Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference

Andrew G. Fraser*†, Ravi S. Kamath*†, Peder Zipperlen*†, Maruxa Martinez-Campos*, Marc Sohrmann‡ & Julie Ahringer*

* *Wellcome/CRC Institute, University of Cambridge, Tennis Court Road, CB2 1QR Cambridge, UK*
‡ *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
† *These authors contributed equally to this work*

Complete genomic sequence is known for two multicellular eukaryotes, the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*, and it will soon be known for humans. However, biological function has been assigned to only a small proportion of the predicted genes in any animal. Here we have used RNA-mediated interference (RNAi) to target nearly 90% of predicted genes on *C. elegans* chromosome I by feeding worms with bacteria that express double-stranded RNA. We have assigned function to 13.9% of the genes analysed, increasing the number of sequenced genes with known phenotypes on chromosome I from 70 to 378. Although most genes with sterile or embryonic lethal RNAi phenotypes are involved in basal cell metabolism, many genes giving post-embryonic phenotypes have conserved sequences but unknown function. In addition, conserved genes are significantly more likely to have an RNAi phenotype than are genes with no conservation. We have constructed a reusable library of bacterial clones that will permit unlimited RNAi screens in the future; this should help develop a more complete view of the relationships between the genome, gene function and the environment.

Complete genomic sequence is an invaluable tool for understanding the molecular mechanisms underlying an organism's development and function. The nematode worm *C. elegans* is one of two multicellular eukaryotes for which essentially complete genomic sequence is known[1,2]. Thirty-six per cent of predicted *C. elegans* genes have a significant human match[1,3] including many genes implicated in human disease[3,4], and functional analysis of the *C. elegans* genome has shed light on many conserved biological processes and molecular pathways. A comprehensive functional analysis of all genes in *C. elegans* would greatly expand our knowledge of conserved gene function. We therefore decided to investigate systematically loss-of-function phenotypes of predicted genes of *C. elegans*, starting with chromosome I.

RNA-mediated interference (RNAi) transiently inhibits the activity of a gene by the introduction of double-stranded RNA (dsRNA) of sequence specific to the targeted gene[5]. The specificity and potency of RNAi make it ideal for investigating gene function beginning only with genomic sequence[6]. Ingestion of dsRNA-expressing bacteria results in RNAi of the targeted gene[7], and we previously established that this technique is at least as effective as the injection of dsRNA for RNAi[8]: embryonic lethal phenotypes are detected with similar efficiency by feeding and injection, but feeding detects over 50% more post-embryonic phenotypes than injection. It is thus possible to make a library of dsRNA-expressing bacteria that can be used for high-throughput genome-wide RNAi screens at very low cost. It is important to note that because RNAi does not efficiently inhibit all genes, an RNAi-based screen will miss some

Table 1 Summary of phentypes arising from RNAi of genes on chromosome I

Phenotype		Number	Per cent
All phenotypes	Total	339	13.9
Embryonic lethal	Emb	226	9.2
Sterile	Ste	82	3.4
	Stp	14	0.6
Developmental delay	Gro/Lva	145	5.9
Larval lethal	Lvl	38	1.6
	Unc	70	2.9
	Pvl	29	1.2
	Bmd	27	1.1
	Dpy	19	0.8
	Clr	14	0.6
Post-embryonic	Him	13	0.5
	Rup	9	0.4
	Mlt	8	0.3
	Pr2	8	0.3
	Egl	5	0.2
	Sck	5	0.2
	Bli	4	0.2
	Muv	2	0.1
	Rol	2	0.1
	Adl	1	<0.1
	Lon	1	<0.1
	Hya	1	<0.1

The number of predicted genes whose targeting through RNAi gave rise to each phenotype is shown. Percentages are given as percentage of total number of clones screened (2,445). Phenotypic classes were defined as described in Methods. the phenotypes are Emb (embryonic lethal), Ste (sterile), Stp (sterile progeny), Gro (slow post-embryonic growth), Lva (larval arrest), Lvl (larval lethality), Unc (uncoordinated), Pvl (protruding vulva), Bmd (body morphological defects), Dpy (dumpy), Clr (clear), Him (high incidence of males), Rup (ruptured), Mlt (molt defects), Prz (paralyzed), Sma (small), Egl (egg-laying defective) Sck (sick), Bli (blistering of cuticle), Muv Multivulva), Rol (roller), Adl (adult lethal), Lon (long) and Hya (hyperactive).

Table 2 Detection of forward genetic loci on chromosome I by RNAi

Phenotype	Genetic loci fed	Possible to detect	RNAi phenotype detected	Published phenotype detected
All phenotypes	62	50	31	25
Embryonic lethal	21	21	19	16
Sterile	3	3	2	2
Sterile progeny	4	4	1	1
Developmental delay	0	0	–	–
Larval lethal	4	4	1	1
Post-embryonic	43	31	14	9

RNAi phenotypes were compared with those of genes that have known loss-of-function phenotypes. 'Genetic loci fed' denotes the number of genes in each category that were analysed by RNAi. 'Possible to detect' denotes the number of genes that have a loss-of-function phenotype that would have been detectable in our screen. 'RNAi phenotype detected' gives the number of genes for which a phenotype was identified. 'Published phenotype detected' gives the number of genes for which the RNAi phenotype matched a published phenotype. See Supplementary Information Table 2 for full data. Some of the differences between RNAi phenotypes and published phenotypes might be because RNAi can reduce both maternal and zygotic gene activities.

relevant genes. Despite this caveat, RNAi is a useful screening tool to complement classical forward genetics.

