Examples of genetic and protein–protein interaction data mining. Genes validated in a secondary screen for regulators of RTK/ERK signaling were cross-referenced with databases of genetic interactions (A) or protein–protein interactions (B) as described in the text.

generation and identifying regulators from HTS for *in vivo* follow-up.

5.3. Potential feedback circuits

RTK/ERK signaling is extensively regulated by positive and negative feedback circuits. Novel regulators that are potential components of such circuits may play an important, conserved role in modulating this pathway *in vivo*. We have taken two approaches to identify such genes in a secondary screen list.

Genes that are transcriptionally regulated by pathway activation and also score in RNAi screens may be components of these feedback loops. Using this principle, we downloaded microarray data from two functional expression analyses related to RTK signaling in *Drosophila*: an analysis of transcriptional changes upon Ras^{V12} expression in hemocytes *in vivo* [20] and a comprehensive study of EGFR/Ras signaling components over-expressed or mutated in the *Drosophila* ovary [21]. Similar datasets may be available for other MAPK pathways.

Additionally, we have used the ScanSite web-based platform [54] to identify hits in our screen that contain ERK consensus docking and phosphorylation sites. Genes with these sites may be direct targets of ERK signaling and, as hits in our screen, components of feedback regulation. Although we have not validated the results from either of these two data mining approaches thus far, this information provides additional evidence supporting characterization *in vivo*.

5.4. Other considerations for gene selection

In addition to these data sources, we also considered the strength of the regulation, the pattern or consistency of effect across cell lines and RTK stimuli (with a bias towards consistent regulators), evolutionary conservation (biased towards those with highly conserved orthologs), and the annotation category (e.g., biased towards kinases and phosphatases). In addition, novel information sources such as literature mining tools [55] could also be used for assembling additional data for each gene.

In summary, a data mining approach towards a high-throughput gene list could direct future functional and *in vivo* characterization towards genes with other data sources linking them to the canonical signaling cascade. Other unbiased data sources such as protein–protein interaction maps and microarray analyses provide the most useful information for this approach. Although we have begun validating a few genes identified through this process *in vivo*, whether using other data sources can truly improve the *in vivo* validation rate compared to that using RNAi screen data alone awaits a large-scale validation effort.

6. Future directions

We have described a high-throughput approach to dissecting the ERK MAPK cascade using a phospho-specific antibody-based RNAi screening platform in *Drosophila*

cells. This approach, which can be extended to many components of other MAPK cascades, provides a significant step beyond traditional methods using forward genetic screens and mammalian cell biology to build the canonical core signaling cassette. Combination of this data with other primary or secondary screens, enhancer/suppressor analysis by co-RNAi, and curationg by data mining of other sources of information on the pathway, should support a systems-level understanding of cellular regulation of this pathway. Similar approaches could be applied to mammalian RNAi screens of MAPK cascades and gain-of-function studies in *Drosophila* or mammalian cell culture. Novel regulators of MAPK cascades isolated from such screens may be attractive targets of drug discovery. Moreover, for existing drugs, combining genome-wide RNAi screens with chemical treatment may help identify their target proteins or pathways [56] as has been demonstrated in *S. cerevisiae* [57–59].

Acknowledgments

We thank L. Kockel for pioneering the phospho-specific antibody HTS approach in our laboratory and B. Mathey-Prevot for critical manuscript review. A.F. is supported by a Medical Scientist Training Program (MSTP) grant. N.P. is an investigator of the Howard Hughes Medical Institute.

