

Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi

Joris Pothof,¹ Gijs van Haaften,¹ Karen Thijssen,¹ Ravi S. Kamath,² Andrew G. Fraser,² Julie Ahringer,² Ronald H.A. Plasterk,^{1,3} and Marcel Tijsterman^{1,4}

¹Hubrecht Laboratory, Center for Biomedical Genetics, 3584 CT, Utrecht, The Netherlands; ²University of Cambridge, Wellcome Trust/Cancer Research UK Institute and Department of Genetics, Cambridge CB2 1QR, UK

An RNA interference (RNAi)-based genome-wide screen was performed to detect genes that contribute to genome stability in somatic cells of *Caenorhabditis elegans*. We identified 61 such genes; these also affect spontaneous mutagenesis in the germ line. Their sequence suggests a role in DNA repair and/or replication, in chromatin remodeling, or in cell cycle control; there are also many novel genes that are highly conserved from yeast to human. Because known mutator genes are causally involved in many hereditary and sporadic human cancers, it is likely that some of these new mutators are equally relevant in cancer etiology.

Received November 20, 2002; revised version accepted December 24, 2002.

The first genetic analyses of mechanisms that protect the genome against spontaneous mutations were performed in bacteria (Friedberg et al. 1995), leading to the identification of DNA mismatch repair genes and the elucidation of the mechanism by which the repair machinery recognizes a DNA replication error and repairs the damage (for review, see Kolodner 1996; Modrich and Lahue 1996). This field of research gained clinical relevance when a hereditary form of cancer, hereditary non-polyposis colon cancer (HNPCC), was found to be the result of germ line mutations in human homologs of DNA mismatch repair genes (e.g., Fishel et al. 1993; Leach et al. 1993). Also, many if not all sporadic human cancer cells display a mutator phenotype, meaning that they have enhanced levels of mutagenesis of some type: point mutations, frame shifts, gross chromosomal rearrangements, or aneuploidy. This has led to the hypothesis that the conversion of a wild-type cell into a mutator cell is an early event in carcinogenesis and that a mutator phenotype is causal to the development of cancer; multiple loss-of-function mutations in a limited number

of genes are required for a regular cell to become a cancer cell (Fearon and Vogelstein 1990; Loeb 1991, 2001). Although the chances of several mutations coinciding in one cell are very low (the product of the chances for each mutation), a mutator's enhancement of the general level of mutagenesis greatly enhances the chance of the combination of mutations to arise.

Here we used RNA interference (RNAi) to systematically screen the nematode *Caenorhabditis elegans* genome for genes that, when inactivated, increase the level of spontaneous mutagenesis. We identified 61 genes with presumed roles in DNA repair, replication, chromatin organization, and cell cycle control. These mutator genes protect the *C. elegans* genome against insertion and deletion mutations in both the soma and the germline, and are thus potential tumor suppressors in higher species.

Results and Discussion

Genome-wide RNAi screens became feasible recently with the generation of a library of bacterial strains that each produce double-stranded RNA (dsRNA) for an individual nematode gene. The current library contains 16,757 bacterial strains that will target ~86% of the 19,427 current predicted genes of the *C. elegans* genome, and the loss-of-function phenotype when performing systemic RNAi on a genome-wide scale is estimated to be ~65% (Fraser et al. 2000; Kamath et al. 2003). Animals fed these bacteria induce RNAi in almost all tissues of the nematode (Timmons and Fire 1998; Kamath et al. 2001).

As a readout for increased DNA instability, we used a recently developed reporter assay that allows the detection of frameshifts and small insertions or deletions in somatic cells of *C. elegans* (Tijsterman et al. 2002). In brief, a *gfp-LacZ* reporter construct is put out of frame by an A₁₇ mononucleotide DNA repeat cloned directly between the initiation ATG and the *gfp-LacZ* open reading frame (ORF; Fig. 1). The presence of such a DNA repeat will enhance the chances of a frameshift mutation to occur and therefore increase the sensitivity of the reporter system.

The transgene is integrated in the *C. elegans* genome as a multicopy array. Only constructs that acquire a mutation that brings the reporter *LacZ* gene back in frame will be visualized by a β -galactosidase staining. The advantage of *LacZ* over other colorimetric indicators (such as *gfp*) is that prolonged staining can visualize relatively low levels of expression. Therefore we can detect a frameshift in only one of the transgenic copies of the integrated high copy array. This shows up as a blue cell, or—if the frameshift occurred early during development—a patch of blue cells. Previously, we found that frequent blue patches, as a result of restoration of *LacZ* expression, are visualized in mismatch repair-deficient animals (*msh-6* mutants). These patches were not observed in repair-proficient animals (Tijsterman et al. 2002).

