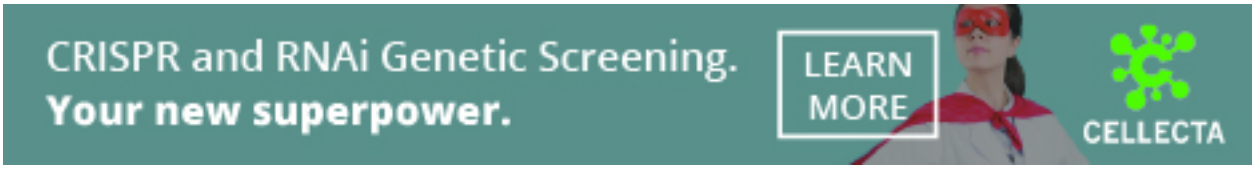
Postsequence genetics of Caenorhabditis elegans.

R H Plasterk

Genome Res. 1996 6: 169-175
Access the most recent version at doi:10.1101/gr.6.3.169



REVIEW

Postsequence Genetics of *Caenorhabditis elegans*

Ronald H.A. Plasterk

Netherlands Cancer Institute, Division of Molecular Biology, 1066CX Amsterdam, The Netherlands

If world oil prices dropped to zero next year, how would it change the world economy? Investments in oil field exploration would lose their value overnight, whereas shares in a factory making environmentally friendly combustion engines might go up. Everybody would feel the need to plan ahead, and many plans would change.

In genetics and molecular biology, DNA sequences are the fuel of research, and their prices are falling dramatically. Within 5 years many complete genomes will be sequenced, and sequence data will be like tap water in Amsterdam—essential for life, but too cheap to measure. A project that was perfectly rational 2 years ago will be a total waste of time tomorrow, and projects that seemed impossible will become feasible.

The aim of this review is to explore the consequences for biology of the wealth of DNA sequence data now becoming available. Several bacterial genomes have been sequenced already (Fleischmann et al. 1995; Fraser et al. 1995). The first animal to feel these changes will be the nematode *Caenorhabditis elegans*, and "the worm" will be the focus of this review. The virtues of *C. elegans* as a model system in biology have recently been sung elsewhere (Hodgkin et al. 1995). In brief, it does everything that makes life interesting (eating, copulating, getting around, and relating to the environment) and manages to do so with only 959 cells, of which 302 form the brain. However, it is likely that much of what is said will apply equally to other species; thus, I hope that the review may also be of some interest outside of the *C. elegans* community.

Falling Shares: Homolog Hunting

In one of his books, Francis Crick tells the amusing story of his visit to an institute where someone was building a model for DNA after the structure for the double helix had already been published. The man was so absorbed in his ongoing work that he refused to give up. I still get phone calls from people who see the possible homolog of their favorite gene in a genomic Southern blot of *C. elegans* and want to clone and sequence the gene. Currently about one-quarter of the genome is sequenced, as are thousands of cDNA clones. Because the genes are not evenly distributed, the sequence consortium has started with gene-rich areas, and 44% of the total 14,000 genes have been sequenced at present (February 1996) and are publicly available (http://www.sanger.ac.uk/~sjj/C.elegans_Home.html/ and http://genome.wustl.edu/gsc/gschmpg.html/). The expectation is that at the current speed of the project, the entire *C. elegans* genome sequence will be available in 1998 (Berks et al. 1995), and for any gene the expected moment of its sequence release lies, on average, about a year from now. The gene sequence will come with endless amounts of flanking DNA and with a precise localization on the genome map, so that no genome walking or gene mapping will be required. Of course, it is possible to clone and sequence a gene quicker than that, perhaps in half a year, but that requires quite a bit of work, and waiting is free. Also, most genes come in families, often spread over the genome, which brings the expected date for the release of one sequenced family member closer. Finally, the genome sequencers are specialized in what they do, and in sequencing accuracy they do a better job than most molecular biology laboratories.

There may be exceptions, but, in general, homolog hunting seems to be a waste of energy. The most efficient way to get a homolog is to wait and perform regular searches in silico (as opposed to vivo or vitro). A recent example of this approach is given in Troemel et al. (1995), where an extensive description of G-protein coupled receptor genes is based almost entirely on a careful analysis of the available genome sequence data. This paper describes what would have been many

E-MAIL rplas@ron.nki.nl; FAX + 31 20 6691383.