Analysis of the function of genes on chromosome I by RNAi

We constructed a library of bacteria expressing dsRNA corresponding to genes on chromosome I. Chromosome I is the second smallest chromosome, has few duplicated gene clusters and has no striking unusual features[1]. Each individual bacterial clone is able to synthesize dsRNA designed to target a single gene; because gene predictions are still changing, a few primer pairs no longer correspond to single genes (see Methods). In total, the resulting library contains 2,445 independent clones, corresponding to 2,416 predicted genes, a total of 87.3% of the 2,769 currently predicted genes of chromosome I.

We screened the library to identify genes whose inhibition gives a clearly detectable phenotype in wild-type worms (see Methods). We were able to assign a phenotype to 13.9% of the analysed genes, raising the number of sequenced genes on chromosome I with known phenotypes from 70 to 378 (Table 1). Many genes have more than one associated phenotype, reflecting that genes frequently have

Table 3 Partial list of RNAi phenotypes of genes on chromosome I

GenePairs		Locus	Description	Emb	Ste	P1	P2	P3	Dev	H
Chromosome dynamics										
B0511.8			CDC1-like	+	-	-	-	-	Lva	+
C37A2.4		cye-1	cyclin e	+	-	Clr	-	-	-	+
C41G7.2	*		kinesin	+	-	Him	-	-	-	+
C53H9.2			chrom stability	+	-	-	-	-	Gro	+
F57B10.12		mei-2	katanin homol	+	-	-	-	-	-	+
M01E11.6	*		kinesin	+	-	Him	-	-	Gro	+
R06C7.8			Bub1-like	+	-	Pvl	Rup	Lvl	-	+
T01G9.5		mei-1	ATPase	+	-	-	-	-	-	+
W09G3.3	*		RCC1 domains	-	-	Him	-	-	-	+
Y110A7A.d			cdc27 homologue	+	-	-	-	-	Gro	+
Y39G10A_246.e			MCM4-like	+	-	-	-	-	-	+
Y39G10A_246.i			INCENP-like	+	-	-	-	-	-	+
Y47G6A_247.i			pombe Rad2 homol	+	-	-	-	-	-	+
Y52B11A.9			Kin17 homol	+	-	Bmd	Rup	-	-	+
Cell structure										
C01G8.5			Ezrin-like	-	-	Unc	Lvl	-	Gro	+
C10H11.1			villin	-	-	Him	-	-	-	+
C17E4.9			PDZ domain	+	+	Unc	-	-	-	+
C32E8.10		unc-11	vesicle reg	+	-	-	-	-	-	+
C45G3.1			actin-binding	+	-	Unc	-	-	-	+
C47B2.3	*	tba-2	tubulin	+	-	-	-	-	-	+
C53D5.a			nuclear import	+	+	-	-	-	Gro	+
C53D5.i			nuclear import	+	-	-	-	-	Gro	+
DY3.2		lam-1	nuclear lamin	+	+	Lvl	-	-	-	+
E03H4.8	*		beta coatomer-like	-	+	Unc	Stp	Clr	Gro	+
F07A5.7		unc-15	paramyosin	-	-	Unc	Prz	Egl	-	+
F20G4.3		nmy-2	non-muscle myosin	+	+	-	-	-	-	+
F21C3.5			MT nucleation	+	-	Unc	Bmd	-	Gro	+
F26B1.3			karyopherin	+	-	Lvl	-	-	-	+
F26E4.8	*	tba-2	tubulin	+	-	-	-	-	-	+
F26H9.6			ras superfamily	+	+	Lvl	-	-	-	+
F28H1.2			calponin domain	+	-	-	-	-	Lva	+
F30A10.6			transporter	-	-	-	-	-	Gro	+
F36H2.1			cation transporter	+	-	-	-	-	Gro	+
F43G9.10			microfib assoc	+	-	-	-	-	Gro	+
F46F11.5			vacuolar ATPase	+	+	Lvl	-	-	-	+
F53B8.1	*		plectrins	+	-	Unc	Prz	Lvl	-	+
F53F10.5			nucleoporin-like	+	-	-	-	-	-	+
F54C1.7			troponin c	-	+	-	-	-	-	+
F55A12.7			UNC-101 homol	-	-	Unc	-	-	Gro	+
F55F8.5			MT associated	-	-	-	-	-	Lva	+
F56F4.5			transporter	-	-	Him	-	-	-	+
H15N14.1			human NSF-like	+	+	-	-	-	-	+
M01A10.3			ribophorin	+	+	Lvl	Unc	-	-	+
R05D11.3			NTF2 homol	+	+	-	-	-	Gro	+
T03F1.9			UNC-89-like	+	-	-	-	-	-	
T19B4.2			NUP153-like	+	+	-	-	-	-	+
T21E12.4		dhc-1	dynein heavy chain	+	-	-	-	-	-	+
T25G3.2			chitin synthases	+	-	-	-	-	Gro	+
T26E3.3		par-6	PDZ domain	+	-	-	-	-	-	+
W02B9.1		hmr-1	cadherin	+	-	Bmd	Unc	Dpy	-	+
W04C9.1			ABC transporter	+	-	-	-	-	Gro	+
Y105E8C.n			γ-adaptin	+	-	Unc	Lvl	-	Gro	+
Y18D10A.17			sup of clathrin defic	+	-	-	-	-	-	+
Y18D10A.20		pfn-1	profilin	+	-	-	-	-	-	+