Using this assay we tested 16,606 genes by RNAi and found 61 genes whose knock-down phenotype results in somatic blue patches similar to those seen in *msh-6*-deficient animals (Table 1), a pattern indicative of a mutator phenotype. Among the identified genes are the four

[Keywords: Carcinogenesis; mutator genes; DNA repair; spontaneous DNA damage; *C. elegans*]

Corresponding authors.

³E-MAIL Plasterk@niob.knaw.nl; FAX 31-30-251-6554.

⁴E-MAIL Tijsterman@niob.knaw.nl; FAX 31-30-251-6554.

Article published online ahead of print. Article and publication date are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1060703>.

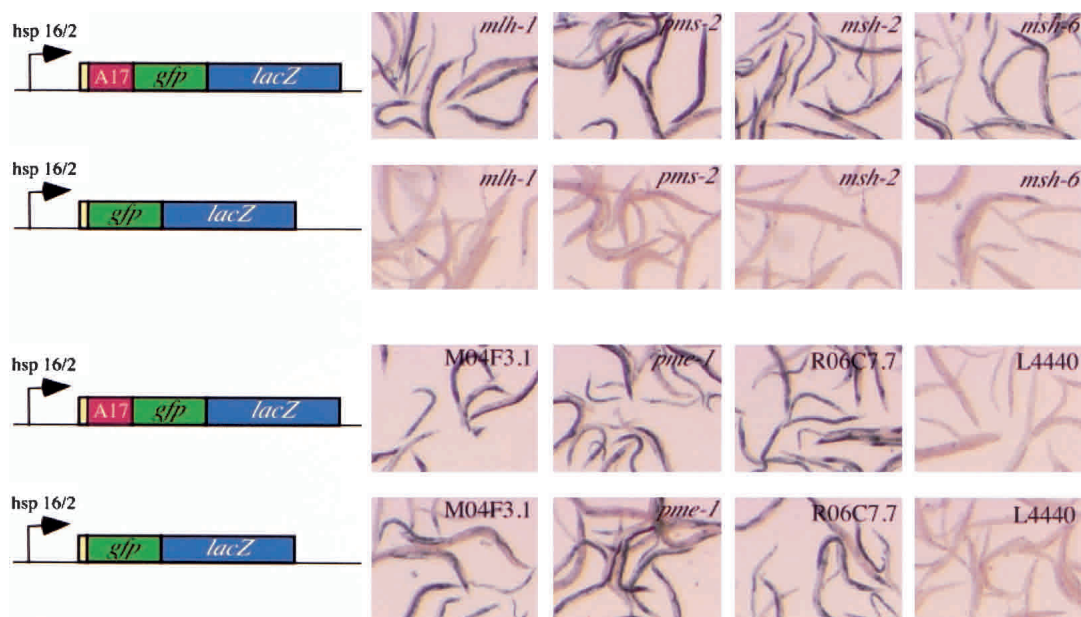


Figure 1. RNAi-mediated DNA instability in *C. elegans*. Strains that contain heat shock promoter-driven out-of-frame *gfp-lacZ* reporter gene fusions were assayed on dsRNA homologous to *mlh-1*, *pms-2*, *msh-2*, *msh-6*, M04F3.1, Y71F9AL.18 (*pme-1*), R06C7.7, and the empty vector control L4440. Two different out-of-frame constructs were tested, one containing a mononucleotide (A₁₇) repeat placed immediately downstream of the ORF's ATG (this reporter has been used to screen the RNAi library), the other identical except that it lacks the mononucleotide repeat.

orthologs of human DNA mismatch repair genes, *msh-2*, *msh-6*, *mlh-1*, and *pms-2* (Fig. 1; note that these were recognized in blinded procedure, without the investigator knowing the gene identity). This four-out-of-four hit rate suggests a good coverage in this screen. The sequence of the identified genes was subsequently inspected for domain structures, and we searched the sequenced genomes of *Homo sapiens*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe* for homologous genes. We termed gene-pairs orthologous if the reverse BLASTP identifies the original *C. elegans* sequence. Based on their sequence homology, we tentatively classified the identified putative mutator genes in five categories: (I) DNA repair and replication, (II) chromatin remodeling, (III) cell cycle, (IV) novel genes with orthologs in many species, and (V) genes that have no clear orthologs in other species.

Genes in the first class are involved in DNA repair and replication. Besides the mismatch repair genes, this category also includes replication protein A2 (M04F3.1), which is known to act in multiple DNA metabolic processes including DNA replication, nucleotide excision repair, and double-strand-break repair (e.g., New et al. 1998; Araujo et al. 2000; Walter and Newport 2000), *pme-1* (Y71F9AL.18), the *C. elegans* Parp [poly(ADP-ribose) polymerase], a polymerase involved in DNA damage response (Burkle 2001), and the *C. elegans* homolog of REV1, which has been shown to function in mutagenic translesion synthesis (Wang 2001).

