

RNA interference: Mechanisms and Applications

Norbert Perrimon, Tony Hyman

23-28 June 2004

In the early 1990s, the startling discovery that small RNAs were able to inhibit the expression of various genes in plants, an observation subsequently extended to animal cells, triggered a flurry of studies that have now completely changed our view of the RNA world. Previously, RNA was considered as a mere intermediate between DNA and proteins, however, the characterization of small classes of RNA molecules, which include small interfering RNAs (siRNAs) and micro RNAs (miRNAs), has revealed that these molecules play a role in a number of basic cellular processes such as developmental control, protection of genomes from foreign DNA, and chromatin organization. The incredible properties of siRNAs have also provided novel tools to create efficient gene knock down, and in combination with the availability of full genome sequences opened up the road to large scale functional genomic projects. Finally, to add to the excitement of this burgeoning field, siRNAs are currently being tested as potential therapeutic tools. Participants in the meeting "RNA interference: Mechanisms and Applications" funded by the Foundation Les Treilles, discussed new exciting findings on the mechanisms underlying RNAi and the various applications that this new methodology has to offer. In addition, a number of presentations discussed ongoing technological efforts that will bring to the field new and more efficient RNAi-based methods. Finally, extensive discussions on how emerging data from RNAi screens should be presented and eventually integrated with other genomic approaches were the subject of much debate.

The completion of the genome sequences of a number of organisms, including yeast, *C. elegans*, *Drosophila* and humans has given scientists the opportunity, for the first time, to look comprehensively at the information globally. The increasing number of fully sequenced genomes from a wide variety of species not only permits the study of gene functions in specific organisms, but also allows one to use the evolutionary conservation between species to obtain important insights into the functional organization of cellular pathways. Thus, the most critical step following the completion of full genome sequences is to develop methodologies that allow the systematic and rapid analysis of the information generated by these large-scale projects.

siRNAs and longer double stranded RNAs (dsRNAs) have emerged in the past few years as exquisite tools to create efficient gene knock down. Various methods have been developed to deliver these reagents to the cells of interest. In the nematode *Caenorhabditis elegans*, specific inactivation of genes containing homologous sequences by RNAi can be achieved *in vivo* by three different techniques: injection of the dsRNA, feeding worms bacteria expressing target-gene dsRNA, and soaking of the worm in a dsRNA solution. In *Drosophila*, dsRNAs can be either injected into the embryo, or expressed as transgenes. In addition, dsRNAs can be efficiently delivered to *Drosophila* cells simply by bathing them into a solution of dsRNAs. These methods have now been widely applied to address specific cell biological and signal transduction questions in both organisms. The efficacy and simplicity of the methods has now allowed the development of highthroughput genome wide screens.

Interestingly, in *C. elegans* dsRNA-induced silencing spreads from the cell(s) where it is initiated to silence the targeted gene throughout the animal. This is most likely mediated by the transport of dsRNA between cells. To understand how dsRNA can be transported into and between cells **Craig Hunter** (Harvard, Boston) and his colleagues constructed a transgenic strain that allows direct visualization of systemic silencing and then used the strain to identify mutants that specifically disrupt systemic silencing without disrupting autonomous RNAi. His group identified over 200 mutants that define five complementation groups. Approximately half the mutants define the gene *sid-1*, and another third the gene *sid-2*. Mechanistic studies indicate that *sid-1* functions as a passive channel for the transport of dsRNA into cells, while *sid-2* functions as a receptor for the uptake of dsRNA from the environment. Understanding of the mechanism of dsRNA intake will have important practical applications as some cell lines, including mammalian cells, do not intake dsRNAs.

A number of large-scale screens based on RNAi have now been performed in *C. elegans* using either injection into embryos or feeding the worm with bacteria expressing dsRNAs. **Tony Hyman** (Max Planck Institute, Dresden) and his colleagues have conducted a genome-wide screen using the one cell stage *Caenorhabditis elegans* embryo as a model system to identify a large number of genes required for cell division processes. In this approach, expression of a given gene in the *C. elegans* embryo is specifically silenced via microinjection of corresponding dsRNA in the gonad of the mother. The resulting one cell stage embryos are then analyzed with time-lapse microscopy for defects in cell division processes, thus identifying genes required for proper cell division. Using these approaches,

together with Cenix Bioscience, Tony Hyman and his colleagues have identified 667 genes required for a successful first division of the embryo. Altogether, their data indicate that at least 80% of the genes that could be identified by classical loss of function genetics have been found using the RNAi approach.