GenePairs		Locus	Description	Emb	Ste	P1	P2	P3	Dev	H
Cell structure										
Y34D9A_152.a			vacuolar sorting	+	-	Unc	Prz	Lvl	Gro	+
Y48G8A_3945.e			adaptin subunit	-	-	Unc	-	-	-	+
Y71A12B.a			gravin-like	+	-	Dpy	Clr	-	-	+
Y71F9A_279.b			NXT1 homol	+	-	Pvl	Unc	-	-	+
Y71F9A_282.b			coatomer subunit	+	+	Unc	-	-	-	+
Y71F9A_290.a			NTF2 homol	-	-	Pvl	Rup	Clr	-	+
Y71G12A_195.e			talin	+	+	Unc	Prz	-	-	+
Y87G2A.s			HuVPS28 homol	-	-	Unc	-	-	-	+
Y87G2A.x			protein trafficking	-	-	Unc	-	-	Gro	+
Y87G2A.y			protein trafficking	-	-	Clr	-	-	-	+
ZK1014.1			vesicle fusion	-	+	Lvl	-	-	-	+
ZK1151.2			spectrin repeats	-	-	Unc	Bmd	Stp	-	+
Specific transcription										
B0025.3			txnl corepressor	+	-	-	-	-	Gro	+
C01G8.7			eyelid-like	+	+	Bmd	Lvl	-	Gro	+
C01G8.8			eyelid-like	+	-	Bmd	-	-	-	+
C01H6.5		nhr-23	nuc horm recep.	-	-	Unc	Lvl	Dpy	-	+
C12C8.3		lin-41	NHL domains	-	+	-	-	-	-	+
C32F10.7		nhr-2	nuc horm recep.	-	-	Him	-	-	-	+
C48E7.3			bZIP	-	-	-	-	-	Gro	+
D1081.2			SRF homol	-	-	Unc	Prz	-	-	+
F43G9.12			TCF-9-like	+	-	-	-	-	Gro	+
F52F12.6			MYT1 homol	-	-	Unc	-	-	-	+
F55F8.4			txnl repression	+	-	-	-	-	Lva	+
F57B10.1			bZIP	+	+	Sma	Dpy	-	Lva	+
K02B12.1		ceh-6	homeobox	-	-	Unc	Mlt	-	-	+
M05B5.5		hlh-2	bHLH	+	+	Unc	Pvl	-	-	+
W02D3.9		unc-37	groucho family	+	+	Unc	-	-	-	+
W03D8.4		pop-1	HMG box	+	+	-	-	-	-	+
Y40B1A.4			Zn finger	-	-	Unc	Bmd	-	-	+
Y54E5B.3			Mediator complex	+	-	Rup	-	-	-	+
Y65B4A_179.b			txnl activator	-	-	Unc	Dpy	-	-	+
ZC247.3		lin-11	LIM homeodomain	-	-	-	-	-	Gro	+
ZK858.4		mel-26	kruppel-like	+	-	-	-	-	-	+
Signalling										
C09D1.1		unc-89	multiple domains	-	-	Unc	-	-	-	+
C10H11.9		let-502	ROCK	+	-	Dpy	Rol	Lvl	-	+
C26C6.2		goa-1	G-a subunit	+	+	Unc	Pvl	Egl	-	+
C32E8.5			FHA domain	+	+	-	-	-	Gro	+
F26E4.1			PP2A reg subunit	+	-	Lvl	-	-	-	+
F55A12.3			PI-4P 5' kinase	-	+	-	-	-	-	+
F55C7.4		unc-73	GEF	+	-	-	-	-	-	+
F55C7.7		unc-73	GEF	-	-	Egl	-	-	-	+
K04G2.8		apr-1	APC homol	-	-	Unc	Bmd	Lvl	-	+
K05C4.6		hmp-2	β-catenin	+	-	Unc	Dpy	Bmd	-	+
K12C11.2			SUMO-1 like	+	-	Pvl	-	-	-	+
T01G9.6		kin-10	CKIIbeta subunit	+	-	Pvl	-	-	Gro	+
T21E3.1	*		PTPase	+	+	-	-	-	-	+
T23D8.1		mom-5	frizzled-like	-	-	Unc	Bmd	-	-	+
T23H2.5	*		rab family	-	-	Sma	Clr	-	-	+
Y106G6E.6			Casein Kinase 1	+	-	Dpy	-	-	Gro	+
Y18D10A.5		gsk-3	GSK-3	+	-	-	-	-	-	+
ZC581.1			NIMA-like kinase	-	-	Unc	Lvl	Rol	-	+
ZK265.6			G prot coup recep	-	-	-	-	-	Gro	+

RNAi phenotypes are shown for genes in the following functional classes: chromosome dynamics and cell cycle; cell structure; specific transcription; and signal transduction. For each gene, the following data are shown: the Research Genetics GenePairs name; whether a number of paralogues might be targeted (asterisk in column 2; see methods for criterion); the corresponding genetic locus name if it exists; a short description of gene function; the RNAi phenotype in which embryonic lethality (Emb), fecundity (Ste), post-embryonic phenotypes (P1–3) and developmental delay (Dev) are shown separately. Emb and Ste are classified into weak (white box, black '+') or strong (black box, white '+') phenotypes. For Emb, weak is 10–80% embryonic lethality, strong is 90% embryonic lethal or more; weak Ste denotes a brood size of 1–10, whereas strong Ste is totally sterile. Column H shows whether there is a match (white box, black '+') or a homologue (black box, white '+') in *Drosophila melanogaster*, *Saccharomyces cerevisiae* or humans. Phenotypic abbreviations are given in legend to Table 1. The GenePairs name does not always correspond with the current predicted gene name as gene predictions change.

several functions in the organism. Furthermore, as we examined both worms that were exposed to dsRNA only as larvae or adults as well as their progeny, we could assign post-embryonic phenotypes to genes that result in sterility or produce 100% embryonic lethal progeny. A summary of these results and a partial listing of the phenotypes obtained are given in Tables 1 and 3. Full results are given in Supplementary Information Table 1 and are publicly accessible in WormBase (http://www.wormbase.org).