6:169–175 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

years of work in any other system—the identification and cloning of >50 membrane-spanning receptor genes, plus determination of the expression patterns of some of them. This work can now be done by analysis of data as they are found in the public data bases. The obvious next step in such studies is to investigate implied functions by targeted gene disruption (see below).

Gene Walking, Gene Sequencing

What is said above is also true for other approaches aimed at isolation and sequencing of genes. The main raison d'être for worms and flies as model organisms is that they can be used for forward genetics: the collection of mutants with interesting phenotypes followed by the identification of the affected gene. The standard approach is to map a gene by genetic crosses onto the genetic map, and then to enter the physical map and perform a genome walk toward the gene and sequence it. In rare cases, one may need to run ahead of the genome sequence, but chances are that the gain is little other than the possibility of comparing one's own sequence with that derived by professional sequencers and re-evaluating the differences (which, in a recent case in my laboratory, resulted in four corrections in a 5-bp sequence determined here and none in the canonical sequence).

Two cautionary notes with respect to the accuracy of the genome sequence: First, although the genome sequence itself is virtually impeccable (>99.9% accuracy), the gene predictions and annotations that appear in the GenBank and ACeDB (a *C. elegans* data base) are not. The program GENEFINDER provides predicted gene structures and is not always correct in detail. Also a gene can show (unpredictable) alternative splices. Therefore, if knowledge about the gene structure/protein product is important for a particular line of investigation, the researcher should confirm the structure, which is done most simply by the sequencing of cDNA products derived from RT–PCR and 5′ RACE products. Second, putative protein functions listed in GenBank and ACEDB have been assigned on the basis of similarity searches performed at the time of release to the public domain, and these assignments must be treated with caution.

Reverse Genetics

Reverse genetics—the isolation and analysis of mutants by targeted alteration of the known sequence of the gene—is a growth market. (Note that until recently this term was used in another sense in the area of human disease gene research: as a synonym of what is now called "positional cloning," the cloning of a gene of which a mutant version is found in patient DNA, using information about the position of the gene on the genetic and physical genome map. Within the terminology as it is used here, positional cloning is forward genetics, because it starts with a disease, a phenotype, and ends with the sequence of the affected gene.)

Gene disruption by homologous recombination with transgenic DNA has thus far not worked as a routine method in *C. elegans*, but target-selected inactivation by transposon insertion, as first done for *Drosophila* (Ballinger and Benzer 1989; Kaiser and Goodwin 1990), is a practical alternative (Plasterk 1992; Rushforth et al. 1993), especially when a transposon insertion mutant is used to isolate a derivative in which the transposon plus gene are deleted (Zwaal et al. 1993). The method is facilitated by the establishment of a frozen transposon insertion mutant library (Zwaal et al. 1993). These studies, when complemented by analysis of the pattern of gene expression, can give a good indication of the function of a gene. The phenotype of a mutant derived by reverse genetics can also be taken as a starting point for a forward mutant hunt for suppressors or enhancers (for review, see Huang and Sternberg 1995).

It happens frequently that a gene knockout in the mouse results in no obvious phenotype, and the most common explanation is that presumably there is redundancy. The animal is thought to contain several versions of the gene, most likely performing slightly different but overlapping functions; thus, inactivation of one gene has little effect, but inactivation of more would presumably have drastic consequences. This remains a hand-waving explanation as long as the genome sequence is not available. Soon in *C. elegans* biology these hypotheses will loose their gratuitous nature, because one will know exactly how many members of a gene family the organism has up its sleeve and can then serially kill them all. Of course the genome sequence will only reveal potential redundancy due to gene family members. It will not help for redundancy due to functionally related proteins that are not sequence homologs.

Because the various genome projects are gen-

erating many thousands of genes of known sequence, reverse genetics is becoming more and more important. One frequently sees researchers enter the *C. elegans* research field because one or more homologs of their favorite genes are found in the nematode genome.

Some caution is justified [also see "Notes for nematode novices" (Plasterk 1995)]. It should be kept in mind that isolation of a worm that has a gene of interest mutated is the beginning and not the end of an analysis of the gene's function. Possible outcomes of a gene knockout are (1) an obvious and expected phenotype [e.g., inactivation of a gene involved in cancer–drug resistance makes an animal sensitive to drugs (Broeks et al. 1995)]; (2) no obvious phenotype, which means that one has to start analyzing the less obvious ones, ranging from aperiodic defecation (Thomas 1990) to altered thermopreference (Mori and Oshima 1995); (3) unexpected phenotypes; and (4) death, and, as someone once said, "death is a boring phenotype." This is not necessarily true, however. By careful analysis of terminal phenotypes, and by (partial) rescue with transgenic DNA, one may extract much information from analysis of lethal mutations, but it is certainly not easy (e.g., cf. Greenstein et al. 1994). In any case, a detailed biological study will be required, based on some experience with the organism.

