

RNA-Mediated Interference as a Tool for Identifying Drug Targets

Nigel J. O'Neil, Rowena L. Martin, Matthew L. Tomlinson, Martin R. Jones,
Alan Coulson and Patricia E. Kuwabara

The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, England

Contents

| | |
|--|----|
| Abstract | 45 |
| 1. Drug Discovery and Genomics | 45 |
| 2. The Power of RNAi and <i>Caenorhabditis elegans</i> Functional Genetics | 46 |
| 3. Background: The Biology of <i>C. elegans</i> | 46 |
| 4. The Genome of <i>C. elegans</i> | 47 |
| 5. <i>C. elegans</i> is a Valid Model for Elucidating Gene Function in Vertebrates | 47 |
| 6. The Mechanism Underlying RNAi | 47 |
| 7. Practical Applications of RNAi | 49 |
| 8. High-Throughput RNAi Facilitates the Global Analysis of Gene Function | 50 |
| 9. Direct Applications of RNAi for Drug Discovery | 50 |
| 10. Corroborating RNAi Phenotypes by Obtaining a Knockout Deletion Mutant | 51 |
| 11. Perspectives and Conclusions | 51 |

Abstract

The nematode *Caenorhabditis elegans* is the first multicellular organism with a fully sequenced genome. As a model organism, *C. elegans* is playing a special role in functional genomic analyses because it is experimentally tractable on many levels. Moreover, the lessons learned from *C. elegans* are often applicable across phyla because many of the key biologic processes involved in development and disease have been well conserved. Many global approaches for analysing gene activity are being pursued in *C. elegans*. RNA-mediated interference (RNAi) is an efficient high-throughput method to disrupt gene function. The basic technique of RNAi involves introducing sequence-specific double-stranded RNA into *C. elegans* in order to generate a nonheritable, epigenetic knockout of gene function that phenocopies a null mutation in the targeted gene. This technique drastically reduces the time needed to jump from the identification of an interesting gene sequence to achieving an understanding of its function. Thus, RNAi facilitates the high-throughput functional analysis of gene targets identified during drug discovery. RNAi can also help to identify the biochemical mode of action of a drug or pesticide and to identify other genes encoding products that may respond or interact with specific compounds.

1. Drug Discovery and Genomics

The rapid growth in numbers of genomic and cDNA sequences deposited in public and private databases in recent years has led to the development of new avenues of exploration and *in silico* approaches for drug discovery. Drug discovery can now be divided into 2 principle categories: 'discovery genetics' and 'discovery genomics'.^[1] Discovery genetics relies on using positional cloning to identify disease-related susceptibility genes. In turn the molecular identification of these disease-associated genes combined with knowledge of the developmental or meta-

bolic pathways that they affect may help researchers to focus their search for compounds that ameliorate disease. The advantage of this scheme is that targets are pre-validated because of their association with a known disease. Unfortunately, this route of gene discovery is limited because the positional cloning of human disease genes has been a lengthy process and relatively few genes are correlated with known diseases. In addition, genetic susceptibility to disease is often polygenic. Moreover, the molecular identification of a disease gene may be initially uninformative because of the absence of associated function.

'Discovery genomics' takes pre-existing information regarding genes known to play important roles in human development or disease and uses their sequences to find other genes or gene families that share extensive sequence similarity. Examples of the types of gene families offering potential for drug discovery include nuclear hormone receptors, G protein-coupled receptors, kinases, and caspases.^[2] Discovery genomics also encompasses other genome-based techniques that are used to identify genes that have similar expression profiles to disease markers or genes encoding proteins that interact physically with one another (fig. 1). Comparative genomic techniques used to analyse gene expression profiles include differential display, semi-quantitative reverse transcriptase polymerase chain reaction (RT-PCR), DNA microarrays, and serial analysis of gene expression (SAGE).^[3-6] Commercially available libraries provide researchers with information about genes that are differentially expressed in normal versus mutant cells.^[7] Protein interaction data is generally obtained by yeast 2-hybrid interaction screens or by co-immunoprecipitation of protein complexes followed by matrix-assisted laser desorption/ionization (MALDI) to identify the constituent components.^[8] Taken together, *in silico* approaches for mining databases and other genomic approaches have provided numerous potential drug targets. However, researchers are still left with the burden of proof to validate the role of these genes in disease processes.

2. The Power of RNAi and *Caenorhabditis elegans* Functional Genetics

This review will focus on the nematode *Caenorhabditis elegans* as a model organism for biomedical studies and highlight the importance of RNAi for learning about gene function. The availability of the whole genome sequence of *C. elegans* has created the challenge to understand the function of each predicted gene and its global role in organismal biology. RNAi has developed into one of the most indispensable tools for analysing gene function in *C. elegans* because it is rapid, efficient, and can be scaled up for high-throughput analyses.^[9-11] There are many examples in the literature where RNAi has been used to target individual genes or groups of genes for knockout to gain clues about gene activity.^[12-18] The basic technique of RNAi involves introducing sequence-specific double-stranded RNA (dsRNA) into *C. elegans* in order to generate a nonheritable, epigenetic knockout of gene function that phenocopies a null mutation in the targeted gene.^[9,19] RNAi can also play a crucial role in drug discovery. RNAi can be used to: (i) validate potential drug targets; (ii) investigate the biologic role of validated targets; and (iii) screen for active compounds.