Our screen was sufficiently effective to identify 90% of known embryonic lethal genes. In addition, we were able to assign phenotypes to 45% of genes with a known post-embryonic phenotype that should have been detectable in our screen (Table 2; and Supplementary Information Table 2). However, we failed to find phenotypes for a number of previously characterized genes. In some cases (for example, *fog-3*), this was not due to an inherent difficulty in inhibiting the genes using RNAi (as we obtained the correct phenotype in a separate experiment), but simply because we overlooked them in the screen. However, only one of eight genes involved in neuronal function gave a detectable RNAi phenotype; this agrees with our finding that neurons appear to be more resistant to RNAi than are other cell types[8]. Similarly, we did not detect phenotypes for several genes involved in sperm development (*fer-1*, *spe-9* and *spe-11*).

The largest phenotypic class, comprising over 60% of the genes, are those whose inhibition by RNAi gives rise to embryonic lethality—the Emb genes; these include a large number of components of the basal cellular machinery. More notably, we find a homologue of the SMN human disease gene[9], a variety of genes encoding RNA-binding proteins (several such proteins have a role in early polarity; reviewed in ref. 10), a number of genes involved in chromosome condensation and separation, components of signal transduction pathways and many conserved genes that have no known biochemical function.

The largest class of post-embryonic phenotype is the Uncoordinated (Unc) class. Unc phenotypes arise from defects in the development or function of the neuromuscular system (reviewed in ref. 11). We find Unc genes encoding proteins involved in vesicle sorting and fusion as well as transcription factors (including a homologue of the zinc finger transcription factor MYT-1 which is only expressed in developing neurons in mammals[12–14]) and components of the cytoskeleton (for example a kakapo[15–18] and a talin[19] homologue).

A number of genes showed a high incidence of males (Him) phenotype. *C. elegans* is usually grown as a self-fertilizing hermaphrodite with males arising at a low frequency in wild-type cultures owing to non-disjunction of the X chromosome (hermaphrodites have two X chromosomes, males only one). An increased number of males is indicative of either the incorrect segregation and maintenance of chromosomes in the germ line (reviewed in ref. 20) or defects in sexual specification. The Him genes that we identified include kinesins, a katanin homologue[21,22] and a nuclear hormone receptor.

Conservation of genes with RNAi phenotypes

We examined the level of cross-species conservation of the genes for which we detected an RNAi phenotype (Fig. 1). To find *C. elegans* genes that are conserved in other species, we identified *C. elegans* genes that have hits with BlastP[23] expectation values below 1.00×10^{-06} in *Saccharomyces cerevisiae*, *Drosophila melanogaster* or humans; we defined these as a 'match'. Hits with BlastP e-values

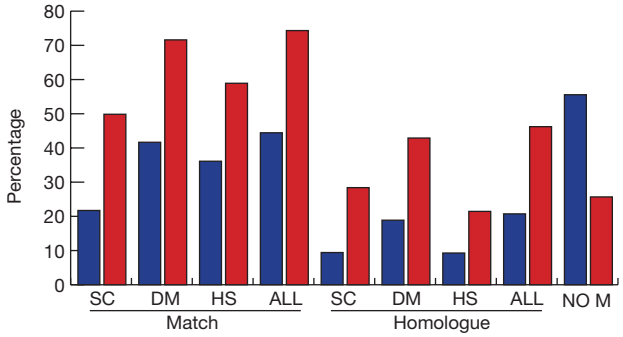

Figure 1 Conservation of genes with an RNAi phenotype. Matches or homologues of *C. elegans* genes were identified as described in the text. Shown are percentages of all genes (blue bars) or genes with RNAi phenotypes (red bars) with matches or homologues in *S. cerevisiae* (SC), *D. melanogaster* (DM), human (HS), or all three combined (ALL), or with no matches in any organism (NO M). The significance of the differences between the percentages of genes and the percentages of genes with RNAi phenotypes that have homologues is $P < 0.001$ in all cases, except for comparison with human homologues for which it is $P < 0.1$.