Does the genome sequence provide the starting point for a global approach toward reverse genetics, that is, should one aim to knock out all 14,000 genes and classify the phenotypes? With current technologies, this is far too much work, but possibly the availability of the genome sequence will trigger new technologies.

Forward Genetics

Even though the first thoughts that come to mind in relation to complete genome sequences concern reverse genetics, there is no reason why the availability of the complete genome sequence should prompt more interest in reverse and less in forward genetics. Both approaches aim to connect genotype to phenotype, by providing the combination of a known gene sequence and a known phenotype. The big advantage of forward over reverse genetics is that it provides an immediate focus on a phenomenon. There are several ways to take full advantage of the complete (or almost complete) genome sequence in forward genetics:

1. *Quick mapping of a mutated gene onto the genome map.* Previously, gene mapping involved measurement of crossover frequencies between the mutation and visible gene markers (Brenner 1974). An often better alternative is to use sequence-tagged site (STS) markers. The genome of some *C. elegans* strains contains many copies (>500) of the transposon Tc*1*, and these can be used as STSs and visualized by PCR of single animals or even embryos or eggs (Williams et al. 1992; Williams 1995). In essence, the mapping is done as it was in the past, but there are several advantages in using these genetic markers: (*A*) They have by themselves no phenotypes, so that in principle their presence does not interfere with the phenotype of the genes under study. This in contrast to many mutants in body morphology or locomotion. (*B*) A related advantage is that many alleles can be studied independently in the same animal. (*C*) These markers provide a direct link between the genetic map, which is a projection of recombination distances, and the physical map, which is the genome sequence. (*D*) The approach may allow mapping of multiple factors that contribute to quantitative traits (Lander and Botstein 1989).

Our laboratory has recently expanded the limited set of 38 STS markers to over 500 (H.C. Korswagen, R.M. Durbin, M.T. Smits, and R.H.A. Plasterk, in prep), which should allow a coverage of on average one STS per 200 kbp. This is a denser map than that of those classic genetic markers that are commonly found useful for gene mapping (excluding lethals, etc.).

The described method does not strictly depend on the genome sequence. All it requires is a high-quality physical map onto which polymorphic STSs can be placed, but both the refinement of the physical map and the mapping of new STSs go hand in hand with the genome sequence project.

2. *Quick mutant gene identification.* Once a gene has been mapped to a region of the genome, the next step, gene identification, is made easier by the "candidate gene approach." This approach, which is already used in human genetics (at great labor cost), was not previously worth the investment in *C. elegans*. However, once the sequence information is available, it will be difficult to resist the temptation to take a quick look at what lies in the region and to make a guess as to which genes might, when mutated, result in the phenotype at hand. In a sense, the sequence might devalue brute force and appreciate intu-

ition and educated guesses in gene identification. For example, researchers have identified a number of suppressors of an activated form of RAS in *C. elegans* (Kornfeld et al. 1995). The suppressor mutations can be mapped meiotically to a 1 map unit interval (1 Mb). Putative proteins encoded in this region, which are related to players in the RAS signaling pathway (e.g., a farnesyl transferase), provide obvious candidates. Whether or not a given candidate gene corresponds to the suppressor locus can be rapidly determined by direct sequencing of genomic DNA from the mutant. In this case, efficient identification of suppressor loci can allow the researcher to focus on novel genes that might be difficult to identify by biochemical methods.

3. *Tagging of the mutant gene.* The most widely used mutagen in *C. elegans* is ethylmethane sulfonate (EMS), which usually results in point mutations and has the most favorable ratio between mutagenicity and other toxicity (Anderson 1995). However, in the age of postsequence genetics, in some cases it may be wise to choose a mutagen that leaves a molecular tag, such as a transposon (Moerman et al. 1986). It has always been an advantage to have an allele that is defined by a transposon, because the transposon can help recognize and clone the affected gene. But now the advantage becomes spectacular. Consider the following protocol: A strain permissive for Tc*1* transposition is used to select mutants of a desired phenotype. The phenotype may also consist of suppression of the phenotype of another mutation. Next, the mutant is crossed a few times with a strain of known low Tc*1* copy number, nonpermissive for transposition (the Bristol strain, commonly used in nematode genetics), and a few independent mutant segregants are picked. It is now likely that the only extra Tc*1* insertion present in all mutants is the one responsible for the mutation.