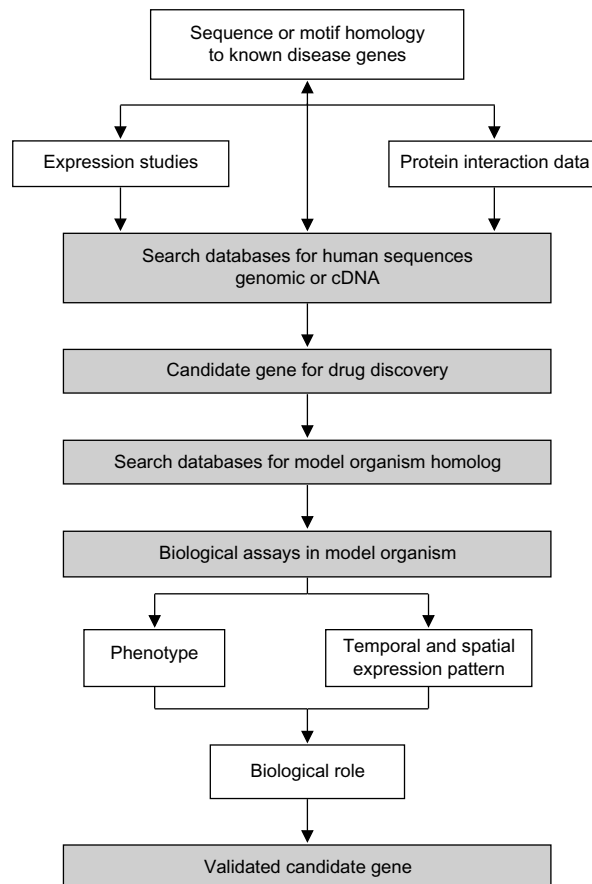


Fig. 1. Flow-chart illustrating how the use of functional genomics can facilitate the identification and validation of drug targets. Genomic data collected from sequence databases, expression studies and protein interaction experiments help to identify potential drug targets. The validation of these targets is facilitated by using model organisms to obtain phenotypic and biologic information about the vertebrate homolog.

3. Background: The Biology of *C. elegans*

C. elegans has many attributes that have helped to establish its role as an excellent model organism for biomedical studies. There are 2 sexes, self-fertilizing hermaphrodites and males. Each hermaphrodite produces approximately 350 embryos that reach maturity in 3 and a half days at 20°C. Adult animals are only one millimetre in length. Worms can be raised on agar plates or in liquid, which can be advantageous for drug-screening. It is also possible to maintain frozen worm stocks in liquid nitrogen.

The biology of *C. elegans* has been extensively characterized. *C. elegans* is transparent during all life-stages so it has been possible to determine the cell lineage.^[20] The invariant cell lineage is comprised of only 959 somatic cells in the adult hermaphrodite and 1031 somatic cells in the male. Although there are relatively

few somatic cells, these cells form epidermal, intestinal, excretory, nerve and muscle tissue. Over the past 3 decades numerous studies have added to our knowledge of biologic processes because *C. elegans* is highly amenable to genetic analysis. A wide variety of mutants have been identified, carrying mutations in genes affecting cell migration, programmed cell death, movement, cell lineage, defecation, longevity, neurobiology, morphology, susceptibility to drugs, behaviour, viability and many other biological processes.^[21] The identification of the genes affected by these mutations and the ability to place them in regulatory pathways have provided major insights into developmental and disease processes.

4. The Genome of *C. elegans*

C. elegans is the first metazoan with a completely sequenced genome.^[22] The small 100 Mb genome of *C. elegans* is distributed across 6 chromosomes and is predicted to encode 19 705 proteins.^[23] Many of these proteins share significant similarity to proteins responsible for human development and disease. Present estimates indicate that as many as 75% of human disease genes have a *C. elegans* homolog.^[22,24] Therefore, many of the potential drug target candidates found in humans are also likely to be represented in the *C. elegans* genome, making it possible to take advantage of the biology of *C. elegans* to validate potential targets and to screen compounds for activity. Table I lists some human disease genes with *C. elegans* homologs.

5. *C. elegans* is a Valid Model for Elucidating Gene Function in Vertebrates

The biology of *C. elegans* has been used successfully to gain insights into the role of the human presenilin genes and their involvement in Alzheimer's disease. Mutations in human presenilin genes *PS1* and *PS2* are responsible for early onset Alzheimer's disease. It was found that the *C. elegans* homologs of *PS1* and *PS2* are encoded by the *sel-12* and *hop-1* genes.^[25,26] *sel-12* was initially identified as a suppressor of an activated form of *lin-12*, which encodes a homolog of the Notch receptor. By itself, the *sel-12* mutant displays an egg laying defective (Egl) phenotype, characterized by bloating due to the inability to expel developing embryos from the uterus.^[27] This phenotype is obviously different than the human phenotype of neurodegeneration, nonetheless it has been found that presenilins from both organisms play similar roles. For example, the human *PS1* and *PS2* can substitute functionally for *sel-12* in *C. elegans*, establishing that *PS1* and *PS2* proteins share a similar function to SEL-12.^[28] Mutant *PS1* alleles that are associated with Alzheimer disease in humans retained partial activity when introduced into *C. elegans*,

suggesting that they are reduced function alleles. Finally, both SEL-12 and the human presenilin appear to play roles in proteolytic processing: SEL-12 affects the intracellular localization and possibly the processing of the LIN-12/Notch-receptor, whereas *PS1* and *PS2* are involved in the proteolytic processing of amyloid precursor protein (APP).^[29,30] This example serves to illustrate that although the phenotypes obtained in the model organism may differ from the human disease phenotype, it is still possible to use model organisms to investigate gene activity.