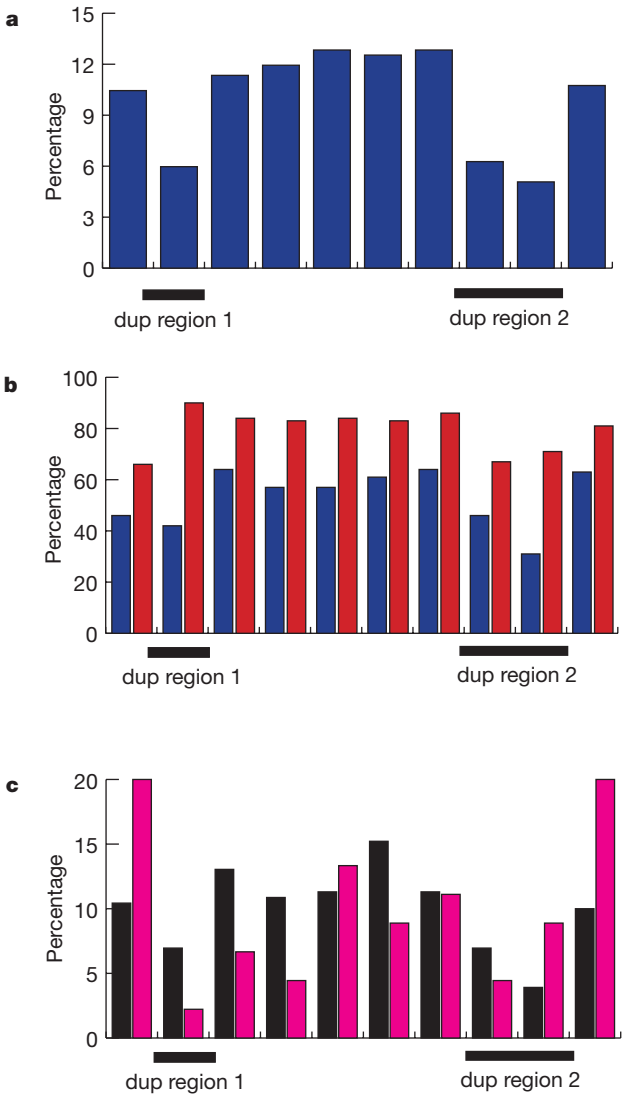

Figure 2 Distribution on chromosome I of genes with RNAi phenotypes and genes with ESTs. In each panel, chromosome I was analysed in 10 consecutive portions, each containing 10% of predicted genes. **a**, Percentage of all genes with an RNAi phenotype that are in each portion. **b**, Percentage of all predicted genes that have an EST (blue bars) or of genes that gave an RNAi phenotype that have an EST (red bars) in each portion. **c**, Percentage of Emb genes (black bars) or genes with viable post-embryonic phenotypes (pink bars) in each chromosomal portion. The 'dup region' indicates the approximate location of regions containing local duplications.

below 1.00×10^{-10} and in which the conservation extends over at least 80% of the *C. elegans* protein length, we defined as 'homologues'; this category includes orthologues. This provides a conservative estimate of the number of genes with regions of conservation (matches) or homologues, respectively.

We found that genes with RNAi phenotypes were much more likely to have a match ($P < 0.001$) compared with all genes (Fig. 1). Most striking is the similarity that we see between *C. elegans* and *Drosophila*: whereas 42% of *C. elegans* genes have a match and 19% have a homologue in *Drosophila*, over 72% of genes with an RNAi phenotype have a *Drosophila* match and 43% have a homologue (Fig. 1). This analysis shows that genes with a required function in *C. elegans* have been highly conserved across eukaryotic evolution. We also find that highly conserved genes are more likely to have an RNAi phenotype than genes that show no conservation: 26% of *C. elegans* genes that have a homologue in one of the organisms examined give an RNAi phenotype, compared with only 5% of genes with no conservation ($P < 0.001$).

Distribution of genes with RNAi phenotypes

Genes for which we identified an RNAi phenotype are evenly distributed across the chromosome with the exception of two regions (corresponding to segments 2 and 8–9 in Fig. 2a) for which there appears to be a drop in number. These two regions correspond to the two regions of chromosome I that contain locally duplicated gene clusters[1]. We suggest that the reduction in the number of phenotypes observed by RNAi in these regions may be due to gene duplication and thus redundancy of function. It is worth noting that some of the predicted genes in the duplicated regions may not be expressed: whereas genes with RNAi phenotypes are equally likely to have an expressed sequence tag (EST) in all regions of the genome (see below), there is a significant drop ($P < 0.05$) in the proportion of total genes with ESTs in the second locally duplicated gene cluster region (Fig. 2b; 39% of genes in the second cluster have an EST compared with 53% over the entire chromosome). We suggest that a portion of the predicted genes in such regions of duplication may in fact be pseudogenes.

Genes that give RNAi phenotypes are much more likely to have an EST than genes on chromosome I in general (82% compared with 53% respectively, $P < 0.001$; Fig. 2b). The relatively high percentage of genes with RNAi phenotypes that have ESTs may reflect that these genes are expressed at higher levels. It may also be that many genes that lack ESTs are only expressed conditionally; we are unlikely to have found phenotypes for such genes.

In *C. elegans*, there is evidence of differences between the chromosome arms and the central regions (the clusters), suggesting that there might be differences in gene type or function across the chromosome[24]. In general, the distribution of genes in any given phenotypic class was similar to that for all genes with an RNAi phenotype (such as Emb genes; compare Fig. 2c and a). However, genes with viable post-embryonic phenotypes (Pep genes)—those that gave a post-embryonic phenotype without any embryonic or post-embryonic lethality, sterility or developmental delay—show a trend toward enrichment at the arms of chromosome I. It has been suggested that the chromosome arms may be more prone to mutation and recombination than the central core portion[24] and, if so, that novel gene functions are more likely to evolve in such regions. Our finding that genes which uniquely affect post-embryonic development cluster at the arms supports this model.

Biochemical function and RNAi phenotype

To explore the relationship between the biochemical function of a gene product and its mutant phenotype, we categorized the sterile (Ste), embryonic lethal (Emb), uncoordinated (Unc) and viable post-embryonic phenotype (Pep) genes into the functional classes shown in Fig. 3a.

Unsurprisingly, genes involved in basal metabolic processes account for ~50% of Ste and Emb genes (Fig. 3a); this confirms