Using the entire battery of PCR tricks available to date, the flanking region of this specific Tc*1* element can be obtained and sequenced in a few days. The simplest approach is to amplify the flanking regions of all of the ~30 transposons in these strains in a shotgun fashion (using a Tc*1*-specific primer and an anchor primer), clone them into M13, and run two or three sequencing gels: The flanking sequences of the transposons present in all Bristol N2 strains are ignored, and the one present only in the mutant is analyzed further. In a slightly more sophisticated scheme, one could filter out the common flanking region before sequencing. From the sequence of the Tc*1* flanking region, one can go directly to the genome sequence and see if a likely candidate gene is recognized. Note that the first step in classic mutant hunts, crosses between mutants to collect different alleles of the same gene into one complementation group, is not necessary, because the genome sequence will tell immediately whether independent mutants are alleles of the same gene.

Gene Expression

Another area of research that currently involves laborious experiments is the analysis of the expression patterns of genes. Current approaches involve analysis of promoter specificity by fusion to reporter genes (Fire et al. 1990; Young and Hope 1993; Chalfie et al. 1994; for review, see Krause 1995; Mello and Fire 1995), detection of RNA by in situ hybridization (Seydoux and Fire 1995; Y. Kohara, unpubl.), or, the final level of analysis, in situ immunochemistry using antibodies against a protein of interest (Miller and Shakes 1995). One may envisage more global approaches, for example, as proposed recently by Schena et al. (1995) and Velculescu et al. (1995), where many different sequenced cDNA clones are tested in parallel for their expression in specific tissues. Here a strength of the nematode—morphology at single-cell resolution—may also be a weakness, because it will not be easy to obtain RNA and make cDNA of each of the 959 somatic cells of the adult animal. It would be a worthy goal, as it would allow the identification of cell-specific expression of all genes in the animal's genome.

At a somewhat cruder level this analysis is certainly feasible; one could screen all sequenced genes in parallel to compare expression levels between different developmental stages, between the two sexes, and so forth.

Proteins

Because *C. elegans* has been studied primarily for genetic rather than biochemical analysis, there has not been a strong emphasis on purification and partial sequencing of nematode proteins. There may now be an extra reason to consider this approach, as a short stretch of an amino acid sequence of a protein will immediately identify the gene and, given the reasonable accuracy of

intron/exon predictions, the likely composition of the entire protein. For example, one can aim to prepare proteins from organelle fractions, from specific developmental stages (eggs, larvae, males, or females), from specific tissues (e.g., sperm), or proteins that are conditionally expressed (e.g., upon irradiation), separate them by two-dimensional electrophoresis, determine a short sequence tract of each protein, and know the sequence. The functional role of the protein can then be investigated, for example, by targeted gene inactivation (see above).

The Power of Tags

The immense power of sequence tags deserves explicit attention. For example, differential display of mRNA or cDNA becomes child's play, as a single read will identify the gene of interest. Analyzing candidates from various types of screens (e.g., the yeast two-hybrid screen) is also greatly simplified, as a single read from each end of a DNA clone will completely identify every clone. This approach is also implicit in the approaches described above for transposon tagging, amino-terminal protein sequencing, and so forth.

Other Areas

This review will not concern itself with the newly emerging science of computerized genome analysis, but it is clear that the storage, analysis, and interpretation of genome data will become vitally important. A useful analysis requires feedback of biological analysis into the computer, and that in turn requires that the software be sufficiently useful for the average biological worker. The ACeDB data base plus software (Eeckman and Durbin 1995) seems to answer these requirements, and is constantly being improved. It not only offers an information resource but is also a strong incentive for collaboration, because it provides a location for everyone working on a given gene to find out which other laboratories have contributed clones, mutants, annotations, and so forth.

The analysis of genome organization will become important. The identification and study of operons in *C. elegans* has been facilitated greatly by the genome sequence (Spieth et al. 1993). The genome sequences of other organisms may also indicate multiple genes expressed from the same promoter. The study of repetitive sequences will start with the sequence, not with the experiments that were done previously to identify them. Many new repeat families will be identified in silico, and their distribution, organization, and variation will be determined.