6. The Mechanism Underlying RNAi

RNAi has emerged as one of the most important tools for elucidating gene function, and efforts to understand the mechanism and physiological function underlying this fascinating phenomenon are providing a stimulating challenge. Study of the *par-1* gene, an essential gene for viability, provided the first evidence that direct injection of RNA molecules into the gonad syncytium of *C. elegans* could lead to inhibition of gene activity throughout the animal.^[31] It was shown that injection of antisense *par-1* RNA produced a mutant phenocopy that mimics the embryonic lethality caused by the *par-1* null mutation. Surprisingly, sense strand RNA also elicited a *par-1* mutant phenocopy. The puzzling observation that either antisense or sense strand RNA could interfere with normal gene activity was subsequently resolved. Fire et al.^[9] showed that double stranded RNA (dsRNA) was responsible for the RNAi effect; it was speculated that dsRNA was generated as a contamination during the *in vitro* synthesis of single-stranded RNA. Several observations indicate that RNAi in *C. elegans* acts at the post-transcriptional level. First, promoter sequences and introns do not elicit RNAi effects.^[9] Second, cDNA sequences appear to be more potent in eliciting RNAi effects than a mixture of intronic and exonic sequences.^[32,33] Third, RNAi can lead to degradation of the target gene mRNA.^[34] Moreover, in *Drosophila* it has been observed that dsRNA is processed into segments 21-23 nucleotides in length, and that these fragments may be guiding the mRNA cleavage.^[35] Finally, there is no indication that the nucleotide sequence of the RNAi targeted gene is modified by processes such as methylation.^[34]

By taking advantage of the genetic tractability of *C. elegans*, mutants have been identified that are defective in RNAi;^[36] these genes have been named *rde-1*, *rde-2*, *rde-3* and *rde-4* (RNA interference-deficient). The molecular identification of the mutated genes has shed additional light on the mechanisms underlying the RNAi phenomenon. Mutations in the *mut-2* and *mut-7* (*mutator*) genes, which lead to the mobilization of transposons in the germline, also reduce the RNAi effect.^[37] Similarly, *rde-2* and *rde-3* mutants, but not *rde-1* and *rde-4*, also have the property

Table I. *Caenorhabditis elegans* homologs of human genes associated with disease. All *C. elegans* proteins corresponding to predicted *C. elegans* open reading frames (*C.e.* ORFs) and corresponding genes (*C.e.* gene) are listed in the Wormpep33 database^a

| Disease | <i>C.e.</i> ORF | <i>C.e.</i> gene | Disease | <i>C.e.</i> ORF | <i>C.e.</i> gene |
|---|-------------------------------|-------------------------------|---------------------------------------|--|------------------|
| Adenylosuccinate lyase deficiency | R06C7.5 | | Lipoid congenital adrenal hyperplasia | ZK177.5 | |
| Alpha-ketoglutarate deficiency | T22B11.5 | | Lowe oculocerebrorenal syndrome | JC8.10 | |
| Alphamethylacetoaceticaciduria | T02G5.8 T02G5.7 | | Lissencephaly | C05D2.6 | |
| Alzheimers disease | F35H12.3 C18E3.8 | <i>sel-12</i> <i>hop-1</i> | Lysinuric protein Intolerance | F27C8.1 | |
| Angiokeratoma | R07B7.11 | | Megaloblastic anemia I | ZC116.3 | |
| Aspartylglucosaminuria | R04B3.2 | | Methemoglobinemia | T05H4.5 T05H4.4 | |
| B cell lymphoma | T07C4.8 | <i>ced-9</i> | Methylmalonicaciduria | ZK1058.1 | |
| Beckwith-Wiedemann syndrome | Y38H8A.5 F45B8.4 ZK84.1 | <i>pag-3</i> | Mevalonicaciduria | Y42G9A.4 | |
| Beta-methylcrotonyl-glycinuria | F32B6.2 | | Molybdenum cofactor deficiency | F49H6.5 | |
| Cortisol 11-beta ketoreductase deficiency | T25G12.7 T25G12.2 | | Nephronophthisis type 1 | M28.7 | |
| Cystic fibrosis | F14F4.3B | | Neurofibromatosis type II | F42A10.2 | |
| Dihydropyrimidinuria | C47E12.8 R06C7.3 | | Niemann-Pick disease type C | F09G8.4 F02E8.6 | |
| Duchenne/Becker muscular dystrophy | F15D3.1 | <i>dys-1</i> | Ornithinemia | C16A3.10 | |
| Dystrophia myotonica I | K08B12.5 | | Osteogenesis imperfecta congenita | F01G12.5A K04H4.1 | |
| Fructosemia | T05D4.1 F01F1.12 | | Phenylketonuria | K08F8.4 | |
| GABA-transaminase deficiency | K04D7.3 | | Phenylketonuria II | T03F6.1 | |
| Galactose epimerase deficiency | C47B2.6 | | Prolidase deficiency | K12C11.1 | |
| Gangliosidosis type I | T19B10.3 | | Propionicaciduria type I | F27D9.5 | |
| Gaucher disease type I | F11E6.1 Y4C6B.6 | | Pseudovitamin D deficiency | Y17G9B.3 | |
| Giant cell hepatitis | ZC8.1 | | Rickets | | |
| Glutaricaciduria type IIC | C05D11.12 | <i>let-721</i> | Pyroglutamicaciduria | M176.2 | |
| Glutathionuria | C53D5.5 | | Refsum disease | ZK550.6 | |
| Granulomatous disease | F53G12.3 | | Sarcosinemia | Y106G6H.5 | |
| Homocystinuria | C06A8.1 | | SSADH deficiency | F45H10.1 | |
| Hyperglycerolemia | R11F4.1 | | Thymine-uraciluria | C25F6.3 | |
| Hyperlysinemia | R02D3.1 | | Trimethylaminuria | H24K24.5 Y39A1A.19 K08C7.2 K08C7.5 F53F4.5 | |
| Hyperoxaluria type I | T14D7.1 | | Vitamin D-resistant rickets | F11A1.3 | |
| Hypokalemicalkalosis | Y37A1C.1 | | Wernicke-Korsakoff syndrome | F01G10.1 | |
| Isovalericacidemia | C02B10.1 | | Xanthinuria | F55B11.1 | |
| King syndrome | T02C5.5 | <i>unc-2</i> | Xeroderma pigmentosum | F53A3.2 | |
| Lecithin-cholesterol acyltransferase deficiency | M05B5.4 | | | | |
| Leigh syndrome | T05H10.6 | | | | |
| Lesch-Nyhan syndrome | Y105E8B.J | | | | |

