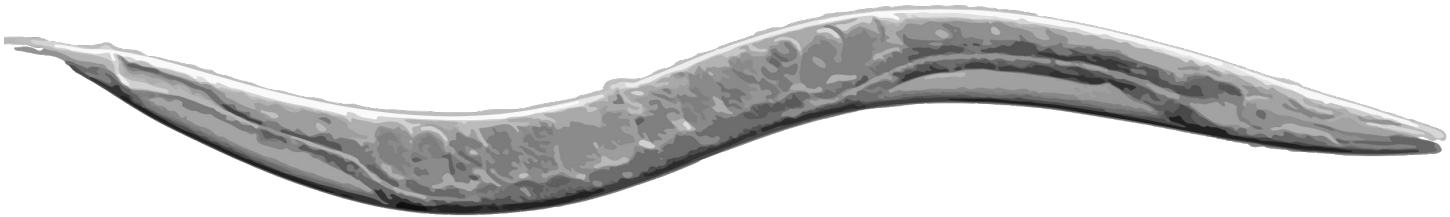
MMG1001 Genomics
Week 3 Tutorial

Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi

TA: Heather Gibling

C. elegans quick facts



- First multicellular organism to have genome sequenced (1998-2002)
- Only organism to have neuronal connections fully described (“connectome”)
- Diploid genome with five autosomes, one sex choromosome
  - 1mb, ~20k genes
  - XX hermaphrodite, X0 male
  - no centromeres, but central clusters of genes with low recombination rates (autosomes)
- Good models for genetic screens

Genetic screens

Experimental technique to **test multiple genes simultaneously** to determine **gene function**, often by looking at phenotypes

In a **screen**, you have to look at (or otherwise examine) a large number of variants – e.g. mutants – in order to find a small number that have a property you are interested in.

Today, we often automate screens (or use reporters, the FACS machine, etc). But, historically, since this is difficult, screens were often staged (and still are, if the assays are challenging or expensive).

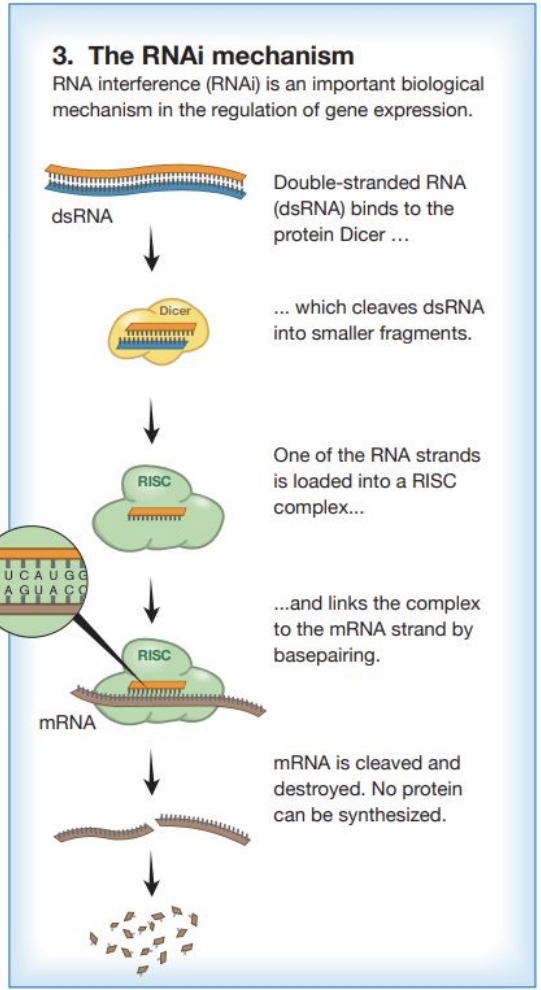
In a **selection**, only a small subset of mutated cells/organisms with a property of interest is able to grow. *Example: antibiotic resistance.*

Tim Hughes

RNA interference (RNAi)

RNA interference (RNAi)

- Inhibition of gene expression or translation by RNA molecules
  - **siRNA**: exogenous RNAs (e.g. viruses)
  - **miRNA**: endogenous RNAs (encoded in genome)
- Used in labs for gene knockdown experiments
- Easy to use in *C. elegans*
  - Engineer *E. coli* with dsRNA for target gene, feed to worms
  - Well-defined libraries available



https://www.nobelprize.org/prizes/medicine/2006/press-release

Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi

Ravi S. Kamath*†, Andrew G. Fraser*†§, Yan Dong*, Gino Poulin*, Richard Durbin‡, Monica Gotta*§, Alexander Kanapin||, Nathalie Le Bot*, Sergio Moreno*¶, Marc Sohrmann‡§, David P. Welchman*, Peder Zipperlen* & Julie Ahringer*

What are the main goals of the paper?

What are the main goals of the paper?

- Perform **genome-wide knockdowns** to determine phenotypic effects of as many *C. elegans* genes as possible
- Determine how genes with related phenotypic effects are **organized** throughout the genome
- Develop and share an **RNAi library** for others to use

How did they do it?

How did they do it?

- Constructed a **library of bacterial strains** (16,757), each capable of expressing **dsRNA** designed to target to a single gene
  - 85% of 19,427 predicted *C. elegans* genes (in 2002)
  - Similar coverage across each chromosome
- Fed worms the bacterial library to knockdown target genes and create **loss of function phenotypes**
- Screened wild-type hermaphrodites to identify genes that result in phenotype classes:
  - **Nonviable**: sterility, embryonic or larval lethality
  - **Growth defects**: slow post-embryonic growth
  - **Viable post-embryonic**: defects in post-embryonic development

What were the major findings?