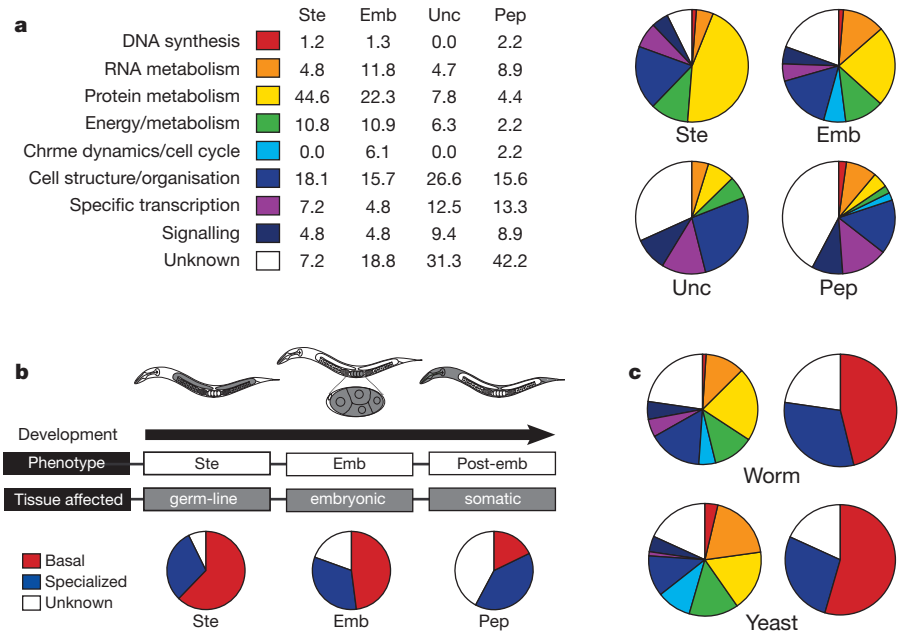

Figure 3 Functional classes of Emb, Ste, Unc and Pep genes. **a**, Predicted products of genes that gave Ste, Emb, Unc or viable post-embryonic (Pep) RNAi phenotypes were placed into functional classes as described in Methods. Genes whose products could not be accurately classified into any of the eight functional classes were placed into the unknown category (white). Numbers denote the percentage of genes in each functional class; pie charts illustrate these numbers graphically. **b**, Pie charts show distributions of predicted gene products grouped as follows: basal metabolic category (red) comprises the classes of DNA, RNA, protein and intermediate metabolism; specialized functions (blue) comprises cell-cycle and chromosome dynamics, cell biology and cellular structure, gene-specific transcription factors and signal transduction. Worms show the tissue affected in each phenotypic class shaded in grey. **c**, Distribution of genes giving rise to non-viable RNAi phenotypes in *C. elegans* (worm) or to non-viable phenotypes following disruption in *S. cerevisiae* (yeast).

that these basic biochemical processes are indeed essential for viability. In contrast, under 20% of Unc and Pep genes encode components of the basal metabolic machinery, whereas more than twice as many encode proteins with more specialized functions (Fig. 3a, b). There is thus a clear difference between the types of gene required for germline function or embryonic viability (which mainly require basal machinery) and those involved in later developmental processes which appear to require proteins of either more specialized functions or as yet unknown function (Fig. 3b).

A second clear trend is that the number of genes of unknown function increases greatly in the Unc and Pep genes, making this the largest overall class for those phenotypes (Fig. 3). This shift underlies the fact that, although we know a great deal about basic metabolic processes of eukaryotic cells (and thus can readily ascribe function to a large proportion of Ste and Emb genes), much is still to be learnt about the complex processes and the genes that regulate the development and function of a multicellular eukaryote. A significant number (~25%) of genes of unknown function have close homologues in *Drosophila* or humans; further study of these may shed light on conserved processes specific to animals.

Essential genes of *S. cerevisiae* and *C. elegans*

Saccharomyces cerevisiae was the first eukaryote to be completely sequenced[25] and reverse genetics has been used extensively to investigate *S. cerevisiae* gene function. In a set of 3,680 genes knocked out by targeted disruption, 890 affect viability[26]; we compared these genes to those that gave different RNAi phenotypes in *C. elegans*. Yeast and worm genes important for viability have a similar distribution within the different functional classes, but are different from the Unc or Pep distributions (Fig. 3c; also compare 3a and b). This suggests that similar types of gene are required for viability of yeast and animal cells. A striking difference ($P < 0.001$) is that only ~1% of the genes required for viability in yeast are transcription factors, whereas for *C. elegans* it is ~4% (a similar percentage of the genomes of yeast[27] and *C. elegans*[2] encode transcription factors, 3.3% and 2.5%, respectively). This suggests that many of the *C. elegans* transcription factors required for viability may be involved in developmental processes.

Estimate of the size of functionally non-redundant genome

What do our data tell us about the size of the functionally non-redundant genome? We screened 12.7% of the *C. elegans* genome and found that 339 genes gave a clearly discernible phenotype. Taking into account the sensitivity of our screen and scaling up to the entire genome, we estimate that ~5,400 genes will be individually required for wild-type *C. elegans* development under standard laboratory conditions (~2,300 genes for embryonic viability and ~3,100 post-embryonically; see Methods for calculation). This is comparable to previous estimates based on forward genetics[28]. We expect that phenotypes for other genes will be identified under different conditions (for example, environmental stress), in other genetic backgrounds, or using more refined and restricted screening conditions.

Discussion

We have taken a systematic approach to identify functions for the predicted genes of *C. elegans* chromosome I. This is the first large-scale reverse genetic analysis of a multicellular organism and has increased by fivefold the number of sequenced genes with known phenotypes on this chromosome.

Although we have identified RNAi phenotypes for many genes, some will have eluded our screen for one of at least two reasons. First, RNAi may have been ineffective against the targeted gene. RNAi does not accurately phenocopy the null phenotype of all genes (such as genes involved in neuronal function), and may result in either partial or no loss of function. It should also be noted that if several genes have regions of identical or near-identical nucleotide sequence, RNAi could target them simultaneously, so that the observed phenotype may be the result of the inhibition of more than one gene. Second, we will not have detected either subtle or conditional phenotypes. However, we anticipate that future RNAi-based screens using specific assays should be able to detect phenotypes for many more genes, thus increasing our understanding of *C. elegans* and hence of metazoan biology in general. As our library consists of bacterial clones that can be replicated, and the feeding protocol is relatively simple compared with injection, the library can be used repeatedly at low cost and high efficiency for such screens. In addition, we expect that a feeding library and database of associated phenotypes will prove valuable for the positional cloning of genes; currently there are over 300 genes on chromosome I identified by mutation but not yet cloned.