The second class of genes is involved in chromatin organization and remodeling. Although at present not much is known about the role of chromatin condensation in protecting DNA against damage, a role for such genes in genome stabilization has been described (Celeste et al. 2002). Note also from this list that three proteins that are known to function in a complex are

independently recognized: SIN3, HDAC1, and HDAC2 (Zhang et al. 1997); all show the mutator phenotype. This is also seen in the third class of genes, which has a putative role in cell cycle control. For example *cdc5* and *cdc14* work together in a mitotic checkpoint, and cell cycle arrest upon DNA damage is dependent on these genes (Stegmeier et al. 2002 and references therein).

In addition, there are 29 genes that have not been described previously (Table 1, categories IV and V). Genes in Table 1, category IV, have clear orthologs in other species and many of them contain recognizable sequence motifs, for example, an exonuclease, which could point to a role in DNA repair and/or DNA synthesis. Other domains such as the GCN5 (GNAT) histone acetyltransferase domain, JmjC domain, and SET domains suggest that they are involved in chromatin remodeling (e.g., Kuo et al. 2000; Lachner and Jenuwein 2002), indicating that the genes categorized in Table 1, category IV, are not a separate class, but based on putative function, could be involved in the same DNA metabolic processes as the genes listed in categories I to III.

Can we be sure that the blue patched phenotype originates from frameshift mutations in the DNA and does not occur as shifts at the transcriptional or translational level? One clue is that most worms have a different pattern of blue patches. If shifts occurred at the transcriptional or translational level, one would expect that all worms would have the same pattern, for example, completely blue worms or one specific tissue. The variable pattern indicates a stochastic event in development. Another way to establish that these mutator genes indeed act at the level of mutations in DNA is to investigate whether RNAi of these genes results in a germline mutator phenotype, as one would expect that mutator genes act in the germline as well as in the soma to protect