Figure 1 A

Phenotypes observed across *C. elegans* genome

a		Chr I (2,445 clones)		Chr II (2,978 clones)		Chr III (2,132 clones)		Chr IV (2,693 clones)		Chr V (4,152 clones)		Chr X (2,357 clones)		Total (16,757 clones)	
	Phe	Num	%	Num	%	Num	%	Num	%	Num	%	Num	%	Num	%
	All	334	13.7	354	11.9	395	18.5	293	10.9	215	5.2	131	5.6	1722	10.3
Nonviable	Nonv	253	10.3	240	8.1	298	14.0	187	6.9	145	3.5	47	2.0	1170	7.0
Growth defects	Grow	37	1.5	41	1.4	59	2.8	70	2.6	46	1.1	23	1.0	276	1.6
Viable post-embryonic	Vpep	44	1.8	73	2.5	38	1.8	36	1.3	24	0.6	61	2.6	276	1.6
Embryonic lethal	Emb	226	9.2	204	6.9	220	10.3	143	5.3	108	2.6	28	1.2	929	5.5
Sterile	Ste	83	3.4	43	1.4	132	6.2	58	2.2	44	1.1	12	0.5	372	2.2
Sterile progeny	Stp	15	0.6	33	1.1	18	0.8	14	0.5	19	0.5	3	0.1	102	0.6
Slow post-embryonic growth / Larval arrest	Gro/Lva	147	6.0	129	4.3	161	7.6	131	4.9	102	2.5	34	1.4	704	4.2
Larval lethality	Lvl	38	1.6	61	2.0	35	1.6	18	0.7	20	0.5	24	1.0	196	1.2
Adult lethal	Adl	3	0.1	19	0.6	34	1.6	7	0.3	7	0.2	16	0.7	86	0.5
Blistering of cuticle	Bli	4	0.2	3	0.1	1	0.0	1	0.0	0	0.0	2	0.1	11	0.1
Body morphological defects	Bmd	27	1.1	89	3.0	32	1.5	17	0.6	15	0.4	14	0.6	194	1.2
Clear	Clr	14	0.6	84	2.8	45	2.1	45	1.7	18	0.4	38	1.6	244	1.5
Dumpy	Dpy	19	0.8	39	1.3	16	0.8	20	0.7	7	0.2	16	0.7	117	0.7
Egg-laying defective	Egl	6	0.2	29	1.0	18	0.8	13	0.5	8	0.2	21	0.9	95	0.6
High incidence of males	Him	12	0.5	4	0.1	1	0.0	2	0.1	1	0.0	2	0.1	22	0.1
Long	Lon	2	0.1	11	0.4	9	0.4	8	0.3	3	0.1	5	0.2	38	0.2
Moult defects	Mlt	8	0.3	8	0.3	4	0.2	2	0.1	2	0.0	13	0.6	37	0.2
Multivulva	Muv	2	0.1	4	0.1	5	0.2	2	0.1	0	0.0	1	0.0	14	0.1
Paralyzed	Prz	18	0.7	41	1.4	15	0.7	11	0.4	9	0.2	25	1.1	119	0.7
Protruding vulva	Pvl	32	1.3	39	1.3	37	1.7	17	0.6	13	0.3	9	0.4	147	0.9
Roller	Rol	2	0.1	5	0.2	1	0.0	2	0.1	1	0.0	2	0.1	13	0.1
Ruptured	Rup	10	0.4	39	1.3	25	1.2	18	0.7	11	0.3	16	0.7	119	0.7
Sick	Sck	6	0.2	32	1.1	67	3.1	44	1.6	25	0.6	6	0.3	180	1.1
Uncoordinated	Unc	72	2.9	111	3.7	55	2.6	46	1.7	40	1.0	64	2.7	388	2.3