Using a different delivery method of dsRNA, **Julie Ahringer** (Wellcome, Cambridge) and her colleagues constructed a library of 16,757 dsRNA-expressing bacterial strains, each designed to target a single predicted gene by RNAi when fed to worms. This feeding technique, pioneered by Timmons and Fire, allows for highthroughput RNAi screening. The library is a reusable resource that can be used for an unlimited number of genetic screens. Initially, they identified 1722 genes with a lethal or visible RNAi phenotype; ~1200 of these were not previously known. By carrying out new screens with different assays the library is now being used for systematic identification of genes involved in different cellular processes.

John Kim (MGH, Boston), in collaboration with Harrison Gabel, Ravi Kamath, Scott Kennedy and Gary Ruvkun, combined an RNAi-bacteria feeding library consisting of 16,757 genes from the Ahringer laboratory with an additional 1,924 unique RNAi clones identified in the Vidal RNAi feeding library. Using the resulting library of 18,681 bacterial strains, which is predicted to individually target 94% of all *C. elegans* genes, they performed a comprehensive screen for *C. elegans* genes required for RNAi. To visualize defects in RNAi, they created a *C. elegans* strain that has a transgene containing a snapback GFP construct introduced into a worm strain expressing GFP under an endogenous promoter (the RNAi sensor strain). When an essential component of the RNAi machinery is silenced, the expression of GFP is de-repressed, resulting in increased levels of GFP that can be visualized using conventional fluorescent microscopy. Therefore, this strain acts as an *in vivo* sensor of RNAi activity. By screening the combined RNAi library with the RNAi sensor strain, they have identified 88 clones that reproducibly result in an RNAi-deficient phenotype upon silencing, including 51 essential genes. Ten of the clones encode genes previously implicated in an RNAi-like phenomena. Among the 71 new factors are 2 Piwi/PAZ proteins, 4 DEAH helicases, 28 RNA binding/processing factors, 11 chromatin factors, 2 MAP kinase pathway members, 3 transcription factors, and 10 nuclear import/export factors. They are also pursuing complementary forward-genetic approaches, as well as recapitulating the RNAi screen in an RNAi-enhanced *eri-1* background, to further expand the constellation of factors that affect small RNA-mediated

processes in worms. Further study of these genes and their interactors should also provide a more complete understanding of the mechanism of RNAi *in vivo*, and may help to identify new connections between RNAi and related biological processes in *C. elegans*

Norbert Perrimon (Harvard Medical School, Boston), **Michael Boutros** (DKFZ, Heidelberg), **Amy Kiger** (Harvard Medical School, Boston) and **Buzz Baum** (UCL, London) presented their work on using *Drosophila* cells to identify gene functions. The Perrimon laboratory in collaboration with Dr. Renato Paro's group (Heidelberg Germany) has generated a library of 21,300 dsRNAs directed against all predicted open reading frames (ORFs). This resource can now be used to conduct highthroughput cell-based RNAi screens to identify genes involved in various assays. dsRNA molecules readily enter *Drosophila* cells in culture, and lead to partial or complete elimination of the corresponding cellular protein. They use robotic technologies to screen the dsRNAs rapidly in specific cell-based assays. These screens provide a powerful methodology to identify rapidly all the proteins encoded by the genome that interfere with a specific assay. For example, a screen that uses luciferase-based read outs and plate readers can be completed in less than a week.

This screening platform is being used to analyze the complexity of signal transduction pathways. Many of the routes that integrate signals have been identified. In particular, much information has been obtained on receptor tyrosine kinases, TGF- β , Wnt, Hedgehog, Notch, JAK/STAT and NF κ B signaling pathways. Although there is a good knowledge of the basic structure of these pathways, it is clear that many components remain to be identified. A number of genome wide RNAi screens to identify new components of these pathways are being conducted. These screens are based on a number of assays that include transcriptional read outs, antibodies against phospho-specific epitopes, and morphological changes. Comparisons of the data sets obtained from these screens allow the identification of factors that are common to many pathways (i.e., ubiquitination, nuclear import, etc), as well as, those that are more pathway-specific. Further, additional information on the structure of the pathways can be obtained by determining the epistatic relationships between the various components of a pathway. This is accomplished by adding combinations of dsRNAs on the cells, and then by analyzing their effects in the read out. Altogether, the analysis of these pathways using RNAi in tissue culture cells generates specific hypotheses that can be validated *in vivo* using standard approaches.