Table 1. Genes that protect the *C. elegans* genome against mutations

<i>C. elegans</i>	MSI	<i>H. Sapiens</i>	<i>S. cerevisiae</i>	<i>S. pombe</i>	ORF description	
I. DNA repair and replication						
<i>mlh-1</i>	T28A8.7	+	MLH1	Mih1	Mih-1	DNA mismatch repair protein
<i>pms-2</i>	H12C20.2a	+	PMS2	Pms1	Pms1	DNA mismatch repair protein
<i>msh-2</i>	H26D21.2	+	MSH2	Msh2	Msh2	DNA mismatch repair protein
<i>msh-6</i>	Y47G6A.11	+	MSH6	Msh6	Msh6	DNA mismatch repair protein
<i>pme-1</i>	Y71F9AL.18 *		ADPRT			Poly (ADP-ribose) polymerase
	Y56A3A.29		UNG1	Ung1	Ung1	Putative uracil-DNA glycosylase, involved in base excision repair
	M04F3.1 *†###		RPA2	Rfa2	Rfa2	Putative replication protein A 2; 32 kDa subunit
	ZK675.2 *	+	REV1	Rev1	SPBC215.16c	Putative deoxycytidyl transferase involved in mutagenic translesion synthesis
	T23G5.6		TdT binding			Putative terminal deoxycytidyl transferase binding protein
<i>rfc-1</i>	C54G10.2		RFC1	Rfc1	Rfc1	Putative replication factor C 140 kDa subunit
<i>rfc-3</i>	C39E9.13		RFC3	Rfc5	SPBC83.14c	Putative replication factor C 38 kDa subunit
	E03A3.2		RECQL5	Sgs1	Hus2	Member of the recQ family of helicases
II. Chromatin organization and remodeling						
	F45E1.6 ###		Histone H3.3	Histone H3	Histone H3.1	Member of the histone H3 protein family
<i>hil-1</i>	C30G7.1		Histone H1 like			Histone H1 like protein
<i>sir-2.2</i>	F46G10.7		SIR2TUII like	Hst2	SPCC132.02	Similarity to members of the SIR2 like protein family
<i>pqn-28</i>	F02E9.4		KIAA0700	Sin3	Sin3	Putative sin3; a transcriptional regulator in Nurd complex
<i>hda-1</i>	C53A5.3 ###		HDAC1	Rpd3	Clr6	Histone deacetylase
	R06C1.1		HDAC1	Rpd3	Clr6	Member of the histone deacetylase protein family
<i>hda-2</i>	C08B11.2		HDAC2	Rpd3	Clr6	Strong similarity to human histone deacetylases HDAC2 and HDAC3
	C34E10.5		JBP1	Hsl17	Skb1	Putative protein methyltransferase
	R06C7.7 *†		H-1-(3)mbt-like			Similarity to human H-1-(3)mbt-like and sex comb on midleg proteins
III. Cell cycle and checkpoint						
	B0511.8 ###	+	MPPE1	Cdc1	Cdc1	Involved in intrachromosomal recombination, and nuclear division
	B0495.2 *†		CDC2L1		SPBC18H10.15	Cyclin-dependent kinase CDC2; PITSLRE protein family
	D1081.8 ###		CDC5	Cef1	Cdc5	Putative cdc5; putative myb transforming protein
	C17G10.4 *†		CDC14A2	Cdc14	Cdc14	Putative cdc14
	F43D2.1 *†	+	Cyclin K	Ume3	Pch1	G1/S specific cyclin C like; strong homology to human cyclin K
<i>cul-1</i>	D2045.6 ###		CUL1	Cdc53	Peu1	Cell cycle control protein, member of the cullin family
<i>cul-5</i>	ZK856.1	+	VACM1	Cdc53	Peu1	Cell cycle control protein, member of the cullin family
<i>hpr-17</i>	F49E12.6		E2F5			Putative E2F transcription factor
	F32A11.2 †		RAD17	Rad24	Rad17	Ortholog of <i>S. pombe</i> and human RAD17; DNA damage checkpoint
<i>mrt-2</i>	K08F4.1		RUVBL	Chl12	Chl12	Putative Chl12; involved in chromosome transmission and telomere maintenance
	Y41C4A.14		RAD1	Rad17	Rad1	Ortholog of human RAD1; DNA damage checkpoint protein
IV. Novel genes with clear orthologs in other species						
	T26A5.5 †		PCCX2			Protein of unknown function; contains jmjC domain
	T05A12.4		AL451145	YLR247C	SPAC144.05	Protein of unknown function; contains SNF2, DEAD/DEAH box helicase, helicase C-terminal, RING finger domains
	C35A5.9		FLJ22237	Rpd3	Clr6	Protein of unknown function; contains HDAC domain
	C06H2.3		FLJ13798			Function unknown; contains GCN5 related N-acetyl transferase and jmjC domain
	F15E6.1		KIAA1757	YPL181W	SPAC22E12.11c	Protein of unknown function; contains SET and PHD finger domains
<i>dnj-15</i>	K08D10.2		AAH00004	YGL018C	SPAC144.08	Protein of unknown function; contains DnaJ domain
	ZK520.3		FLJ_3127			Protein of unknown function; contains WD40 repeats
	F29C4.6		LOC90353	YGL211W	SPBC2G5.03	Protein of unknown function; contains UPF0021 domain
	Y113G7A.9		HSPC015	YOR173W	SPBP4H10.20	Protein of unknown function
	K09B11.2	+	FLJ23323	YLL035W	SPCC830.03	Protein of unknown function
	Y76A2B.5	+	HSPC161			Protein of unknown function
	T28D6.4 †		KIAA1223			Protein of unknown function; contains ankyrin repeats
	B0412.3 *†		FLJ12716			Protein of unknown function
	C32D5.6		KIAA0452	YKR018C	SPAC11E3.14	Protein of unknown function; contains TPR repeat
	H14A12.3 ###	+	FLJ22386			Protein of unknown function
	R06F6.8a		KIAA1432			Protein of unknown function
	Y39A3CL.4a			YOR3263W	SPAC1851.04c	Protein of unknown function
	Y56A3A.33	+	KIAA1138	YLR107w	SPAC1782.08c	Protein of unknown function; contains EXO III domain
V. Genes with homology to other genes (no clear orthologs)						
<i>pme-6^b</i>	AC8.1		ADPRT			Putative poly (ADP-ribose) polymerase; Paralog of <i>C. elegans</i> Y71F9AL.18
	F19G12.2	+	RIR2	Rnr4	Rir2	Similarity to the ribonucleotide reductase small subunit
	R03D7.2		KIAA1404	Sen1	SPAC16C9.06c	Contains helicase domain. Possible paralog of <i>C. elegans</i> K08D10.5
	Y43F4B.1		FLJ23414	Set2	Clr4	Protein of unknown function; contains SET domain
	F02D10.7		HIF-1	Set1	SPCC306.04c	Protein of unknown function; contains SET and PHD finger domain
<i>cul-6</i>	K08E7.7 ###†		CUL1	Cdc53	Peu1	Member of the cullin protein family of cell cycle control proteins
	Y111B2A.1		CLK2	KSS1	C1D4.11c	Protein of unknown function; contains kinase domains
	F54D10.1 *†		SKP1	Skp1	Skp1	Homolog of SKP1; 1 of 21 <i>C. elegans</i> SKP related genes
<i>skr-15</i>	F57C12.2		FIP-1	Gtr1	SPBC337.13c	Protein with strong similarity to GTP binding protein
	W07A8.3 ¹⁾ or		KIAA0473	YDR320C	SPAC17A5.12	Involved in clathrin-mediated endocytosis; contains DnaJ domain /
<i>dnj-25</i>	W07A8.5					Member of uncharacterized gene family
	F14F9.5					Protein of unknown function; contains AP endonuclease domain

¹⁾ The RNAi clone that was used contains 102 bp. of sequence homologous to W07A8.5 (1550 bp. directed against W07A8.3).