1,722 genes had observable phenotypes

~10.7% of targeted genes

Many genes associated with multiple phenotypes

Figure 1 B

Chromosome X has predominantly viable phenotypes

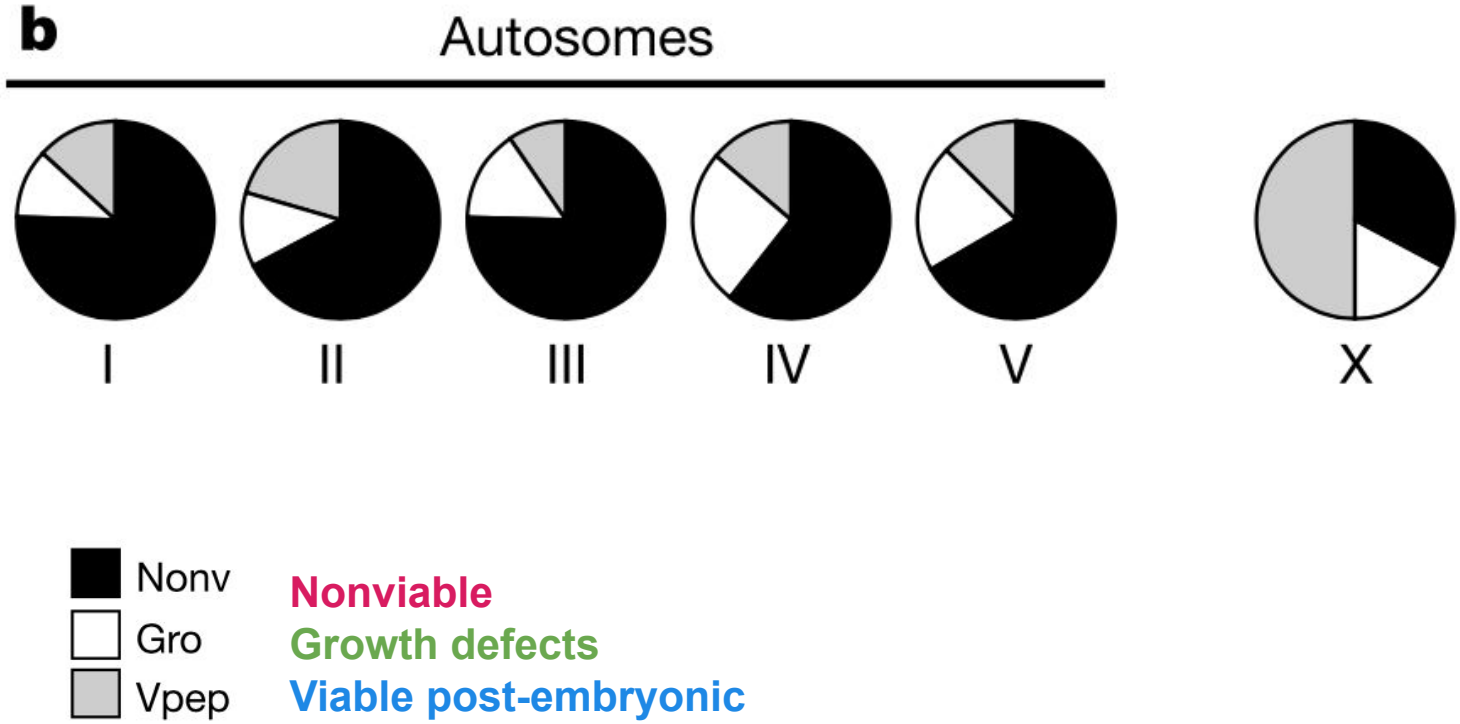
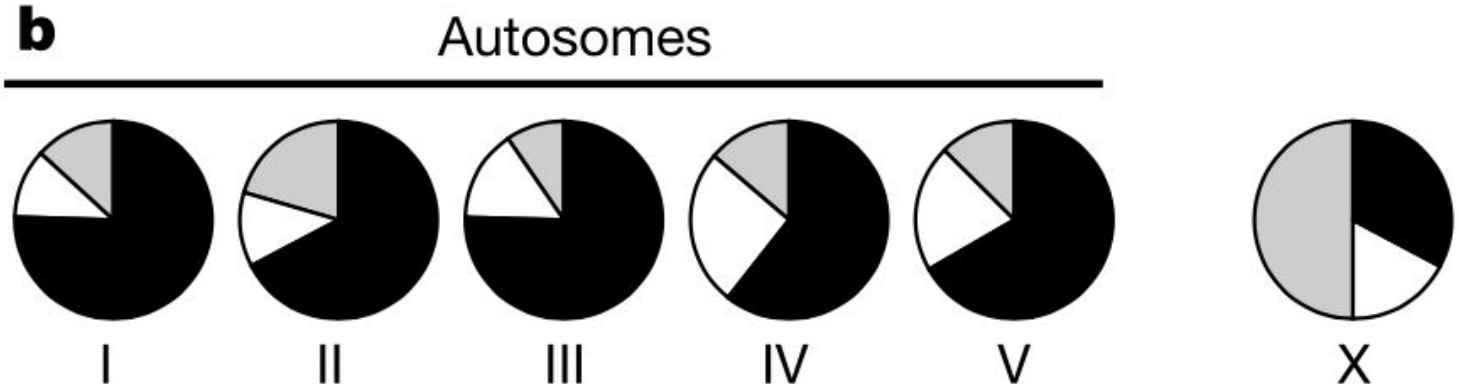


Figure 1 B

Chromosome X has predominantly viable phenotypes



Nonv Nonviable
Gro Growth defects
Vpep Viable post-embryonic

Autosomes had similar distributions of phenotypes, with mostly nonviable phenotypes.

What does this say about selection on chrom X?