^a ORF names correspond to the cosmid or YAC clone carrying a given gene. If more than one gene is present on a cosmid or YAC, each is given a unique identifying number (n). *C.e.* gene names are provided for those ORFs associated with a known genetic mutation.

of mobilizing transposons in the germline. Based on these data it has been hypothesised that RNAi may function physiologically as a defence mechanism designed to prevent transposon activation or viral invasion.^[37]

The *rde-1* gene encodes a novel member of a gene family that includes the translational initiation factor eIF2c, suggesting a link between RNAi and translation.^[36] The *mut-7* gene encodes a protein with homology to bacterial RNaseD, which is implicated in mRNA degradation.^[37] Another gene identified as essential for RNAi in *C. elegans* is *ego-1*, which encodes a homolog of RNA-directed RNA polymerases.^[38] The involvement of *ego-1* in *C. elegans* RNAi provides a link between the gene silencing mechanisms called quelling in *Neurospora crassa* and post-transcriptional gene silencing (PTGS) in plants. It is hypothesized that these phenomena share a conserved core mechanism that may depend on the generation of dsRNA by RNA-directed RNA polymerases.^[39]

7. Practical Applications of RNAi

On a practical level, RNAi is one of the most powerful techniques available for probing gene function in *C. elegans*. All that is required is a DNA template from which to synthesize dsRNA. Sequence similarity should be considered when deciding on a dsRNA template because cross-interference can occur when genes share significant (greater than 70%) nucleotide similar-

ity.^[40] To enhance specificity, unique nucleotide sequences of the mRNA should be targeted, such as the 3' UTR. It is generally preferable to use cDNA sequences as templates for generating dsRNA because dsRNA generated from genomic DNA may affect the penetrance of the RNAi effect,^[32,33] however, in the absence of cDNA, exon-rich genomic DNA regions can be used. It is often convenient to generate DNA templates of 250 to 2000 bp for *in vitro* transcription using the polymerase chain reaction (PCR) and gene-specific primers flanked with T3 and/or T7 RNA polymerase promoter sequences (fig. 2). Subsequently, RNA can be generated *in vitro* using one of the many commercially available kits containing T3 or T7 RNA polymerase; the products generated are annealed to form dsRNA. Agarose gel electrophoresis is used to assess the efficiency of the synthesis and annealing reactions.

RNAi effects have been elicited in worms using 4 different methods (fig. 2). The first technique involves microinjecting dsRNA into the syncytial gonad of adult animals.^[9] This method requires a knowledge of *C. elegans* anatomy, specialized equipment, and skill in microinjection. Surprisingly, it has been found that the RNAi effect is capable of crossing tissue boundaries, hence the exact location of microinjection may not be crucial.^[9]

Non-invasive methods for eliciting RNAi include soaking and feeding (fig. 2) Both techniques simplify the administration of dsRNA and can facilitate high-throughput RNAi analyses