Although the time needed for an RNAi screen using our bacterial library is similar to that for a classical genetic screen, the two approaches have different advantages and will yield different results. Both approaches can be used to screen the entire genome for genes involved in a particular process, and both may identify complete or partial loss-of-function phenotypes. Classical forward genetics generates stable mutant lines that can be maintained indefinitely; furthermore, whereas some genes are resistant to RNAi, all genes are sensitive to mutagens (albeit to a greater or lesser degree) and could thus be cloned using a classical screen. Also, some mutants isolated by forward genetics are due to gain-of-function mutations, which cannot be generated by RNAi. However, the positional cloning of a gene is often slow and laborious. RNAi has the disadvantages mentioned above, but it also has the key advantage of all reverse genetics: the sequence of the gene is already known, and thus any mutant phenotype observed is automatically connected to a known sequence.

In the future, we aim to extend our library construction and functional analysis to the entire *C. elegans* genome and anticipate that the possibility of genome-wide RNAi screening, in conjunction with other functional genomics approaches such as expression analyses using microarrays[29] and two-hybrid experiments[30] will accelerate *C. elegans* research. □

Methods

Generation and cloning of PCR products

PCR products were synthesized using BioTaq polymerase (Bioline) in a reaction containing 25 ng of *C. elegans* genomic DNA, 20 pmol of *C. elegans* GenePairs primers (Research Genetics) and 100 μM dNTPs: 34 cycles of (94 °C 30 s, 58 °C 30 s, 72 °C 90 s) were followed by an extension of 1 h at 72 °C to enhance A-tailing of products. Products were ligated into linearized T-tailed L4440 vector[7] and transformed into the HT115(DE3) bacterial strain (L. Timmons and A. Fire, personal communication) using standard methods. Colonies containing correct sized insert were identified by PCR using vector-specific oligos, and the cloned inserts confirmed by PCR using the original Research Genetics primer pair. Primer sequences are available at http://cmgm.stanford.edu/~kimlab/primers.12-22-99.html.

RNAi screening

RNAi was performed essentially as described in ref. 8, where feeding data were reported on 86 of the 2,445 genes described here. Briefly, 4 wells of a 12-well plate containing NGM agar + 1 mM IPTG + 25 μg ml^{-1} carbenicillin were inoculated with bacterial cultures grown 8–18 h for each targeted gene. Between 10 and 15 L3–L4 stage worms were placed in the first of the 4 wells for each gene and left for 72 h at 15 °C. Three worms, now young adults, were removed and individually placed on three remaining wells for each gene and allowed to lay embryos for 24 h at room temperature; the three worms were then removed (t = 0). The phenotypes of adults and progeny remaining in the first well were scored as well as of the progeny in wells 1–3. Our screen was not ideal for detection of phenotypes visible only in adults (for example, egg-laying defective and progeny sterile); we will have missed some of these. Phenotypic analysis of lethality/sterility was carried out at t = 24 h and post-embryonic phenotypes were analysed by two independent observers at t = 36, 48, 60 and 72 h. Phenotypic classes were defined as follows. Embryonic lethal (Emb) reproducibly has 10–100% embryonic lethality; sterile (Ste) has a brood size of less than or equal to 10 (wild-type worms in these conditions typically give over 50); progeny sterile (Stp) has a brood size of less than or equal to 10 in the progeny of fed worms. Post-embryonic phenotypes require at least 10% of the analysed worms to display a given phenotype; phenotypic classes are given in the legend to Table 1. A full listing of phenotypes obtained is given in Supplementary Information Table 1; genes that we did not clone, and thus did not analyse, are given in Supplementary Information Table 3. Thus,

any GenePairs absent from both Supplementary Tables 1 and 3 was fed, and did not give a detectable mutant phenotype.

Bioinformatic analyses and categorization of genes

Analyses were carried out on GenePairs predictions rather than currently predicted genes as although gene predictions change, phenotypes will always match the GenePairs. About 95% of GenePairs genes have a one-to-one match with a currently predicted gene. Current gene predictions that are targeted for RNAi by the primer pairs were identified by comparing electronic PCR (ePCR) fragments (generated using the ePCR program (ftp.ncbi.nlm.nih.gov/pub/schuler/e-PCR)[31] on the whole chromosome DNA files from the WS9 release of ACeDB (ftp.sanger.ac.uk/pub/wormbase) to gene predictions in ACeDB. To identify additional genes that might be targeted for RNAi by a particular clone we found those with an overlap of 200 base pairs or more with greater than 80% nucleotide identity with the predicted PCR product (asterisks in column 2 of Table 3 denote GenePairs that have such matches); however, it is not yet known what level of identity is required for RNAi.

To find *C. elegans* genes with conservation in other organisms, BlastP[23] was carried out for each individual *C. elegans* gene on chromosome I against *S. cerevisiae, Drosophila melanogaster* and human sequences. The databases used were as follows: *C. elegans* (18,337 entries), *S. cerevisae* (6,191 entries) and *D. melanogaster* (13,743 entries) downloaded on 1 June 2000 (http://www.ebi.ac.uk/proteome); and *Homo sapiens* (35,723 entries, confirmed peptides) downloaded on 1 June 2000 (http://www.ensembl.org). NCBI-Blast2 was used (BLASTP 2.0.6) with the SEG filter, and the search space was set to 7,947,758.

We defined 'sequenced genes with a known phenotype' as being those with a named entry in ACeDB that also have a known phenotype entered in ACeDB, WormBase (http://www.wormbase.org) or the Proteome database (http://www.proteome.com). EST data were supplied by the Sanger Centre on 21 June 2000.

Predicted gene products were placed into functional classes by manual inspection, primarily using data from Proteome, InterPro and Blast analysis[23,32]. The functional classes are: (1) DNA synthesis; (2) RNA synthesis and processing including general transcription machinery, splicing/processing, RNA binding and regulation of chromatin; (3) protein synthesis and proteolysis including translation, degradation and folding; (4) metabolism including energy production and intermediary metabolism; (5) cell-cycle and chromosome dynamics; (6) cell biology and cellular structure including cell junction/adhesion, cytoskeleton, ion channels, protein trafficking and vesicle regulation, and cell polarity; (7) gene-specific transcription; and (8) signal transduction including kinases, phosphatases and components of signal transduction pathways. The 'unknown' functional class contains either genes that have motifs about which there is insufficient information to assign a function, or genes with no significant matches in any organism.