Chromosome-specific phenomena can be examined, for example. (1) The X chromosome must be "counted" in a way that is distinct from the autosomes as part of the primary sex determination mechanism in *C. elegans*—the X/A ratio. Additionally, the X-chromosome dosage compensation machinery utilizes proteins (i.e., Dpy-27) that bind specifically to the X chromosome. What features of the X-chromosome sequence are distinct from the autosomes? (2) Questions about the organization of genes/sequences can be addressed on a whole-chromosome basis. Is there a system for the strand specificity of genes? Do genes of common expression pattern, timing, or function cluster in any way?

Other areas of research that can profit from the *C. elegans* genome sequence but will not be discussed in detail here include population biology. For example, how are sequence polymorphisms geographically distributed? Can short polymorphic sequence tracts be used to keep track of the gene flow within the species worldwide? Evolutionary biology studies will also benefit. In this area, comparison to the related species *Caenorhabditis briggsae* can help identify conserved and, therefore, presumably, important sequences (Prasad et al 1991; Kuwabara and Shah 1994), and comparison to, on the one hand, simpler genomes such as that of yeast, and, on the other hand, more complex genomes can hint at which genes are minimally required for an organism to be a eukaryote, a metazoan, or a vertebrate.

It has taken the biological community several years to appreciate fully the powers of new technologies such as gene cloning and PCR. The same phenomenon will probably occur after the genomes are sequenced. Within 5–10 years, we will all have free access to the complete encoded information that makes a human, worm, and plant what they are. The entire genotype will be known, as will the phenotype in terms of morphology and physiology. The remaining, but daunting, task is to explain the one in terms of the other.

ACKNOWLEDGMENTS

I thank the people in my laboratory for comments on the manuscript. I am also particularly grateful to two reviewers

for very good suggestions and additions. Work in my laboratory is supported by grants from the Netherlands Organization for Scientific research (NWO), the National Institutes of Health–NCRR (RR10082-02), and the Human Frontier Science Project Organization (RG-309/92).

REFERENCES

Anderson, P. 1995. Mutagenesis. *Caenorhabditis elegans*: Modern biological analysis of an organism (ed. H.F. Epstein and D.C. Shakes). *Methods Cell Biol.* **48:** 584–607.

Ballinger, D.G. and S. Benzer. 1989. Targeted gene inactivation in *Drosophila*. *Proc. Natl. Acad. Sci.* **86:** 9402–9406.

Berks, M. and the *C. elegans* Genome Mapping and Sequencing Consortium. 1995. The *C. elegans* genome sequencing project. *Genome Res.* **5:** 99–104.

Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77:** 71–94.

Broeks, A., H.W.R.M. Janssen, J. Calafat, and R.H.A. Plasterk. 1995. A P-glycoprotein protects *Caenorhabditis elegans* against natural toxins. *EMBO J.* **14:** 1858–1866.

Chalfie, M., Y. Tu, G. Euskirchen, W. Ward, and D. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science* **263:** 802–805.

Eeckman, F.H. and R. Durbin. 1995. ACeDB and Macace. *Caenorhabditis elegans*: Modern biological analysis of an organism (ed. H.F. Epstein and D.C. Shakes). *Methods Cell Biol.* **48:** 584–607.

Fire, A., S.W. Harrison, and D. Dixon. 1990. A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93:** 189–198.

Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–498.

Fraser, C.M., J.D. Gocayne, O. White, M.D. Adams. R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kerlavage, G. Sutton, J.M. Kelley, et al. 1995. The minimal gene complement of Mycoplasma genitalium. *Science* **270:** 397–403.

Greenstein D., S. Hird, R.H.A. Plasterk, Y. Andachi, Y. Kohara, B. Wang, M. Finney, and G. Ruvkun. 1994. Targeted mutations in the *Caenorhabditis elegans* POU homeo box gene *ceh-18* cause defects in oocyte cell cycle arrest, gonad migration, and epidermal differentiation. *Genes & Dev.* **8:** 1935–1948.

Hodgkin, J., R.H.A. Plasterk, and R.H. Waterston. 1995. The nematode *Caenorhabditis elegans* and its genome. *Science* **270:** 410–414.

Huang, L.S. and P.W. Sternberg. 1995. Genetic dissection of developmental pathways. *Caenorhabditis elegans*: Modern biological analysis of an organism (ed. H.F. Epstein and D.C. Shakes). *Methods Cell Biol.* **48:** 98–122.