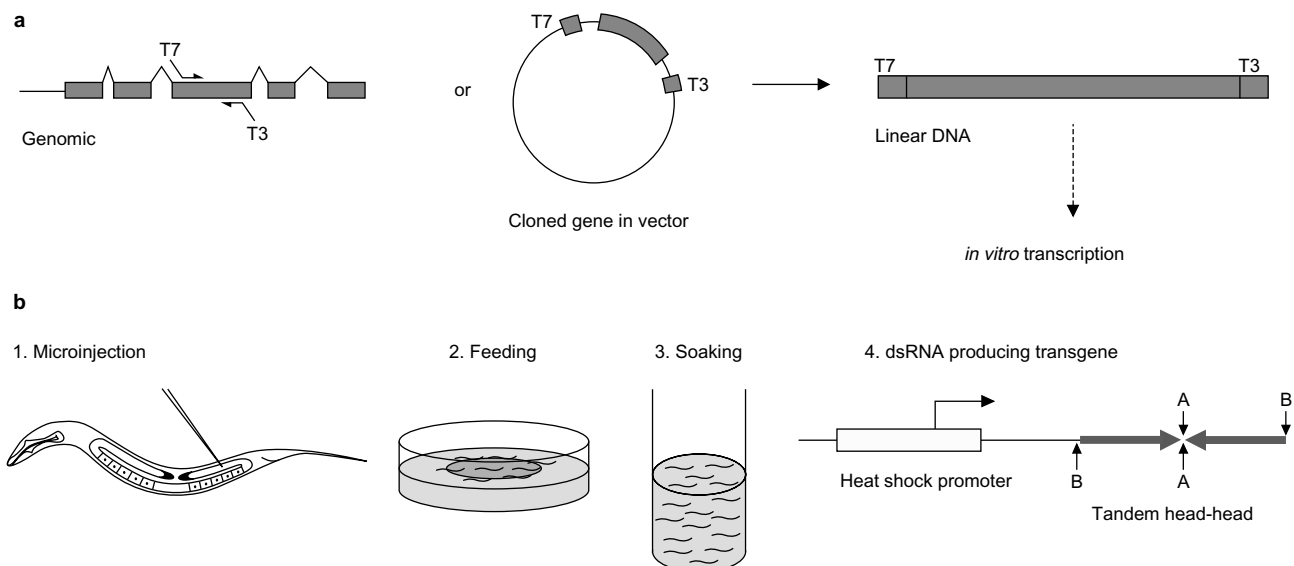


Fig. 2. Generating dsRNA for performing RNAi in *Caenorhabditis elegans*. **(a)** The DNA source for generating *in vitro* transcription templates can be genomic DNA or a plasmid carrying either genomic or cDNA sequence. DNA templates for *in vitro* transcription are conveniently generated by polymerase chain reaction (PCR) using sequence-specific primers flanked with promoter sequences for T7 or T3 RNA polymerase. dsRNA is generated by *in vitro* transcription. **(b)** Methods for performing RNAi. RNAi is performed by administering dsRNA using (1) microinjection (2) feeding worms bacteria expressing both sense and anti-sense RNA (3) soaking worms in dsRNA solutions or (4) expressing a 'snap-back' dsRNA from a transgene under the control of an inducible promoter.

(discussed in section 8). When a worm is soaked in a solution of dsRNA, it ingests the dsRNA, which is subsequently absorbed by intestinal cells to produce an RNAi response.^[15] It has also been shown that RNAi effects can be elicited by feeding worms *E. coli* carrying plasmids expressing both sense and anti-sense RNA driven from inducible promoters.^[41] This technique has the advantage that the costs associated with synthesizing dsRNA are reduced and the bacteria carrying the dsRNA producing plasmids become renewable resources that can be used for multiple purposes.

The methods for applying RNAi thus far discussed produce only transient phenotypes. However, it is possible to generate RNAi inducing transgenes that are stably inherited in *C. elegans*. This method was developed to investigate the mutant phenotypes of genes that play a role in neuronal development.^[42] Many neuronal genes had been refractory to other forms of RNAi. Tavernarakis et al.^[42] expressed dsRNA in neuronal cells utilizing a transgene under the control of the heat-shock promoter (fig. 2). The snap-back transgene carries 2 copies of a gene-specific sequence oriented in a head-to-head inverted repeat; the RNA molecule expressed from this transgene is thus capable of folding back on itself to generate dsRNA. The heat shock promoter drives expression of the transgene in most somatic cells, including neurons. The direct expression of dsRNA in neurons appears to overcome the RNAi resistance previously encountered using other modes of RNAi. This result suggests that the RNAi effect may be unable to cross neuronal cell boundaries.^[42]

As observed with neuronal genes, the method by which dsRNA is delivered to cells and the nature of the gene being studied may affect the ability to detect an RNAi phenotype. Thus it is advisable to attempt different methods for applying RNAi if a mutant phenocopy is not immediately apparent. For genes that act in the embryo, injection into the syncytial gonad of the parent may be the most efficient method for eliminating both maternal and zygotic gene activities in the progeny. For genes expressed later in development, it has been reported that feeding results in 50% more post-embryonic phenotypes than injection.^[10] The increased efficacy of feeding may be due to the fact that the animals have a continuous input of dsRNA while feeding.

There are additional considerations to note when performing RNAi. Although RNAi usually leads to the loss of gene function, this loss of activity may range from a partial to a complete loss of activity. In addition, tissues may reveal a differential susceptibility to RNAi. It has also been reported that dsRNA templates derived from different regions of the same gene may produce a gradation of RNAi responses.^[18]

8. High-Throughput RNAi Facilitates the Global Analysis of Gene Function

The suitability of RNAi as a method for analysing genes on a global scale has been demonstrated in 2 independent reports that used RNAi to explore the mutant phenotypes of all predicted genes on chromosome I^[10] and chromosome III.^[11] The studies used different methods to administer the dsRNA: Fraser et al.^[10] administered dsRNA from feeding vectors expressing dsRNA corresponding to each predicted gene on chromosome I, whereas Gonczy et al.^[11] microinjected dsRNA corresponding to each predicted gene on chromosome III. Both reported similar levels of success in detecting mutant phenotypes; 13.9% of tested genes produced a visible phenotype. Gonczy et al.^[11] also found that only 1 or 2 dsRNA species should be included per microinjection to prevent repression of the RNAi phenomenon. The relative efficiency of RNAi in producing mutant phenotypes was tested by comparing 50 genes with known loss of function phenotypes to the RNAi results. 31 of the 50 genes tested resulted in a visible RNAi phenotype and 81% of these phenotypes matched the published phenotype for a corresponding mutant.^[10] Embryonic phenotypes were also found to be more susceptible to RNAi than post-embryonic phenotypes. 19 of 21 tested embryonic genes exhibited an RNAi phenotype whereas only 14 of 31 post-embryonic genes exhibited an RNAi phenotype. These studies establish that RNAi is an effective method for screening large numbers of genes to obtain functional information.

9. Direct Applications of RNAi for Drug Discovery

RNAi phenotypes form the basis for screening compound libraries in order to identify compounds that enhance or ameliorate specific phenotypes (fig. 3). For example, dsRNA directed against the *C. elegans* RAD51 homolog results in increased germline apoptosis.^[43] Mutations in the DNA checkpoint/repair enzyme *mrt-2* prevent *Ce-rad-51*-induced apoptosis. Therefore, it is now possible to screen for compounds that antagonise the DNA checkpoint/repair pathway by screening for compounds that ameliorate the apoptotic phenotype of *Ce-rad-51(RNAi)*.