One of the challenges of highthroughput RNAi screens is that they require a relatively expensive and sophisticated technology. A solution to

this problem was presented by **Bernard Mathey-Prevot** (Harvard Medical School, Boston) who discussed the establishment of an academic centralized facility for genome-wide screen in *Drosophila* cells at Harvard Medical School (*Drosophila* RNAi Screening Center, DRSC; <http://flyrnai.org>). The goal of DRSC is to make this technology available to the community. The rationale for having a centralized RNAi screening center is as follows: 1. DRSC makes relatively expensive and sophisticated technology accessible to all interested in the community. Since DRSC is generously supported by NIGMS, the cost to a visiting researcher associated to the screen is minimal and limited mostly to supplies; 2. DRSC diminishes the variables associated with screens, permitting valuable functional comparisons across many studies; and 3. It allows the creation of a database of information that will be eventually made public by submission to Flybase or other community resources in a standardized format. Researchers interested in genome-wide RNAi screening submit a short application to be reviewed by a DRSC committee that will assess the scientific merit of the screen. The long term scientific direction and challenges faced by the Center was discussed in the context of the evolving technology and emergence of siRNA screens in the mammalian systems.

Although screens in the 384 well plate format are successful and effective, they are quite costly. Thus, new technological advances would be welcome to facilitate the application of RNAi. **Anne Carpenter** (Whitehead, Boston) provided an overview of her work as well as other ongoing work in David Sabatini laboratory at the Whitehead Institute for Biomedical Research. In 2001, the Sabatini lab published the development of 'living cell microarrays' in which each feature of the array is a cluster of 100-1000 cells expressing a cDNA or small interfering RNA (siRNA). The arrays can be made at densities of up to 7,000 cell clusters per standard slide and can be screened with any technique that is compatible with cells grown on a surface, including immunofluorescence, autoradiography, and *in situ* hybridization. The most recent unpublished advances have extended this technology to small molecules, lentivirally-expressed shRNA, and *Drosophila* dsRNA. Anne Carpenter showed the proof-of-principle results from these new applications, including the testing of a 400-gene set of dsRNAs for *Drosophila* RNA interference to investigate cell growth signaling, a process linked to several human diseases. The lab plans to have genome-wide screens in *Drosophila* under way shortly using living cell microarrays and is working on techniques to generate genome-wide mammalian RNAi cell microarrays.

Highthroughput screens that rely on microscopy require imaging hardware that allows the rapid collection of thousands of high-resolution images of cells. Extracting information from these images by eye is impossible due to the large numbers of images to be analyzed. In addition, visual analysis is only useful for identifying unusual samples rather than recording a quantitative measurement for each sample, and subtle but important differences between samples are undetectable. Whereas genetics has typically relied on dramatic phenotypes, where normal physiology is grossly disturbed, there is likely valuable functional information present amongst phenotypes that are less dramatic. However, available image analysis software suffers from limitations. Towards this goal, Anne Carpenter developed CellProfilerTM, a cell image analysis software that allows biologists without training in computer vision or programming to quantitatively measure complex cellular phenotypes from thousands of images in a highthroughput manner. The software is modular and therefore can be adapted to any cell type and to any visual phenotype. She showed preliminary results of the identification of nuclei and cells from *Drosophila* and Human HeLa cells. The software will be made available to interested collaborators this year and will be freely available to academics after publication early next year, in the spirit of previous widely useful software projects like NIH Image.

Results from RNAi screens in tissue culture need to be eventually validated *in vivo*. In *Drosophila*, extensive collections of transposon-induced mutations are already available for approximately half of the genes. In addition, **Barry Dickson** (IMP, Vienna) described his progress at building a collection of transgenic flies that express hairpin constructs directed against every ORFs. The ambitious goal of creating more than 13,000 UAS-hairpin lines is well under way and the preliminary results of the efficacy of the methodology are encouraging. This resource will allow the analysis of gene function in cells and tissues of interest as the hairpin is under UAS control and thus can be expressed precisely using the appropriate Gal4 line.