References

[1] J. Schlessinger, Cell 103 (2000) 211–225.
[2] I. Rebay, Dev. Biol. 251 (2002) 1–17.
[3] G. Pearson, F. Robinson, T. Beers Gibson, B.E. Xu, M. Karandikar, K. Berman, M.H. Cobb, Endocr. Rev. 22 (2001) 153–183.
[4] G.L. Johnson, R. Lapadat, Science 298 (2002) 1911–1912.
[5] S. Madhusudan, T.S. Ganesan, Clin. Biochem. 37 (2004) 618–635.
[6] R.A. Hilger, M.E. Scheulen, D. Strumberg, Onkologie 25 (2002) 511–518.
[7] M.R. Green, N. Engl. J. Med. 350 (2004) 2191–2193.
[8] S. Sasagawa, Y. Ozaki, K. Fujita, S. Kuroda, Nat. Cell Biol. 7 (2005) 365–373.
[9] T.G. Smith, D. Sweetman, M. Patterson, S.M. Keyse, A. Munsterberg, Development 132 (2005) 1305–1314.
[10] A.S. Yoo, C. Bais, I. Greenwald, Science 303 (2004) 663–666.
[11] M. Iwanami, Y. Hiromi, M. Okabe, Genes Cells 10 (2005) 743–752.
[12] W. Xiong, J.E. Ferrell Jr., Nature 426 (2003) 460–465.
[13] K. Nybakken, S.A. Vokes, T.Y. Lin, A.P. McMahon, N. Perrimon, Nat. Genet. 37 (2005) 1323–1332.
[14] G.H. Baeg, R. Zhou, N. Perrimon, Genes Dev. 19 (2005) 1861–1870.
[15] R. DasGupta, A. Kaykas, R.T. Moon, N. Perrimon, Science 308 (2005) 826–833.
[16] J. Moffat, D.M. Sabatini, Nat. Rev. Mol. Cell Biol. 7 (2006) 177–187.
[17] S.H. Yang, A.D. Sharrocks, A.J. Whitmarsh, Gene 320 (2003) 3–21.
[18] T.L. Tootle, P.S. Lee, I. Rebay, Development 130 (2003) 845–857.
[19] E.M. O'Neill, I. Rebay, R. Tjian, G.M. Rubin, Cell 78 (1994) 137–147.
[20] H. Asha, I. Nagy, G. Kovacs, D. Stetson, I. Ando, C.R. Dearolf, Genetics 163 (2003) 203–215.
[21] K.C. Jordan, S.D. Hatfield, M. Tworoger, E.J. Ward, K.A. Fischer, S. Bowers, H. Ruohola-Baker, Dev. Dyn. 232 (2005) 709–724.
[22] A. Kiger, B. Baum, S. Jones, M. Jones, A. Coulson, C. Echeverri, N. Perrimon, J. Biol. 2 (2003) 27.
[23] G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A.J. Levine, M.B. Elowitz, U. Alon, Nat. Genet. 36 (2004) 147–150.
[24] A. Sorkin, M. McClure, F. Huang, R. Carter, Curr. Biol. 10 (2000) 1395–1398.
[25] N. Mochizuki, S. Yamashita, K. Kurokawa, Y. Ohba, T. Nagai, A. Miyawaki, M. Matsuda, Nature 411 (2001) 1065–1068.
[26] A. Harding, T. Tian, E. Westbury, E. Frische, J.F. Hancock, Curr. Biol. 15 (2005) 869–873.
[27] M.J. Clague, S. Urbe, J. Cell Sci. 114 (2001) 3075–3081.
[28] S.R. Hingorani, D.A. Tuveson, Curr. Opin. Genet. Dev. 13 (2003) 6–13.
[29] B. Nicke, J. Bastien, S.J. Khanna, P.H. Warne, V. Cowling, S.J. Cook, G. Peters, O. Delpuech, A. Schulze, K. Berns, J. Mullenders, R.L. Beijersbergen, R. Bernards, T.S. Ganesan, J. Downward, D.C. Hancock, Mol. Cell 20 (2005) 673–685.
[30] T.F. Westbrook, E.S. Martin, M.R. Schlabach, Y. Leng, A.C. Liang, B. Feng, J.J. Zhao, T.M. Roberts, G. Mandel, G.J. Hannon, R.A. Depinho, L. Chin, S.J. Elledge, Cell 121 (2005) 837–848.
[31] J.C. Clemens, C.A. Worby, N. Simonson-Leff, M. Muda, T. Maehama, B.A. Hemmings, J.E. Dixon, Proc. Natl. Acad. Sci. USA 97 (2000) 6499–6503.
[32] R. Schweitzer, M. Shaharabany, R. Seger, B.Z. Shilo, Genes Dev. 9 (1995) 1518–1529.
[33] P. Duchek, K. Somogyi, G. Jekely, S. Beccari, P. Rorth, Cell 107 (2001) 17–26.
[34] S. Armknecht, M. Boutros, A. Kiger, K. Nybakken, B. Mathey-Prevot, N. Perrimon, Methods Enzymol. 392 (2005) 55–73.
[35] M. Boutros, A.A. Kiger, S. Armknecht, K. Kerr, M. Hild, B. Koch, S.A. Haas, H.F. Consortium, R. Paro, N. Perrimon, Science 303 (2004) 832–835.
[36] N. Malo, J.A. Hanley, S. Cerquozzi, J. Pelletier, R. Nadon, Nat. Biotechnol. 24 (2006) 167–175.
[37] HomoloGene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene (2006).
[38] M. Remm, C.E. Storm, E.L. Sonnhammer, J. Mol. Biol. 314 (2001) 1041–1052.
[39] InParanoid, http://inparanoid.cgb.ki.se/ (2006).
[40] P. Hong, W.H. Wong, BMC Bioinformatics 6 (2005) 20.
[41] S. Li, C.M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P.O. Vidalain, J.D. Han, A. Chesneau, T. Hao, D.S. Goldberg, N. Li, M. Martinez, J.F. Rual, P. Lamesch, L. Xu, M. Tewari, S.L. Wong, L.V. Zhang, G.F. Berriz, L. Jacotot, P. Vaglio, J. Reboul, T. Hirozane-Kishikawa, Q. Li, H.W. Gabel, A. Elewa, B. Baumgartner, D.J. Rose, H. Yu, S. Bosak, R. Sequerra, A. Fraser, S.E. Mango, W.M. Saxton, S. Strome, S. Van Den Heuvel, F. Piano, J. Vandenhaute, C. Sardet, M. Gerstein, L. Doucette-Stamm, K.C. Gunsalus, J.W. Harper, M.E. Cusick, F.P. Roth, D.E. Hill, M. Vidal, Science 303 (2004) 540–543.
[42] I. Lee, S.V. Date, A.T. Adai, E.M. Marcotte, Science 306 (2004) 1555–1558.
[43] Drosophila RNAi Screening Center, www.flyrnai.org (2006).
[44] C.J. Echeverri, N. Perrimon, Nat. Rev. Genet. (2006).
[45] Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani, Behav. Brain Res. 125 (2001) 279–284.
[46] Y. Benjamini, D. Yekutieli, Genetics 171 (2005) 783–790.
[47] M.A. Simon, D.D. Bowtell, G.S. Dodson, T.R. Laverty, G.M. Rubin, Cell 67 (1991) 701–716.
[48] L. Kockel, J. Zeitlinger, L.M. Staszewski, M. Mlodzik, D. Bohmann, Genes Dev. 11 (1997) 1748–1758.
[49] I. Flockhart, M. Booker, A. Kiger, M. Boutros, S. Armknecht, N. Ramadan, K. Richardson, A. Xu, N. Perrimon, B. Mathey-Prevot, Nucleic Acids Res. 34 (2006) D489–D494.
[50] FlyBase, http://www.flybase.org (2006).
[51] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y.L. Hao, C.E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C.A. Stanyon, R.L. Finley Jr., K.P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R.A.