In a similar manner, it is possible to screen for compounds that phenocopy or mimic an RNAi phenotype. In the case of the *C. elegans* presenilin homolog, SEL-12, it is possible to screen for compounds that phenocopy the *sel-12* egg laying defective phenotype. Compounds resulting in an egg laying block may act in the same pathway as the presenilins and function as inhibitors of presenilin.

The availability of the entire *C. elegans* genomic sequence coupled with the ability to generate dsRNA against any given sequence provides researchers with the opportunity to generate

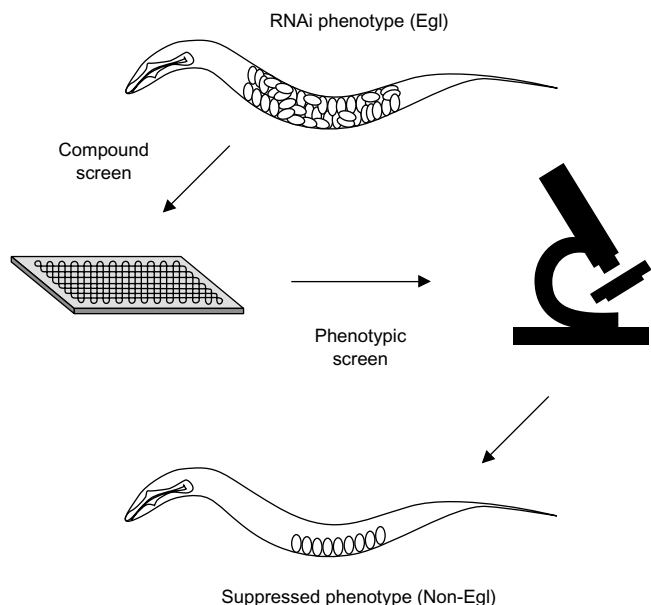


Fig. 3. Screening compounds that affect a specific *Caenorhabditis elegans* phenotype. Mutant worms displaying an easily scored phenotype, such as egg-laying defective (Egl), are targets for assaying the ability of pharmaceuticals to ameliorate or enhance the selected phenotype. Mutants can be generated either genetically or by using RNAi. Worms are easily propagated in multiwell plates in order to facilitate the parallel screening of multiple compounds.

global bacterial libraries expressing dsRNA corresponding to all predicted genes in the *C. elegans* genome. Such libraries can be used for multiple purposes. For example, worms grown on these bacteria can be treated with a compound that elicits a known phenotype in order to search for genes that enhance or suppress the activity of the compound. It also becomes possible to identify loci that are resistant to the effects of a drug or pesticide and in turn, the identification of these resistant genes can provide rapid information about the possible mode of action of the compound. Screens have been performed using classical genetics to identify mutants that are resistant to the action of drugs such as fluoxetine^[44] and the anthelmintic ivermectin.^[45] In the case of fluoxetine, resistant mutants were identified that defined a novel gene family encoding multipass transmembrane proteins. These results may have clinical implications relating to the mechanism of fluoxetine action and may identify new targets for drug discovery.

10. Corroborating RNAi Phenotypes by Obtaining a Knockout Deletion Mutant

Presently, *C. elegans* lacks an efficient system for site-directed homologous recombination and the search for a gene deletion can be laborious, especially when compared to RNAi. However, once sufficient incentive is generated to pursue a target gene of inter-

est, it is usually advisable to generate a chromosomal gene knockout before pursuing further study. Additionally, in cases where there are few potential targets of clear importance it may be more expedient to obtain a genetic knockout of the target gene rather than utilizing RNAi. In the case of Niemann-Pick type C disease, there are only 2 homologs of the human *NPC1* gene, designated *npc-1* and *npc-2*.^[46] Instead of using RNAi to phenocopy the two *C. elegans* homologs, which might have proven difficult due to functional redundancy or cross-interference, Sym et al.^[46] generated knockout mutations in both *npc-1* and *npc-2* using the PCR-mediated gene knockout technique described below.

The primary strategy for generating a gene knockout uses high-throughput PCR to detect a chromosomal rearrangement in the gene of interest by screening complex combinatorial libraries of worms (fig. 4). Normally the libraries are divided in half in order to prepare DNA and also to save the remaining live worms as frozen stocks. This strategy was initially developed to detect the presence of transposon insertions in the target gene of interest using PCR and a combination of gene-specific and transposon-specific primers. The desired insertion/deletion worm was identified after multiple rounds of sibling selection in which pools testing positive for a rearrangement were continuously subdivided into pools until the desired single worm carrying the mutant allele was found.^[47] Unfortunately, transposon insertions are often silent because they preferentially integrate in A/T rich sequences, such as introns. Therefore, this method often requires a second step to screen for an imprecise excision event that either removes exonic sequences or creates a frameshift mutation.

To circumvent the need to perform two rounds of PCR screening in order to find a gene rearrangement that produces a mutant phenotype, a direct mutagenesis approach is now being followed. The mutagen trimethylpsoralen (TMP) coupled with long wave ultraviolet light (UV) appears to be the most effective method for generating small deletions.^[48] Thus, combinatorial libraries of TMP/UV mutagenized worms have been generated.^[48] A deletion allele is detectable during the PCR screening of these libraries because the synthesis of the shorter deletion product is generally favoured over the longer wild-type product. Moreover, because deletions are generated at random, it is usually possible to screen for a specific deletion event and to recover a single worm carrying this allele (fig. 4).