Although RNAi provides a powerful tool for functional genomics, RNAi results in a reduction of target mRNA levels and hence loss-of-function phenotypes. However, nearly all proteins have multiple functional domains. Hence it is often difficult or impossible to independently assess the multiple functions of individual proteins by using RNAi. To address this challenge **Craig Mello** (Worcester) has undertaken a large-scale chemical mutagenesis of *C. elegans* to identify temperature-sensitive mutations that perturb development. Their analysis of several phenotypic categories identified in this screen suggests that the collection of ~4000 mutant strains

is likely to include alleles in 10 to 25% of the essential *C. elegans* genes. For example, four of 12 known Wnt-signaling components were found in the screen. Importantly, analysis of the alleles defined in the screen includes those sought-after non-null alleles that alter specific domains within important regulatory proteins. For example, alleles of CDK-1 and its binding partner CKS-1 were identified that do not perturb the cell-cycle, but result in the stabilization of a specific target protein, and a consequent change in a polarized cell division. The goal is that in the future conditional genetics will help to put the flesh on the skeleton of interactions defined by proteomics and RNAi-based functional genomics.

Another means of interfering with gene function is to block protein activity using small molecules. However, one of the drawbacks of chemical genetics is that small molecules usually lack specificity and identification of the compound targets is a major challenge. **Ulrike Eggert** (ICCB, Boston) presented her efforts at identifying genes involved in *Drosophila* cytokinesis using both small molecules and RNAi screening. She performed a parallel chemical genetic and genome-wide RNAi screens in *Drosophila* cells, identifying 50 small molecule inhibitors of cytokinesis and 224 genes important for cytokinesis, including a new protein in the Aurora B pathway. By comparing small molecule and RNAi phenotypes, she identified a small molecule that inhibits the Aurora B kinase pathway.

The large amount of data generated from RNAi screens need to be stored into databases that allow data mining and further integration with other data sets generated from other "omic" approaches such as protein-protein interaction and expression genomics. **Kris Gunsalus** and **Fabio Piano** (NYU, New York) presented their RNAi database (RNAiDB) and progress at integrating functional genomic data. In particular excellent progress has already been made toward a phenome map of *C. elegans* embryogenesis. The integration of RNAi data sets with protein networks were discussed by **Marc Vidal** (DFCI, Boston), **Joel Bader** (John Hopkins, USA), and **Stan Letvosky** (Harvard, Boston). Joel bader described his efforts to use biological networks as frameworks for understanding the information generated by full-genome RNAi screens. Protein interaction networks provide a valuable framework, and his group has mapped ~10-15% of the protein interaction network of *Drosophila* using the yeast two-hybrid method. They enriched the map by transferring information cross-species from yeast and worm screens. Extracting molecular machines (ribosome, proteasome, etc.) hit by RNAi screens is feasible; extracting more subtle sub-networks remains a challenging area for algorithm development and network mapping. Towards this end, they are preparing a release of human

protein interactions identified from ~280,000 bait-prey pairs from two-hybrid screens. Finally, Joel Bader described work with collaborators in yeast (Jef Boeke, Forrest Spencer) to use multi-gene perturbations to map networks. These experiments use yeast knockouts to identify the cumulative effect of removing two genes from a network; these experiments are the yeast analog of two-gene RNAi experiments that are now becoming feasible. An important conclusion is that genetic interactions identified through this approach are usually orthogonal to the typical conception of a pathway. Thus, rather than linking genes that have a genetic interaction, they link genes with a shared pattern of genetic interaction partners (i.e., gene deletions that give similar phenotypes in a series of different genetic backgrounds). These patterns, signatures, or motifs, are similar to the phenoclusters described by others. They have generated a series of these motifs from yeast synthetic lethality data. Remarkably, these motifs are predictive of phenotypes distinct from synthetic lethality. For example, they show that distance from a motif defined by known landmark genes provides a high-quality predictor of quantitative nuclear migration defect rates.