RNAi phenotypes for 33 human homologs/diseases

Table 1 **Thirty-three human disease gene homologues with an RNAi phenotype**

Predicted gene	*C. elegans* locus	Human disease	Human gene	BlastP *E* value	RNAi phenotype
B0035.5		G6PD deficiency	*G6PD*	1×10^{-176}	Emb, Clr, Gro
B0350.2A	*unc-44*	Hereditary spherocytosis	*ANK1*	0.00	Slu
C01G6.8	*cam-1/kin-8*	Insulin-resistant diabetes mellitus	*INSR*	6×10^{-55}	Unc, Pvl, clear patch
C01G8.5A		Neurofibromatosis	*NF2*	1×10^{-123}	Unc, Lvl, Gro
C06A1.1		Zellweger syndrome	*PEX1*	3×10^{-67}	Emb, Bmd, Sck, Gro
C07H6.7	*lin-39*	MODY, type IV	*IPF1*	5×10^{-14}	Egl, Vul, Muv
C17E4.5		Oculopharyngeal muscular dystrophy	*PABPN1*	3×10^{-41}	Emb, Unc, Lva
C29A12.3	*lig-1*	DNA ligase I deficiency	*DNA ligase1*	1×10^{-167}	Emb
C48A7.1	*egl-19*	Long QT syndrome 3	*SCN5A*	2×10^{-64}	Egl, Clr
C50H2.1		Leydig cell hypoplasia	*LHCGR*	9×10^{-76}	Gro
D2045.1		Spinocerebellar ataxia 2	*SCA2*	7×10^{-09}	Emb
F01G10.1		Wernicke–Korsakoff syndrome	*TKT*	0.00	Emb, Clr, Gro
F07A5.7	*unc-15*	Tuberous sclerosis	*TSC1*	1×10^{-07}	Unc, Prz, Egl
F11C1.6	*nhr-25*	Pseudohyperaldosteronism	*NR3C2*	7×10^{-24}	Unc, Prz, Clr, Egl
F11H8.4	*cyk-1*	Nonsyndromic sensorineural deafness	*DFNA1*	9×10^{-49}	Emb, Adl, Rup, Clr
F20B6.2	*vha-12*	Renal tubular acidosis	*ATP6B1*	0.00	Emb, Ste, Adl, Lvl, Prz
F54D8.1		Ehlers–Danlos syndrome, type IV	*COL3A1*	1×10^{-06}	Dpy
F53G12.3		Chronic Granulomatous Disease	*X-CGD*	3×10^{-34}	Bli, Mlt, Lvl
F58A3.2A	*egl-15*	Multiple venous malformations	*VMCM*	1×10^{-62}	Egl
K04G2.8A	*apr-1*	Adenomatous polyposis of the colon	*APC*	9×10^{-34}	Unc, Bmd, Lvl
K07A1.12	*rba-2*	Cockayne syndrome	*CKN1*	6×10^{-13}	Emb, Pvl, Lvl
K08A8.2		Gonadal dysgenesis	*SRY*	3×10^{-31}	Unc, Egl
K08C7.3	*epi-1*	Usher syndrome 2a	*USH2A*	1×10^{-112}	Ste, Unc, Muv, Dpy, Pvl, Rup
K11D9.2A		Darier–White disease	*SERCA*	0.00	Ste, Sck
M02A10.2		Hyperinsulinism	*KCNJ11*	4×10^{-78}	Unc
R107.8	*lin-12*	Alagille syndrome	*JAG1*	2×10^{-90}	Egl
R12B2.1	*sma-4*	Pancreatic carcinoma	*MADH4*	2×10^{-39}	Sma, Dpy
T03F6.5	*lis-1*	Miller–Dieker lissencephaly syndrome	*PAF*	1×10^{-148}	Emb
W05E10.3	*ceh-32*	Holoprosencephaly	*SIX3*	1×10^{-69}	Unc
W10G6.3	*ifa-2*	Keratoderma	*KRT9*	7×10^{-26}	Unc, Lvl, Mlt
Y47D3A.6A	*tra-1*	Grieg cephalopolysyndactyly syndrome	*GLI*	6×10^{-58}	Rup, clear patch
Y76A2A.2		Menkes disease	*ATP7A*	0.00	Prz, Adl, Unc
ZC506.4	*mgl-1*	Hypercalcemia	*CASR*	2×10^{-77}	Gro

C. elegans genes with a human disease gene homologue are defined as those with a BlastP *E* value less than 1.0×10^{-6}, taken from refs 38, 39. Shown are those with an RNAi phenotype. The phenotypes are defined in Methods. MODY, maturity onset diabetes of the young. G6PD, glucose-6-phosphate dehydrogenase.

Table 1

RNAi phenotypes for 33 human homologs/diseases

Table 1 **Thirty-three human disease gene homologues with an RNAi phenotype**

Predicted gene	*C. elegans* locus	Human disease	Human gene	BlastP *E* value	RNAi phenotype
B0035.5		G6PD deficiency	*G6PD*	1×10^{-176}	Emb, Clr, Gro
B0350.2A	*unc-44*	Hereditary spherocytosis	*ANK1*	0.00	Slu
C01G6.8	*cam-1/kin-8*	Insulin-resistant diabetes mellitus	*INSR*	6×10^{-55}	Unc, Pvl, clear patch
C01G8.5A		Neurofibromatosis	*NF2*	1×10^{-123}	Unc, Lvl, Gro
C06A1.1		Zellweger syndrome	*PEX1*	3×10^{-67}	Emb, Bmd, Sck, Gro
C07H6.7	*lin-39*	MODY, type IV	*IPF1*	5×10^{-14}	Egl, Vul, Muv
C17E4.5		Oculopharyngeal muscular dystrophy	*PABPN1*	3×10^{-41}	Emb, Unc, Lva
C29A12.3	*lig-1*	DNA ligase I deficiency	*DNA ligase1*	1×10^{-167}	Emb
C48A7.1	*egl-19*	Long QT syndrome 3	*SCN5A*	2×10^{-64}	Egl, Clr
C50H2.1		Leydig cell hypoplasia	*LHCGR*	9×10^{-76}	Gro
D2045.1		Spinocerebellar ataxia 2	*SCA2*	7×10^{-09}	Emb
F01G10.1		Wernicke–Korsakoff syndrome	*TKT*	0.00	Emb, Clr, Gro
F07A5.7	*unc-15*	Tuberous sclerosis	*TSC1*	1×10^{-07}	Unc, Prz, Egl
F11C1.6	*nhr-25*	Pseudohyperaldosteronism	*NR3C2*	7×10^{-24}	Unc, Prz, Clr, Egl
F11H8.4	*cyk-1*	Nonsyndromic sensorineural deafness	*DFNA1*	9×10^{-49}	Emb, Adl, Rup, Clr
F20B6.2	*vha-12*	Renal tubular acidosis	*ATP6B1*	0.00	Emb, Ste, Adl, Lvl, Prz
F54D8.1		Ehlers–Danlos syndrome, type IV	*COL3A1*	1×10^{-06}	Dpy
F53G12.3		Chronic Granulomatous Disease	*X-CGD*	3×10^{-34}	Bli, Mlt, Lvl
F58A3.2A	*egl-15*	Multiple venous malformations	*VMCM*	1×10^{-62}	Egl
K04G2.8A	*apr-1*	Adenomatous polyposis of the colon	*APC*	9×10^{-34}	Unc, Bmd, Lvl
K07A1.12	*rba-2*	Cockayne syndrome	*CKN1*	6×10^{-13}	Emb, Pvl, Lvl
K08A8.2		Gonadal dysgenesis	*SRY*	3×10^{-31}	Unc, Egl
K08C7.3	*epi-1*	Usher syndrome 2a	*USH2A*	1×10^{-112}	Ste, Unc, Muv, Dpy, Pvl, Rup
K11D9.2A		Darier–White disease	*SERCA*	0.00	Ste, Sck
M02A10.2		Hyperinsulinism	*KCNJ11*	4×10^{-78}	Unc
R107.8	*lin-12*	Alagille syndrome	*JAG1*	2×10^{-90}	Egl
R12B2.1	*sma-4*	Pancreatic carcinoma	*MADH4*	2×10^{-39}	Sma, Dpy
T03F6.5	*lis-1*	Miller–Dieker lissencephaly syndrome	*PAF*	1×10^{-148}	Emb
W05E10.3	*ceh-32*	Holoprosencephaly	*SIX3*	1×10^{-69}	Unc
W10G6.3	*ifa-2*	Keratoderma	*KRT9*	7×10^{-26}	Unc, Lvl, Mlt
Y47D3A.6A	*tra-1*	Grieg cephalopolysyndactyly syndrome	*GLI*	6×10^{-58}	Rup, clear patch
Y76A2A.2		Menkes disease	*ATP7A*	0.00	Prz, Adl, Unc
ZC506.4	*mgl-1*	Hypercalcemia	*CASR*	2×10^{-77}	Gro