11. Perspectives and Conclusions

As more potential drug targets are uncovered by genomic approaches, techniques must be developed to rapidly validate good candidate targets. Model organisms such as *Drosophila*, *S. cerevisiae*, and *Caenorhabditis elegans* will play a significant role in

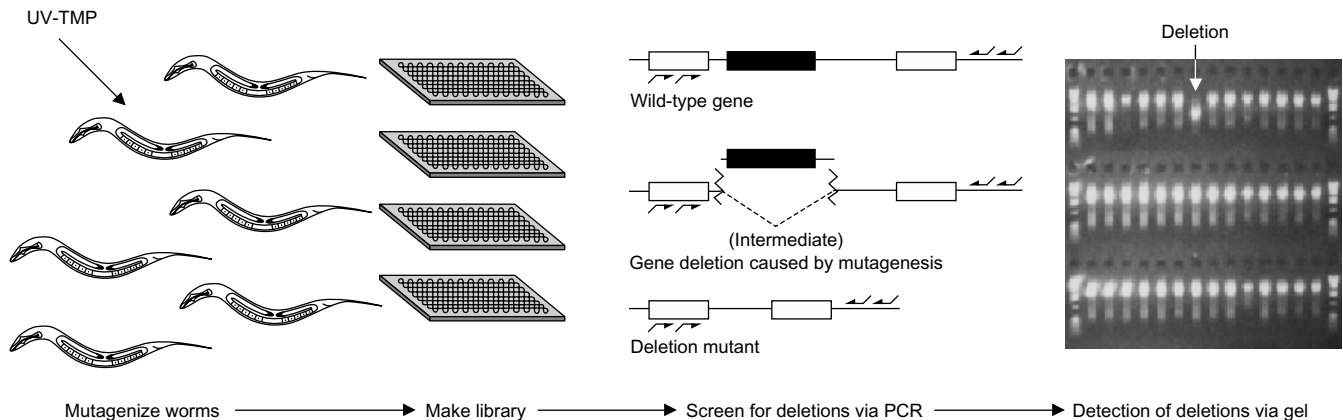


Fig. 4. A strategy for identifying a *Caenorhabditis elegans* deletion mutant. The F1 progeny of animals mutagenized with trimethyl psoralen (TMP) and long wave UV light (UV) are cultured in multiwell plates to generate a worm library. Half of the library is used to generate genomic DNA and the remaining half is stored to maintain a population of live worms. The isolated DNA is pooled combinatorially and PCR is performed using nested primers to identify a deletion in the targeted gene. The deletion allele is detected by agarose gel electrophoresis as a band that migrates with higher mobility than the wild-type allele. A single worm carrying the deletion allele is isolated by sib-selection from the worm library address identified from the PCR screen.

validating these targets. The sequenced genome of *C. elegans* has led to several global techniques for assaying the biological roles of the genes encoded in the genome, including microarray analysis of gene expression,^[49] whole genome RNAi screening,^[10,11] genome-wide 2 hybrid analyses of protein interactions^[50] and high-throughput screening for PCR mediated gene knockouts.^[48] The high throughput capacity of RNAi makes it a particularly attractive method for rapidly screening and validating targets identified by protein-protein interaction, or microarray analyses or *in silico* gene predictions. With such a robust validation system available, it is possible to use RNAi to determine the gene function of entire classes of predicted proteins with the potential to be involved in human disease, for example phosphatases, kinases, or nuclear hormone receptors. In conclusion, if a *C. elegans* homolog of a disease associated gene exists, then there is a strong likelihood that the biology of *C. elegans* may be advantageous in providing information crucial for drug discovery and development.

Acknowledgements

We thank the MRC Laboratory of Molecular Biology, the Biotechnology and Biological Sciences Research Council and Janssen Pharmaceutica N.V. for support.

References

- Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000; 405: 857-65
- Cockett M, Dracopoli N, Sigal E. Applied genomics: integration of the technology within pharmaceutical research and development. *Curr Opin Biotechnol* 2000; 11: 602-9
- Madden SL, Wang CJ, Landes G. Serial analysis of gene expression: from gene discovery to target identification. *Drug Discov Today* 2000; 5 (9): 415-25
- Zweiger G. Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends Biotechnol* 1999; 17 (11): 429-36
- Rae FK, Stephenson SA, Nicol DL, et al. Novel association of a diverse range of genes with renal cell carcinoma as identified by differential display. *Int J Cancer* 2000; 88 (5): 726-32
- Dasmahapatra AK, Wimpee BA, Trewin AL, et al. Demonstration of 2,3,7,8-tetrachlorodibenzo-p-dioxin attenuation of P450 steroidogenic enzyme mRNAs in rat granulosa cell in vitro by competitive reverse transcriptase-polymerase chain reaction assay. *Mol Cell Endocrinol* 2000; 164 (1-2): 5-18
- Walker MG, Volkmut W, Spinzak E, et al. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. *Genome Res* 1999; 9: 1198-1203
- Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 2000; 405: 837-46
- Fire A, Xu S, Montgomery MK, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998; 391 (6669): 806-11
- Fraser AG, Kamath RS, Zipperlen P, et al. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 2000; 408: 325-30
- Gonczy P, Echeverri C, Oegema K, et al. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 2000; 408: 331-6
- Chase D, Serafinas C, Ashcroft N, et al. The Polo-like kinase PLK-1 is required for nuclear envelope breakdown and the completion of meiosis in *Caenorhabditis elegans*. *Genesis* 2000; 26: 26-41
- Boxem M, Srinivasan DG, van den Heuvel S. The *Caenorhabditis elegans* gene *ncc-1* encodes a cdc2-related kinase required for M phase in meiotic and mitotic cell divisions, but not for S phase. *Development* 1999; 126: 2227-39
- Fay DS, Stanley HM, Han M, et al. A *Caenorhabditis elegans* homologue of hunchback is required for late stages of development but not early embryonic patterning. *Development Biol* 1999; 205: 240-53
- Tabara H, Grishok A, Mello CC. RNAi in *C. elegans* – soaking in the genome sequence. *Science* 1998; 282: 430-1
- Moore LL, Morrison M, Roth MB. HCP-1, a protein involved in chromosome segregation, is localized to the centromere of mitotic chromosomes in *Caenorhabditis elegans*. *J Cell Biol* 1999; 147: 471-9