²⁾ On the nucleotide level the genomic sequence of *pme-6* (Panel V) shows > 90% identity to *pme-1* (listed in panel I).

The genes were identified on the basis of a mutator phenotype, as measured by the reversion of *LacZ* expression, upon growth on dsRNA homologous to the cognate *C. elegans* gene. MSI (microsatellite instability) indicate the situation where the reversion to proper *LacZ* expression depended on the presence of the mononucleotide repeat. Based on protein domain structure and sequence homology to known proteins, the identified genes were placed into five categories (homologous proteins in *Homo sapiens*, *S. cerevisiae* and *S. pombe*, are depicted). Panels I (DNA repair and replication), II (chromatin organization and remodeling) and III (cell cycle and check point) contain genes which at some level have been described in other species. Panel IV: Novel genes with clear orthologs in mammalian species. Panel V: Genes that have some level of homology to genes in other species or to *C. elegans* family members, but at present lack clear orthologous genes. Other notable RNAi induced phenotypes, i.e., a reduced broodsize (*), embryonic lethality (#10–30%; ##30–60%, ###60–100%), and slow growth (†) are also indicated. The absence of such marking reflects that this phenotype has not been observed despite careful inspection.

progeny from harmful mutations. We performed two different assays to validate a possible germline mutator phenotype. (1) For three genes, we quantified the fre-

quency of induced recessive lethal mutations in a balanced region of the genome (Rosenbluth et al. 1983); RNAi against M04F3.1 (encoding replication protein

A2), *pme-1*, and R06C7.7 resulted in 14-, 9- and 8-fold elevated germline mutation frequencies above the wild-type control (data not shown), indicating that these genes protect the germline of *C. elegans* against mutations. (2) For 24 genes, we determined the mutation frequency using an endogenous monitor gene (Greenwald and Horvitz 1980; see also below). As shown in Figure 2, 23 genes have a significantly increased germline mutation frequency compared to the wild-type control, indicating that these genes (and thus probably 96% of the genes that were identified) also protect the germline against mutations. Of this subset, only C17G10.4(RNAi), the putative ortholog of *cdc14*, does not show an increased germline mutation frequency, suggesting that this gene may only act in the soma to protect the DNA against mutations.

With this experimental approach, we expected to identify genes whose loss-of-function phenotype is the frequent occurrence of frameshifting mutations in particular DNA substrates (note the presence of the mono A repeat in the reporter) and which thus display not only a mutator phenotype but also a so-called microsatellite instability (MSI) phenotype. To test this presumption, we tested for all clones whether the reversion to proper *LacZ* expression depended on the presence of the mononucleotide repeat or whether this reversion would also occur in an out-of-frame *lacZ* lacking a repeat. As shown in Figure 1, RNAi against the mismatch repair genes *mlh-1*, *pms-2*, *msh-2*, and *msh-6* does not result in a blue-patched phenotype in the absence of this repeat,

which is in good agreement with the MSI phenotype of tumor cells associated with a defect in DNA mismatch repair (MMR). In addition to these four, nine other RNAi clones (Table 1) induced the MSI phenotype, making these genes logical candidates to constitute the class of yet unidentified genes that are causally mutated in HNPCC or HNPCC-like patients.

Additionally, the screen resulted in 48 clones, for example, M04F3.1, *pme-1*, and R06C7.7 that induced reversion to correct *lacZ* expression independent of the presence of the mononucleotide repeat in the reporter (Fig. 1), suggesting that the molecular lesion that leads to such reversion to in-frame *LacZ* expression occurs at alternative DNA substrates. Perhaps small insertions or deletion mutations are induced in DNA sequences that are positioned between the ATG and the *LacZ* coding region; the reporter contains an upstream GFP ORF fused to the *LacZ* ORF. To test this possibility directly, we determined the mutation spectrum for three of these genes, M04F3.1, Y71F9AL.18 (*pme-1*), and R06C7.7. RNAi against these genes was performed on *unc-93* worms that carry a dominant mutation in the *unc-93* gene rendering these worms immobile and egg-laying defective (Greenwald and Horvitz 1980). Loss-of-function mutations in *unc-93* results in wild-type movement and egg laying, and thus loss-of-function mutations in this mutational target can easily be isolated and subsequently identified by sequencing the *unc-93* ORF.