Shimkets, M.P. McKenna, J. Chant, J.M. Rothberg, Science 302 (2003) 1727–1736.

[52] E. Formstecher, S. Aresta, V. Collura, A. Hamburger, A. Meil, A. Trehin, C. Reverdy, V. Betin, S. Maire, C. Brun, B. Jacq, M. Arpin, Y. Bellaiche, S. Bellusci, P. Benaroch, M. Bornens, R. Chanet, P. Chavrier, O. Delattre, V. Doye, R. Fehon, G. Faye, T. Galli, J.A. Girault, B. Goud, J. de Gunzburg, L. Johannes, M.P. Junier, V. Mirouse, A. Mukherjee, D. Papadopoulo, F. Perez, A. Plessis, C. Rosse, S. Saule, D. Stoppa-Lyonnet, A. Vincent, M. White, P. Legrain, J. Wojcik, J. Camonis, L. Daviet, Genome Res. 15 (2005) 376–384.

[53] Flynet Server, http://macbeth.clark.jhu.edu/cgi-bin/flynew.pl (2005).

[54] J.C. Obenauer, L.C. Cantley, M.B. Yaffe, Nucleic Acids Res. 31 (2003) 3635–3641.

[55] L.J. Jensen, J. Saric, P. Bork, Nat. Rev. Genet. 7 (2006) 119–129.

[56] U.S. Eggert, A.A. Kiger, C. Richter, Z.E. Perlman, N. Perrimon, T.J. Mitchison, C.M. Field, PLoS Biol. 2 (2004) e379.

[57] K. Baetz, L. McHardy, K. Gable, T. Tarling, D. Reberioux, J. Bryan, R.J. Andersen, T. Dunn, P. Hieter, M. Roberge, Proc. Natl. Acad. Sci. USA 101 (2004) 4525–4530.

[58] G. Giaever, P. Flaherty, J. Kumm, M. Proctor, C. Nislow, D.F. Jaramillo, A.M. Chu, M.I. Jordan, A.P. Arkin, R.W. Davis, Proc. Natl. Acad. Sci. USA 101 (2004) 793–798.

[59] P.Y. Lum, C.D. Armour, S.B. Stepaniants, G. Cavet, M.K. Wolf, J.S. Butler, J.C. Hinshaw, P. Garnier, G.D. Prestwich, A. Leonardson, P. Garrett-Engele, C.M. Rush, M. Bard, G. Schimmack, J.W. Phillips, C.J. Roberts, D.D. Shoemaker, Cell 116 (2004) 121–137.