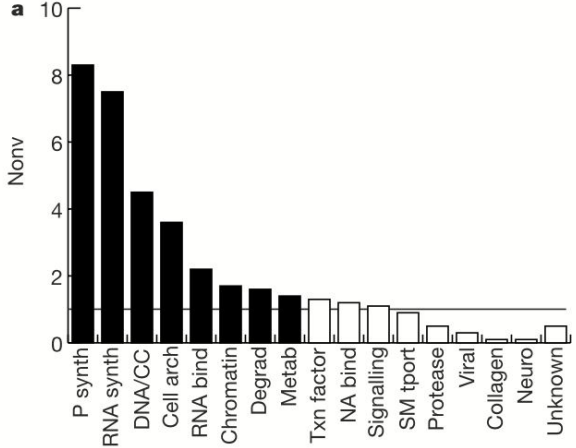
C. elegans genes with a human disease gene homologue are defined as those with a BlastP *E* value less than 1.0×10^{-6}, taken from refs 38, 39. Shown are those with an RNAi phenotype. The phenotypes are defined in Methods. MODY, maturity onset diabetes of the young. G6PD, glucose-6-phosphate dehydrogenase.

Many of these (50%) had viable post-embryonic phenotypes

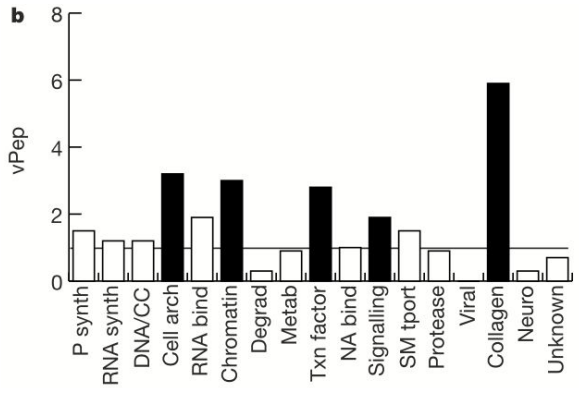
Why might it be useful to identify C. elegans homologs for disease-causing genes in humans?

Phenotype classes have different over representations in functional classes of genes

Nonviable



Viable post-embryonic



Chromosome X

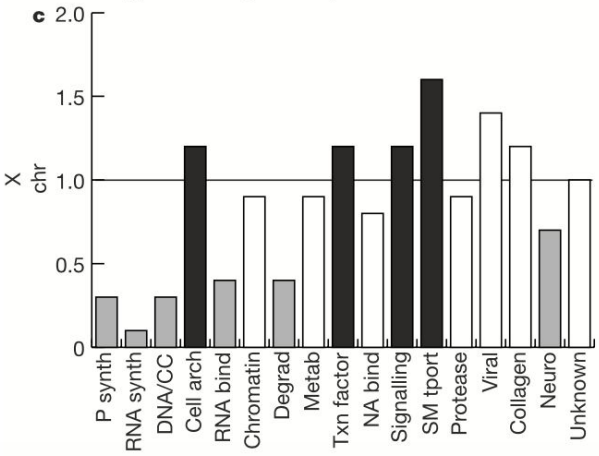
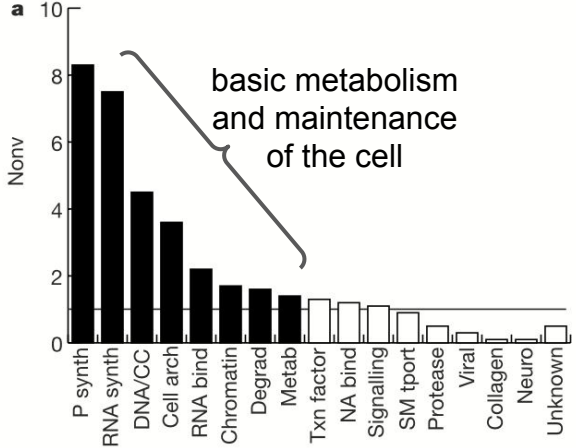