Although most of the meeting was centered on RNAi applications in *C. elegans* and *Drosophila*, two talks from **Josh LaBaer** (Harvard Medical School, Boston) and **Andy McMahon** (Harvard, Boston) discussed ongoing efforts in their own laboratories and the community at applying RNAi methodologies both in mammalian cells and *in vivo* in the mouse. Josh LaBaer discussed approaches to upregulate and downregulate gene expression, which are accomplished by overexpression or ectopic expression of the gene in the former and RNAi mediated inhibition of gene expression for the latter. In order to test the effect of gene overexpression on cell behavior, Josh LaBaer described his efforts at developing full-length cDNA clones in a protein expression-ready format. Highthroughput gene capture cloning has been used to create a growing repository of human genes in recombinational cloning vectors, which allow the highthroughput transfer of the coding regions into virtually any expression vector for use in a broad variety of experiments. One challenge for this repository especially notable for mammalian genes regards the identification and naming of genes. Often multiple names or symbols are attached to the same gene and the same name is given to more than one gene. Thus, when users request genes, it is often difficult to determine if the requested gene is present in the collection because it may be present under a different name. A generic searching tool was developed to overcome this disambiguation challenge. It is based upon reducing any gene query to cDNA sequence which is then used to BLAST

the relevant databases to find any potential matches. All matches are presented, along with the alignments.

A crucial challenge of building clone collections of mammalian cDNAs is quality control. Given the time and cost of doing mammalian experiments, the reagents must be well-characterized. The highthroughput cloning methods are by nature mutagenic. For this reason, multiple clonal isolates are selected for each attempted gene and these are then analyzed to find the best clone (least mutations). The process of full-length sequence verification is tedious and time consuming. Software has been developed to automate the process of comparing clone sequences to expected reference sequences. The software takes into account the quality of the sequence reads, the effects of discrepancies on the predicted protein sequence and whether or not the discrepancy correlates with a polymorphism. The software will automatically sort the clones based on user specifications into several bins: automatically accepted clones, automatically rejected clones and clones that need manual review.

Early screening experiments in mammalian cells have looked for genes that can induce cancer-like behavior in immortalized mammary epithelial cells. Several different behaviors were measured including inducing cell proliferation, cell migration, and altered morphogenesis of acinar structures in three dimensional cell cultures. To test a set of genes with high likelihood of scoring in these screening experiments, a collection of 1000 clones representing human genes linked to breast cancer was produced.

To identify genes linked to human diseases, the MedGene database was produced, which examines all entries in the Medline database and looks for publications that co-cite diseases and genes. Thus, all disease-gene relationships published in abstracts or titles are represented in MedGene. Users can search the database to find all genes associated with any disease. A statistical score was devised to rank the output based on the strength of the linkage. Users can also submit lists of genes to the database and sort them based on their linkage to a disease. MedGene has identified over 2400 genes that have been linked to breast cancer. A new version of MedGene has been produced that links gene to other biological concepts such as signal transduction, cell cycle control and drug response. Approximately 300 of the breast cancer 1000 genes have been tested by retroviral introduction into mammary epithelial cells and hits have been scored in all of the assays. The hit rate was approximately 10% in these assays. These hits include genes already known to play a role in the tested processes as well as novel candidates for these activities. New assays are underway to test the effect of

gene expression on the response of cells to drugs. These experiments involve the development of good retroviral expression vectors and automation of the cell culture methods.

Andy McMahon described the current state of genetic approaches for functional analysis of mammalian development. Current methods need to be complemented by alternative approaches if the goal is rapid, "highthroughput" screening of multiple genes in a specific process. Initial applications of RNAi-based methodologies in the developing mouse suggest that these strategies will be an effective genetic tool in this model. Generating conditional RNAi approaches and focusing on embryo stem cell based RNAi transmission is projected to enable the temporal and/or spatial specific analysis of gene activity significantly faster, albeit with lower fidelity, than the existing gold-standard, homologous recombination mediated gene-targeting.

In conclusion, the meeting provided a comprehensive description of the current applications of RNAi-based methodologies. The general feeling was that the major applications of RNAi are now well defined and that the next few years will see an explosion of results that will emerge from this revolutionary tool. Clearly, one of the issues that need to be dealt urgently deals with the proper management of the large amount of data that are emerging from RNAi studies. Coordination between the various RNAi databases that are being built will be necessary to ensure that data generated from various organisms can be integrated in meaningful ways. Further, the combination of powerful statistical analyses and data sets of high quality will provide excellent means to build correlation between gene products and contribute to our understanding of protein networks present in the cell.