Table 2 lists the molecular nature of the mutations that are induced in M04F3.1(RNAi), Y71F9AL.18(RNAi), and R06C7.7(RNAi), which are primarily small deletion or insertion events. Strikingly, these DNA alterations did not preferentially occur at sequences that have a repeated nature, which explains why, in the somatic mutator assay, *LacZ* expression is restored in transgenes that lack the mononucleotide repeat (Fig. 1). This outcome differs from the frameshift-resulting mutations in mismatch repair-deficient animals—those occur solely in repeated nucleotide tracts present in the *unc-93* mutational target (Tijsterman et al. 2002)—and suggests that the processes (or their disturbance) that can trigger such mutational events are functionally different for these different types of mutator genes.

In this study we systematically screened the *C. elegans* genome for genes that protect the genetic information against mutations. This screen was certainly not saturated: the RNAi library covers 87.2% of all predicted genes, and RNAi does not always efficiently result in a loss-of-function phenotype [estimates suggest knock-out phenotypes for 65% of the genes if all phenotypes are taken into account (Fraser et al. 2000; Kamath et al. 2003)]. The screen we describe here does not depend on the silencing of a gene in a given tissue or stage of development: as long as some sublineages are hit, one will see blue patches. Thus it would not be surprising if the success rate is >65%. To evaluate the hit rate of this screen further, we retested a collection of genes (in triplicate) that are known or expected to be involved in various types of DNA metabolic processes (Boulton et al. 2002 and references therein; e.g., base excision repair, nucleotide excision repair, homologous and nonhomologous recombination, DNA damage checkpoint, and DNA replication). Of 73 genes tested, 17 clones scored positive: 13 were previously identified in the genome-wide RNAi screen, but 4 were initially missed, confirming the notion of a certain degree of interexperimen-

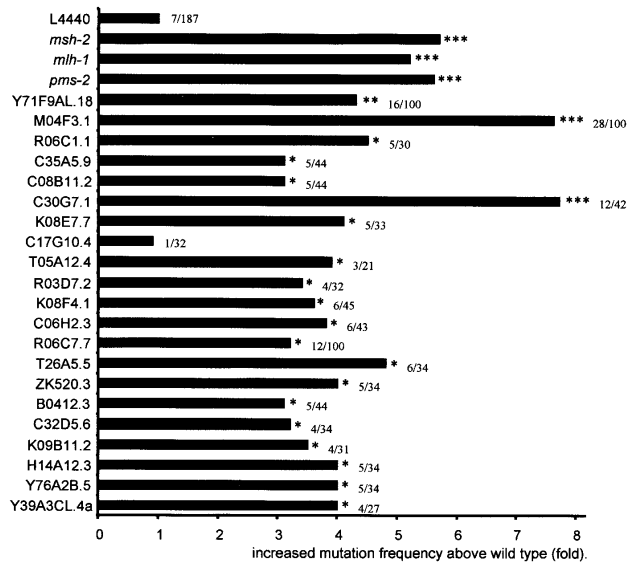


Figure 2. RNAi-mediated germline mutator phenotype in *C. elegans*. An endogenous mutational target, *unc-93* (*e1500*), in which mutations can be recognized based on altered movement of the animal (Greenwald and Horvitz 1980) is used to determine the level of spontaneous mutagenesis levels in the *C. elegans* germline. This assay was performed for 24 genes (Y-axis), several from each category listed in Table 1. As a control, the empty vector L4440 was assayed. The X-axis represents the number of independent *unc-93* (*e1500*) cultures that showed revertant worms. *, $p < 0.05$; **, $p < 0.001$; ***, $p < 0.0001$; using the Chi-square test. The absolute number of cultures that contained revertants versus the total number of cultures analyzed is also indicated. The mutational loads in *msh-2*(RNAi), *mlh-1*(RNAi), and *pms-2*(RNAi) being 18/187, 16/183, and 16/168, respectively, were determined in a separate experiment with a reduced background level of mutations: 3/177.

Table 2. Loss of function of M04F3.1, Y71F9AL.18 and R06C7.7 results in frequent insertion/deletion mutations