Estimates of non-redundant genome size were done as follows. We detected 90.5% of genes known to give an embryonic lethal phenotype and 32.6% of genes known to give a post-embryonic phenotype. After screening 87.3% of the genes on chromosome I, we identified 226 Emb genes and 113 genes that only gave a post-embryonic RNAi phenotype (including steriles); adjusting for our efficiencies of detection, we estimate that on chromosome I, 286 genes should be required for viability and ~397 for post-embryonic processes. We screened 12.7% of the genome, and thus for the entire genome we expect 2,250 Emb genes and 3,130 total genes to have a post-embryonic phenotype.

Received 20 July; accepted 18 October 2000.

1. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282,** 2012–2018 (1998).
2. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster. Science* **287,** 2185–2195 (2000).
3. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287,** 2204–2215 (2000).
4. Ahringer, J. Turn to the worm! *Curr. Opin. Genet. Dev.* **7,** 410–415 (1997).
5. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans. Nature* **391,** 806–811 (1998).
6. Tabara, H., Grishok, A. & Mello, C. C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282,** 430–431 (1998).
7. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395,** 854 (1998).
8. Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *C. elegans. Genome Biol.* (in the press).
9. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80,** 155–165 (1995).
10. Bowerman, B. & Shelton, C. A. Cell polarity in the early *Caenorhabditis elegans* embryo. *Curr. Opin. Genet. Dev.* **9,** 390–395 (1999).
11. Bargmann, C. I. Neurobiology of the *Caenorhabditis elegans* genome. *Science* **282,** 2028–2033 (1998).
12. Kim, J. G. & Hudson, L. D. Novel member of the zinc finger superfamily: A C2-HC finger that recognizes a glia-specific gene. *Mol. Cell Biol.* **12,** 5632–5639 (1992).
13. Armstrong, R. C., Kim, J. G. & Hudson, L. D. Expression of myelin transcription factor I (MyTI), a "zinc-finger" DNA- binding protein, in developing oligodendrocytes. *Glia* **14,** 303–321 (1995).
14. Kim, J. G. *et al.* Myelin transcription factor 1 (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system. *J. Neurosci. Res.* **50,** 272–290 (1997).
15. Strumpf, D. & Volk, T. Kakapo, a novel cytoskeletal-associated protein is essential for the restricted localization of the neuregulin-like factor, vein, at the muscle-tendon junction site. *J. Cell Biol.* **143,** 1259–1270 (1998).
16. Gregory, S. L. & Brown, N. H. *kakapo*, a gene required for adhesion between and within cell layers in *Drosophila*, encodes a large cytoskeletal linker protein related to plectin and dystrophin. *J. Cell Biol.* **143,** 1271–1282 (1998).
17. Lee, S., Harris, K. L., Whitington, P. M. & Kolodziej, P. A. *short stop* is allelic to *kakapo,* and encodes rod-like cytoskeletal-associated proteins required for axon extension. *J. Neurosci.* **20,** 1096–1108 (2000).
18. Prokop, A., Uhler, J., Roote, J. & Bate, M. The *kakapo* mutation affects terminal arborization and central dendritic sprouting of *Drosophila* motorneurons. *J. Cell Biol.* **143,** 1283–1294 (1998).
19. Burridge, K. & Connell, L. Talin: a cytoskeletal component concentrated in adhesion plaques and other sites of actin-membrane interaction. *Cell Motil.* **3,** 405–417 (1983).
20. Hodgkin, J. Sex, cell death, and the genome of *C. elegans. Cell* **98,** 277–280 (1999).
21. McNally, F. J. & Vale, R. D. Identification of katanin, an ATPase that severs and disassembles stable microtubules. *Cell* **75,** 419–429 (1993).
22. Srayko, M., Buster, D. W., Bazirgan, O. A., McNally, F. J. & Mains, P. E. MEI-1/MEI-2 katanin-like microtubule severing activity is required for *Caenorhabditis elegans* meiosis. *Genes Dev.* **14,** 1072–1084 (2000).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
24. Barnes, T. M., Kohara, Y., Coulson, A. & Hekimi, S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans. Genetics* **141,** 159–179 (1995).
25. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274,** 563–567 (1996).
26. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28,** 37–40 (2000).
27. Svetlov, V. V. & Cooper, T. G. Review: compilation and characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast* **11,** 1439–1484 (1995).
28. Hodgkin, J. & Herman, R. K. Changing styles in *C. elegans* genetics. *Trends Genet.* **14,** 352–357 (1998).
29. Chen, J. J. *et al.* Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* **51,** 313–324 (1998).
30. Walhout, A. J. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287,** 116–122 (2000).
31. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res.* **7,** 541–550 (1997).
32. Altschul, S. F. *et al.* Gapped BLAST and PSI–BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature.*

Acknowledgements

We would like to thank A. Fire and L. Timmons for providing protocols, feeding vectors and the HT115(DE3) bacterial strain; and R. Durbin, D. St. Johnston and B. Schumacher for critical reading of the manuscript. A.G.F. was supported by a US Army Breast Cancer Research Fellowship, R.S.K. by a Howard Hughes Medical Institute Predoctoral Fellowship, M.M.-C. by an EC-TMR Network Grant, P.Z. by a Wellcome Trust Studentship, M.S. by a Swiss National Science Foundation fellowship and J.A. by a Wellcome Trust Senior Research Fellowship.

Correspondence and requests for materials should be addressed to J.A. (e-mail: jaa@mole.bio.cam.ac.uk).