17. Shi Y, Mello C. A CBP/p300 homolog specifies multiple differentiation pathways in *Caenorhabditis elegans*. *Genes Dev* 1998; 12: 943-55
18. Kuwabara PE. Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. *Genetics* 1996; 144: 597-607
19. Fire A, Albertson D, Harrison SW, et al. Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development* 1991; 113: 503-14
20. Sulston J. Cell lineage. In: Wood WB, editor. *The nematode Caenorhabditis elegans*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1988: 123-55
21. Riddle DL, Blumenthal T, Meyer BJ, et al. editors. *C. elegans II*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1997
22. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998; 282 (5396): 2012-8
23. Available from: URL: http://www.sanger.ac.uk/Projects/C_elegan/wormpep/download.shtml [Accessed Feb 2001]
24. Kuwabara PE, O'Neil NJ. The use of functional genomics in *C. elegans* as it relates to studying human development and disease. *J Inherit Metabol Dis*. In press
25. Levitan D, Greenwald I. Facilitation of *lin-12*-mediated signalling by *sel-12*, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. *Nature* 1995; 28 (377): 351-4
26. Li X, Greenwald I. HOP-1, a *Caenorhabditis elegans* presenilin, appears to be functionally redundant with SEL-12 presenilin and to facilitate LIN-12 and GLP-1 signalling. *Proc Natl Acad Sci U S A* 1997; 94 (22): 12204-9
27. Levitan D, Doyle TG, Brousseau D, et al. Assessment of normal and mutant human presenilin function in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 1996; 93 (25): 14940-4
28. Levitan D, Greenwald I. Effects of SEL-12 presenilin on LIN-12 localization and function in *Caenorhabditis elegans*. *Development* 1998; 125: 3599-606
29. Xia W, Zhang J, Ostaszewski BL, et al. Presenilin 1 regulates the processing of beta-amyloid precursor protein C-terminal fragments and the generation of amyloid beta-protein in endoplasmic reticulum and Golgi. *Biochemistry* 1998; 37 (47): 16465-71
30. Palacino JJ, Berechid BE, Alexander P, et al. Regulation of amyloid precursor protein processing by presenilin 1 (PS1) and PS2 in PS1 knockout cells. *J Biol Chem* 2000; 275 (1): 215-22
31. Guo S, Kemphues KJ. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 1995; 81 (4): 611-20
32. Subramaniam K, Seydoux G. *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* 1999; 126: 4861-71
33. Kraemer B, Crittenden S, Gallegos M, et al. NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. *Curr Biol* 1999; 9: 1009-18
34. Montgomery MK, Xu SQ, Fire A. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc Nat Acad Sci U S A* 1998; 95: 15502-7
35. Zamore PD, Tuschl T, Sharp PA, et al. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 2000; 101 (1): 25-33
36. Tabara H, Sarkissian M, Kelly WG, et al. The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 1999; 99: 123-32
37. Ketting RF, Haverkamp THA, van Luenen HGAM, et al. *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 1999; 99: 133-41
38. Smardon A, Spoerke JM, Stacey SC, et al. EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr Biol* 2000; 10: 169-78
39. Maine EM. A conserved mechanism for post-transcriptional gene silencing? *Genome Biol* 2000; 1 (3): 1-4
40. Zhang B, Gallegos M, Puoti A, et al. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 1997; 390: 477-84
41. Timmons L, Fire A. Specific interference by injected dsRNA. *Nature* 1998; 395: 854
42. Tavernarakis N, Wang SL, Dorovkov M, et al. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nature Genet* 2000; 24 (2): 180-3
43. Gartner A, Milstein S, Ahmed S, et al. A conserved checkpoint pathway mediates DNA damage-induced apoptosis and cell cycle arrest in *C. elegans*. *Mol Cell* 2000; 5: 435-43
44. Choy RKM, Thomas JH. Fluoxetine-resistant mutants in *C. elegans* define a novel family of transmembrane proteins. *Mol Cell* 1999; 4: 143-52
45. Dent JA, Smith MM, Vassilatis DK, et al. The genetics of ivermectin resistance in *Caenorhabditis elegans*. *Proc nat Acad Sci U S A* 2000; 97: 2674-9
46. Sym M, Basson M, Johnson C. A model for Niemann-Pick type C disease in the nematode *Caenorhabditis elegans*. *Curr Biol* 2000; 10: 527-30
47. Zwaal RR, Broeks A, van Meurs J, et al. Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc Nat Acad Sci U S A* 1993; 90: 7431-5
48. Liu LX, Spoerke JM, Mulligan EL, et al. High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res* 1999; 9: 859-67
49. Reinke V, Smith HE, Nance J, et al. A global profile of germline expression in *C. elegans*. *Mol Cell* 2000; 6: 605-16
50. Walhout AJM, Boulton SJ, Vidal M. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. *Yeast* 2000; 17: 88-94

Correspondence and offprints: Dr Nigel J. O'Neil, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, England.
E-mail: njo@sanger.ac.uk