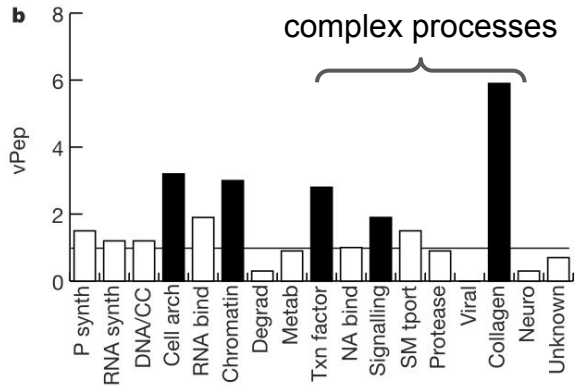
Figure 2

Phenotype classes have different over representations in functional classes of genes

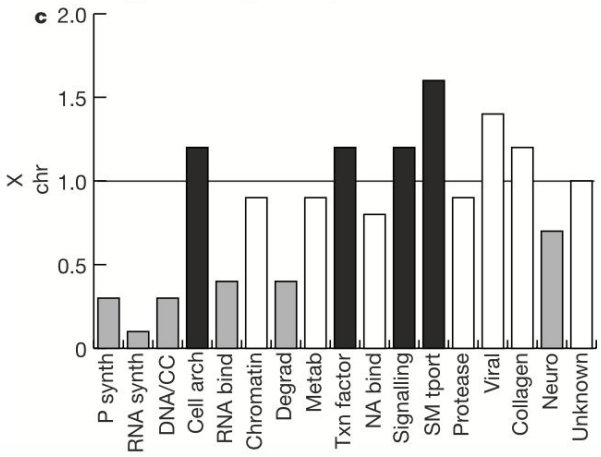
Nonviable



Viable post-embryonic



Chromosome X



Gene classes involved with **basic metabolism and maintenance of the cell** are enriched for nonviable phenotypes; more **complex processes** classes enriched for viable post-embryonic phenotypes.

What is observed for chromosome X?

Several abundant protein domains have significant associations with particular RNAi phenotypes

Table 2 **InterPro domains associated with RNAi phenotypes**

Nonv only	
	Elongation factor, GTP-binding
	Cyclin
	Ubiquitin domain
	TPR repeat
	Zinc-finger, CCHC type
	Myb DNA-binding domain
	Laminin-type EGF-like domain
	DEAD/DEAH box helicase
	Ubiquitin-associated domain
	Zinc-finger, C_2H_2 type
	Mitochondrial substrate carrier
	Protein kinase C, phorbol ester/DAG binding
Gro only	
	Glycosyl transferase, family 2
	Zinc-finger, RING
	Phosphotyrosine interaction domain
	Proline-rich extensin
Nonv and Gro	
	G-protein β-subunit WD40 repeat
	AAA ATPase
	KH domain
	Zinc-finger, C-X_8-C-X_5-C-X_3-H type
	RNA-binding region RNP-1 (RNA recognition)
Vpep	
	Immunoglobulin/major histocompatibility complex
	Collagen triple helix repeat
	Immunoglobulin-like
	EGF-like calcium-binding
	Aspartic acid and asparagine hydroxylation site
	Fibronectin, type III
	Worm-specific repeat type 1

Nonviable
Growth defects
Viable post-embryonic

Most viable post-embryonic associated domains are found in flies and humans, but not budding yeast or *Arabidopsis* (plants).

What does this suggest about animal-specific domains?

Figure 3

Genes with a nonviable phenotype are enriched for being in the ancient gene class

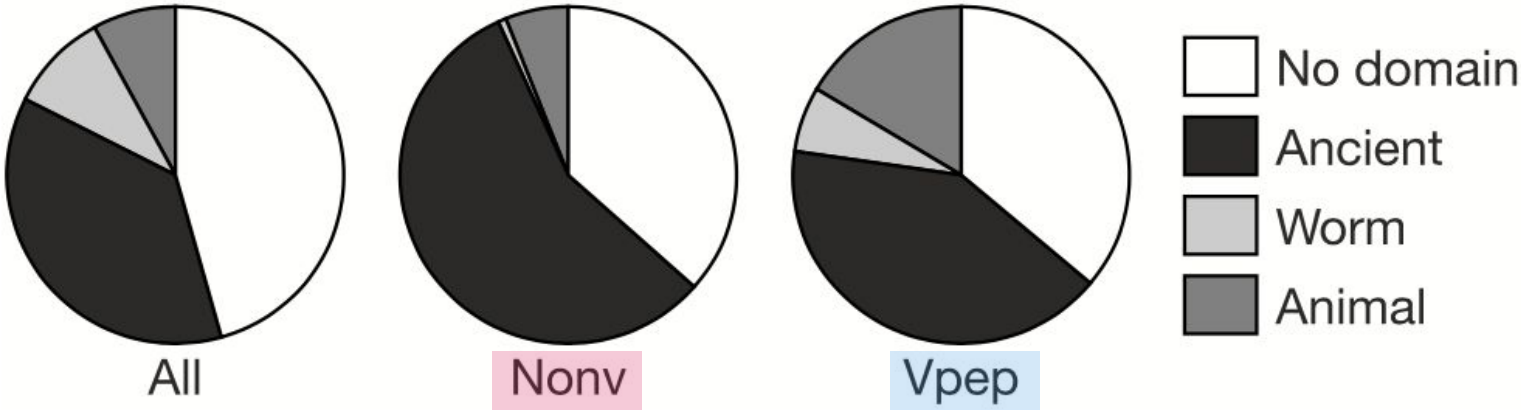
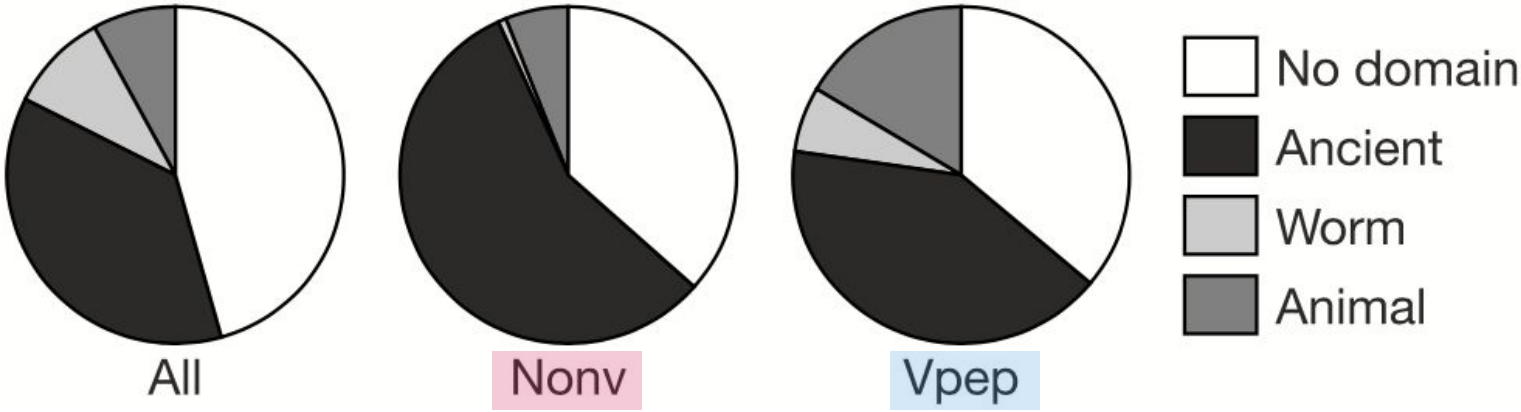