Kaiser, K. and S.F. Goodwin. 1990. "Site selected" transposon mutagenesis of *Drosophila*. *Proc. Natl. Acad. Sci.* **87:** 1686–1690.

Kornfeld, K., D.B. Hom, and H.R. Horvitz. 1995. The ksr-1 gene encodes a novel protein kinase involved in Ras-mediated signaling in *C. elegans*. *Cell* **83:** 903–913.

Krause M. 1995. Techniques for analyzing transcription and translation. *Caenorhabditis elegans*: Modern biological analysis of an organism (ed. H.F. Epstein and D.C. Shakes). *Methods Cell Biol.* **48:** 513–533.

Kuwabara, P.E. and S. Shah. 1994. Cloning by synteny: identifying *C. briggsae* homologs of *C. elegans* genes. *Nucleic Acids Res.* **22:** 4414–4418.

Lander, E.S. and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121:** 185–199.

Mello, C. and A. Fire. 1995. DNA transformation. *Caenorhabditis elegans*: Modern biological analysis of an organism. (ed. H.F. Epstein and D.C. Shakes) *Methods Cell Biol.* **48:** 452–482.

Miller, D.M. and D.C. Shakes. 1995. Immunofluorescence microscopy. *Caenorhabditis elegans*: Modern biological analysis of an organism (ed. H.F. Epstein and D.C. Shakes). *Methods Cell Biol.* **48:** 365–394.

Moerman, D.G., G.M. Benian, and R.H. Waterston. 1986. Molecular cloning of the muscle gene unc-22 in *Caenorhabditis elegans* by Tc1 transposon tagging. *Proc. Natl. Acad. Sci.* **83:** 2579–2583.

Mori, I. and Y. Ohshima. 1995. Neuronal regulation of thermotaxis in *Caenorhabditis elegans*. *Nature* **376:** 344–348.

Plasterk, R.H.A. 1992. Reverse genetics of *C. elegans*. *BioEssays* **14:** 629–633.

Plasterk, R.H.A. 1995. Reverse genetics: From gene sequence to mutant worm. *Caenorhabditis elegans*: Modern biological analysis of an organism. *Methods Cell Biol.* **48:** 59–80.

Prasad, S.S., L.J. Harris, D.L. Baillie, and A.M. Rose. 1991. Evolutionarily conserved regions in *Caenorhabditis elegans* transposable elements. *Genome* **34:** 6–12.

Rushforth, A.M., B. Saari, and P. Anderson. 1993. Site-selected insertion of the transposon Tc1 into a *Caenorhabditis elegans* myosin light chain gene. *Mol. Cell. Biol.* **13:** 902–910.

Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression

patterns with a complementary DNA microarray. *Science* **270:** 467–470.

Seydoux, G. and A. Fire. 1995. Whole-mount in situ hybridization for the detection of RNA in Caenorhabditis elegans embryos. *Caenorhabditis elegans*: Modern biological analysis of an organism. (ed. H.F. Epstein and D.C. Shakes) *Methods Cell Biol.* **48:** 323–339.

Spieth, J., G. Brooke, S. Kuersten, K. Lea, and T. Blumenthal. 1993. Operons in C. elegans—Polycistronic messenger RNA precursors are processed by transsplicing of SL2 to downstream coding regions. *Cell* **73:** 521–532.

Thomas, J.H. 1990. Genetic analysis of defecation in *C. elegans*. *Genetics* **124:** 855–872.

Troemel, E.R., J.H. Chou, N.D. Dwyer, H.A. Colbert, and C.I. Bargmann. 1995. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell* **83:** 207–218.

Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler. 1995. Serial analysis of gene expression. *Science* **270:** 484–486.

Williams, B.D. 1995. Genetic mapping with polymorphic sequence-tagged sites. *Caenorhabditis elegans*: Modern biological analysis of an organism. (Ed. H.F. Epstein and D.C. Shakes) *Methods Cell Biol.* **48:** 81–96.

Williams, B.D., B. Schrank, C. Huynh, R. Shownkeen, and R.H. Waterston. 1992. A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* **131:** 609–624.

Young, J.M and I.A. Hope. 1993. Molecular markers of differentiation in *Caenorhabditis elegans* obtained by promoter trapping. *Development* **196:** 124–132.

Zwaal, R.R., A. Broeks, J. Van Meurs, J.T.M. Groenen, and R.H.A. Plasterk. 1983. Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl. Acad. Sci.* **90:** 7431–7435.