RNAi against gene	Frameshift mutations ^a	Mutation type (del = deletions; ins = insertions)	Mutated sequence: <i>unc-93</i> mutational target
M04F3.1	10/12	del. nt 21 del. nt 27 del. A G→ACT TA→ACT G→AT ins. A ins. A ins. T ins. T	TGGCTTTTTCCAGAT actcttctatgatccctgctt CAATATTCTGTGGAA ATATTGACGCGC aatttcatactcctctcagtttggc CACTAAATGCAGC TATGTAGGTATGTAGG a TATGTAGGTATGTA TCTACATTTTGCTA (G→ACT) TTGACAACTTT ACTAACCTTTGTA (TA→ACT) AACTCCAAC GACCAATTTATTAC (G→AT) TGGAAGTCTTC AAAATACTTTGGGCTC A CTCGAAAAACTCT AACTGATTTTTTTGAT A TTTTAATTATTTTC CATGCATTCCGGTAGAT T GATGAAGATAATG AAACATCGGTGGCCG T AGTTTTCCAATTTTC
Y71F9AL.18	6/9	del. nt 1184→ins. nt 523 del. AAAA TA→GAGCCCAAA ins. AC ins. A ins. A	AAATTGAA (del. nt 2147–3331→ins. nt 4358–4880) GGCCACCAC TTTGGAGCGATTCCA aaaa AGAAGGATGATAG CGAGGAGCCCAAA (TA→GAGCCCAAA) TTTTCCTTTTCGAG TTCAGTGGGAAGTTC AC GAAATGATCGTCGT TTTCTCTTCCTGTTC A CTGCATTTAACGGAC AAGAGATTCACAAAA A TCGATATTTAGTGAA TATCACTTGCCATAT catcat TATTTGTTCCATC TATTTTCTATACATT gtcatt AATTTGAGGCCGA AATTGGTCACAATGC aaac TCTCATGCATTCG TTTTTCCAGATACTCT tct ATGATCCCCTGCTTC AAAACCTTCGATT (catcttttgc→GT) AAAAAATTTG TCATATCTCC (tctcagttt→TTTTTTTTTTTTTTTT) GGCCACTAAA CTACATACCTACATAT T CTACAAACCTACAT
R06C7.7	7/10	del. CATCAT del. GTCATT del. AAAC del. TCT del. N10→ins. GT del. N10→ins. T16 ins. T	TATCACTTGCCATAT catcat TATTTGTTCCATC TATTTTCTATACATT gtcatt AATTTGAGGCCGA AATTGGTCACAATGC aaac TCTCATGCATTCG TTTTTCCAGATACTCT tct ATGATCCCCTGCTTC AAAACCTTCGATT (catcttttgc→GT) AAAAAATTTG TCATATCTCC (tctcagttt→TTTTTTTTTTTTTTTT) GGCCACTAAA CTACATACCTACATAT T CTACAAACCTACAT

^aThe remainder are single base pair substitutions (not depicted).

tal variation in high-throughput genome-wide RNAi screens. Also note that the effectiveness of RNAi in knocking down gene expression can be variable for individual genes, and the mutation frequency for the genes we analyzed are therefore not the mutation frequency of null alleles per se. In support, the elevated mutation frequency by *msh-6*(RNAi) is lower than the increase in mutation frequency as a result of a genetic knockout: *msh-6(pk2504)*.

One advantage of screening an RNAi library is that essential genes can also be recognized (Table 1); these are usually overlooked by normal forward screens. In these cases, the blue patches are seen either in sick or sterile animals, or in “escapers”: animals that have escaped the embryonic lethal phenotype but have inherited the RNAi effect, leading to reduction of function of the relevant gene in somatic tissues. In the complete set of genes we found 19 mutator genes that also cause embryonic lethality, reduced brood size, or a severe growth delay (Table 1).

In total, 61 genes were identified. Of these, 82% have clear orthologs in the human genome. This is much higher than the estimated value of all genes that have a counterpart encoded by the human genome (only 53% of the *C. elegans* genes that have been studied genetically or biochemically show homology to genes in other species; Reinke et al. 2000), confirming that the genes protecting the genome’s integrity are ancient and well conserved. Forward mutator screens have been done with bacteria and yeast. An advantage of screening with *C. elegans* is that as a multicellular organism, it can have supplemental mechanisms for DNA protection, which may not exist in unicellular organisms. Indeed, we recognized several genes that are present in human and fly but not in yeast (Table 1), suggesting that multicellular

organisms have indeed evolved additional mechanisms to protect the integrity of the DNA.

Normal somatic cells transform into malignant cancer cells after multiple gene inactivations. These gene inactivations are probably inflicted by the mutator phenotype that many or all cancer cells have. In, for example, mismatch repair-deficient tumors, there is an increased mutation frequency that is often seen as DNA repeat instability. There are still tumor suppressor genes to be discovered; we suggest that among the genes identified in this screen are somatic mutators that cause human cancers.

Materials and methods

Genome-wide RNAi screen

The RNAi library and the RNAi feeding protocol are described by Ahringer and coworkers (Kamath et al. 2001). *E. coli* clones containing dsRNA producing feeding constructs were grown overnight at 37°C in LB containing ampicillin. Agar plates were seeded with these cultures, incubated overnight at room temperature and then seeded with 2–4 worms of strain NL3400: *pkIs1604 [rol-6(su1006) hsp16/2::ATG(A)₁₇gfp/lacZ (out of frame)]* at the L3/L4 developmental stage. Nematodes were grown at 20°C for 5 d, washed off the plates, fixed, and stained for the presence of β-galactosidase with X-gal. Positive feeding clones were tested again in triplicate and with strain NL3401: *pkIs1605 [rol-6(su1006) hsp16/2::gfp/lacZ(out of frame)]*. For all clones that induced the mutator phenotype, plasmids were isolated and the dsRNA encoding DNA sequence was sequenced.