Figure 3

Genes with a nonviable phenotype are enriched for being in the ancient gene class

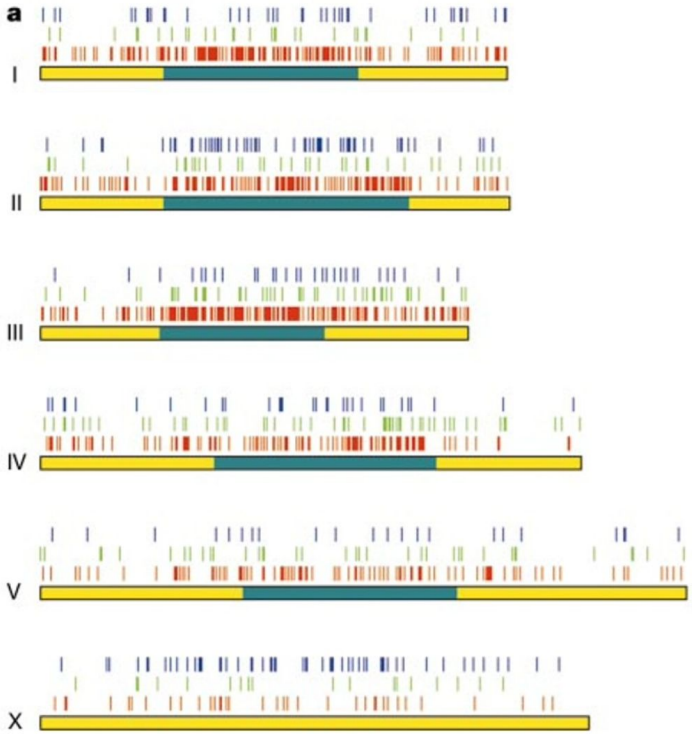


Almost none of the worm class genes have a nonviable phenotype.

What does this suggest about these worm-specific genes?

Figure 4 A

RNAi phenotypes are enriched in central chromosome cluster regions

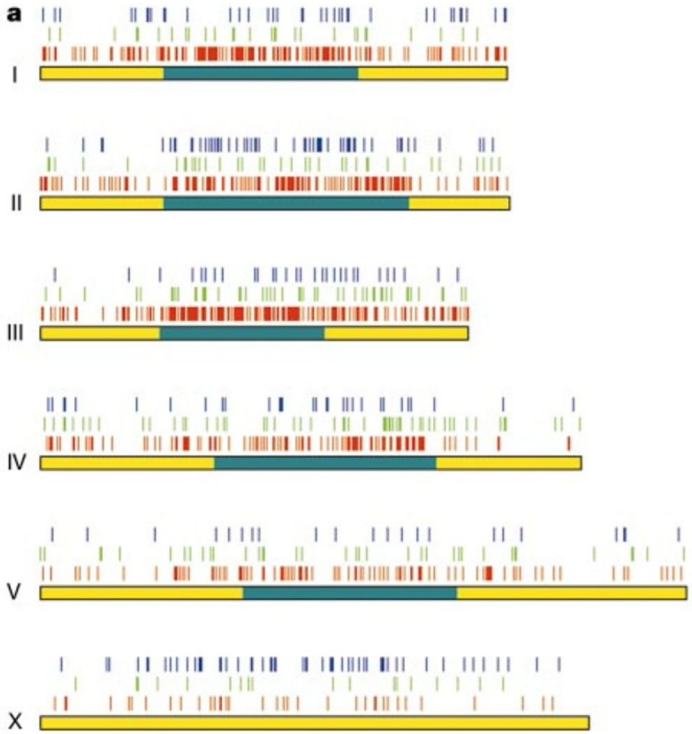


Viable post-embryonic
Growth defects
Nonviable

Chromosome arm regions
Chromosome cluster regions

Figure 4 A

RNAi phenotypes are enriched in central chromosome cluster regions



Of genes with phenotypes:

- 14.9% in clusters
- 7.6% in arms

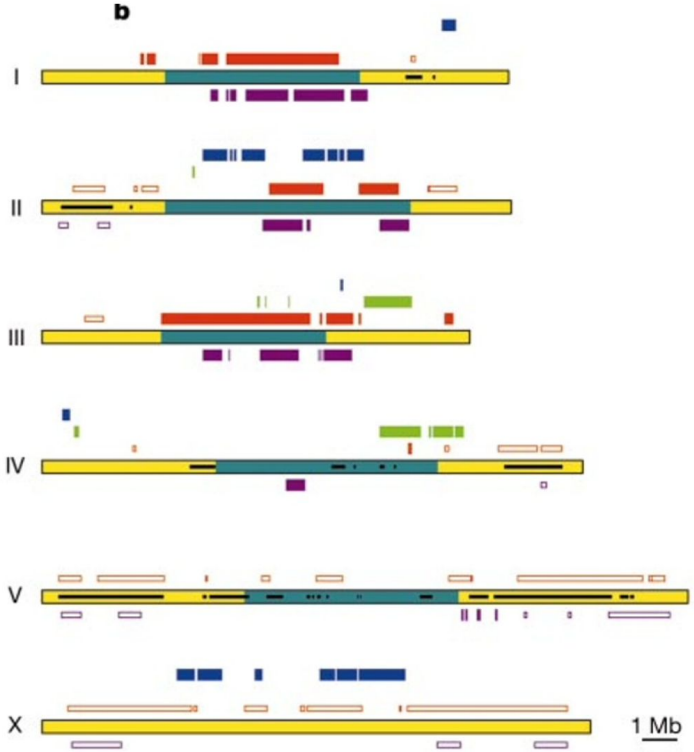
Where does recombination tend to occur?
Which areas are more conserved?

Viable post-embryonic
Growth defects
Nonviable

Chromosome arm regions
Chromosome cluster regions

Figure 4 B

Nonviable genes over-enriched in three chromosome clusters, under-enriched in chromosome X



Over-enrichment
Under-enrichment

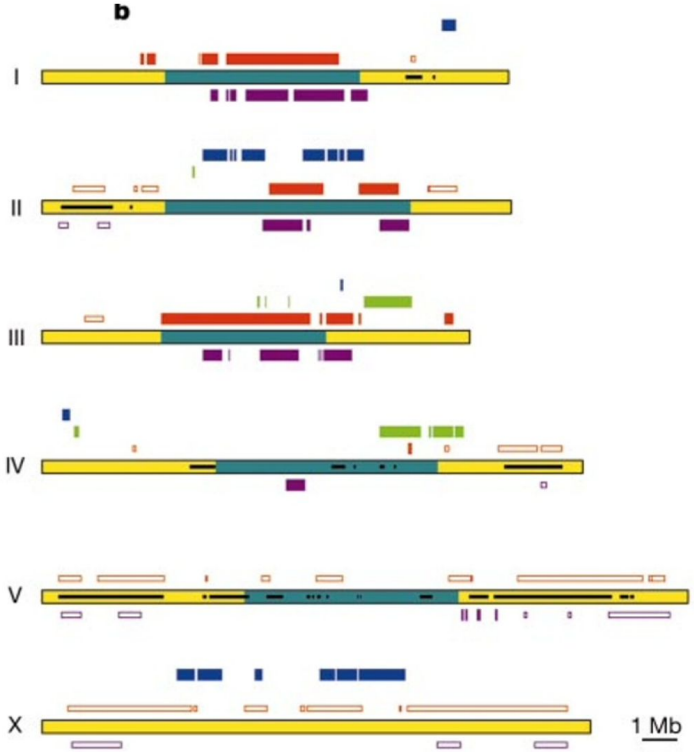
Viable post-embryonic
Growth defects
Nonviable

Sig. over/under-enriched in mount (microarray set) 11

Enriched for duplicated genes

Chromosome arm regions
Chromosome cluster regions

Nonviable genes over-enriched in three chromosome clusters, under-enriched in chromosome X



36% of genes with nonviable phenotype enriched strongly enriched in clusters of chromosomes I, II, and III (13% of genome)

What might explain the under-enrichment of phenotypic genes in the chromosome arms?

Over-enrichment
Under-enrichment

Viable post-embryonic
Growth defects
Nonviable

Sig. over/under-enriched in mount (microarray set) 11
Enriched for duplicated genes
Chromosome arm regions
Chromosome cluster regions

Other fun findings

- *C. elegans* genes with an **ortholog** in another eukaryote are **more likely to have a detectable RNAi phenotype** than all other genes (21% versus 6%)
- Highly conserved genes present as **single copy** in the *C. elegans* genome are **more than twice as likely to have phenotype** as those present in more than one copy (31% versus 12%)
    - Suggests that many **recently duplicated paralogs are at least partially functionally redundant** or have specialized functions that are not detectable in this screen
- **Highest cross-species conservation** seen among genes with a **nonviable** phenotype (52% have an ortholog in another eukaryote)
    - Similar essential basal cellular machinery common to all eukaryotes

Why does it matter?

Why does it matter?

- First systematic functional analysis of a metazoan genome
- Over two-thirds of genes annotated with an RNAi phenotype in this study had previously not been associated with a biological function *in vivo*
- Created genome-wide RNAi library for others to use
  - Ahringer library for *C. elegans*
  - influenced library development for other species

Assignment: Using RNAi screens to predict phenotypes of orthologous genes

Assignment: Using RNAi screens to predict phenotypes of orthologous genes

Homologs: Genes that are descended from a common ancestor. (vs. *Analog*, which has a similar function, but different origin)

Orthologs: Derived from a single ancestral gene, arising due to speciation.

Paralogs: Homologous sequences that are separated by gene duplication.