unc-93 reversion assay

On 50 assay plates (9-cm Petri dishes seeded with RNAi food), 30 *unc-93 (e1500)* animals of the L1 stage were placed, grown until food deprivation (the F1 was also directly inspected for revertants to exclude spontaneous mutations in earlier generations; none were found), and subsequently transferred twice to fresh RNAi assay plates (similar numbers of worms

were transferred). Then, cultures were scored for the presence of worms that had reverted to a wild-type phenotype. The number of plates that contain such revertants was taken as a measurement for the germline mutation frequency. For RNAi against genes M04F3.1, Y71F9AL.18, and R06C7.7, this procedure was repeated three times to obtain a sufficient number of mutations that are within the *unc-93* gene (note that inactivation of at least five different loci suppress the *unc-93* phenotype). In all animals, the *unc-93* gene was sequenced.

Acknowledgments

We thank members of the Plasterk lab for discussions and the *Caenorhabditis* Genetic Stock Centre for providing some of the strains used in this study.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Araujo, S.J., Tirode, F., Coin, F., Pospiech, H., Syvaaja, J.E., Stucki, M., Hubscher, U., Egly, J.M., and Wood, R.D. 2000. Nucleotide excision repair of DNA with recombinant human proteins: Definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes & Dev.* **14**: 349–359.
- Boulton, S.J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D.E., and Vidal, M. 2002. Combined functional genomic maps of the *C. elegans* DNA damage response. *Science* **295**: 127–131.
- Burkle, A. 2001. Poly(ADP-ribosylation), a DNA damage-driven protein modification and regulator of genomic instability. *Cancer Lett.* **163**: 1–5.
- Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J., et al. 2002. Genomic instability in mice lacking histone H2AX. *Science* **296**: 922–927.
- Fearon, E.R. and Vogelstein, B. 1990. A genetic model for colorectal tumorigenesis. *Cell* **61**: 759–767.
- Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**: 1027–1038.
- Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. 2000. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**: 325–330.
- Friedberg, E.C., Walker, G.C., and Siede, W. 1995. *DNA repair and mutagenesis*. Am. Soc. Microbiol., Washington, DC.
- Greenwald, I.S. and Horvitz, H.R. 1980. *unc-93*(e1500): A behavioral mutant of *Caenorhabditis elegans* that defines a gene with a wild-type null phenotype. *Genetics* **96**: 147–164.
- Kamath, R.S., Martinez-Campos, M., Zipperlen, P., Fraser, A.G., and Ahringer, J. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2** research0002.1–research0002.10.
- Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrman, M., et al. 2003. Systematic functional analysis of the *C. elegans* genome using RNAi. *Nature* **421**: 286–287.
- Kolodner, R. 1996. Biochemistry and genetics of eukaryotic mismatch repair. *Genes & Dev.* **10**: 1433–1442.
- Kuo, M.H., vom Baur, E., Struhl, K., and Allis, C.D. 2000. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Mol. Cell* **6**: 1309–1320.
- Lachner, M. and Jenuwein, T. 2002. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**: 286–298.
- Leach, F.S., Nicolaidis, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M., et al. 1993. Mutations of a *mutS* homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**: 1215–1225.
- Loeb, L.A. 1991. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**: 3075–3079.
- . 2001. A mutator phenotype in cancer. *Cancer Res.* **61**: 3230–3239.
- Modrich, P. and Lahue, R. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65**: 101–133.
- New, J.H., Sugiyama, T., Zaitseva, E., and Kowalczykowski, S.C. 1998. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* **391**: 407–410.
- Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S., et al. 2000. A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**: 605–616.
- Rosenbluth, R.E., Cuddeford, C., and Baillie, D.L. 1983. A rapid eukaryotic mutagen test system using the reciprocal translocation, *eT1*(III;V). *Mutat. Res.* **110**: 39–48.
- Stegmeier, F., Visintin, R., and Amon, A. 2002. Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. *Cell* **108**: 207–220.
- Tijsterman, M., Pothof, J., and Plasterk, R.H. 2002. Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient *Caenorhabditis elegans*. *Genetics* **161**: 651–660.
- Timmons, L. and Fire, A. 1998. Specific interference by ingested dsRNA. *Nature* **395**: 854.
- Walter, J. and Newport, J. 2000. Initiation of eukaryotic DNA replication: Origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase alpha. *Mol. Cell* **5**: 617–627.
- Wang, Z. 2001. Translesion synthesis by the UmuC family of DNA polymerases. *Mutat. Res.* **486**: 59–70.
- Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. 1997. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89**: 357–364.