

# Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*

B. Sönnichsen<sup>1</sup>, L. B. Koski<sup>1\*</sup>, A. Walsh<sup>1</sup>, P. Marschall<sup>1\*</sup>, B. Neumann<sup>1\*</sup>, M. Brehm<sup>1</sup>, A.-M. Alleaume<sup>1\*</sup>, J. Artelt<sup>1\*</sup>, P. Bettencourt<sup>1\*</sup>, E. Cassin<sup>2\*</sup>, M. Hewitson<sup>1</sup>, C. Holz<sup>1</sup>, M. Khan<sup>1</sup>, S. Lazik<sup>1</sup>, C. Martin<sup>1</sup>, B. Nitzsche<sup>1\*</sup>, M. Ruer<sup>2</sup>, J. Stamford<sup>2</sup>, M. Winzi<sup>1</sup>, R. Heinkel<sup>1\*</sup>, M. Röder<sup>1\*</sup>, J. Finell<sup>1\*</sup>, H. Häntschi<sup>1</sup>, S. J. M. Jones<sup>3</sup>, M. Jones<sup>4\*</sup>, F. Piano<sup>5</sup>, K. C. Gunsalus<sup>5</sup>, K. Oegema<sup>2\*</sup>, P. Gönczy<sup>2\*</sup>, A. Coulson<sup>4\*</sup>, A. A. Hyman<sup>2</sup> & C. J. Echeverri<sup>1</sup>

<sup>1</sup>Cenix BioScience GmbH, Tatzberg 47-51, D-01307 Dresden, Germany

<sup>2</sup>Max-Planck-Institute for Cell Biology and Genetics (MPI-CBG), Pfotenhauerstrasse 108, D-01307 Dresden, Germany

<sup>3</sup>Genome Sciences Centre, British Columbia Cancer Research Centre, 570 West 7th Avenue, Vancouver, British Columbia V5Z 4E6, Canada

<sup>4</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

<sup>5</sup>Department of Biology, Centre for Comparative Functional Genomics, New York University, 1009 Silver Centre, 100 Washington Square East, New York, New York 10003, USA

\* Present addresses: Centre Robert Cedergren, Centre de Recherche en Bioinformatique et en Sciences Génomiques de l'Université de Montréal, 2900 Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada (L.B.K.); Max Planck Institute of Infection Biology, Schumannstrasse 21/22, 10117 Berlin, Germany (P.M.); European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, D-69117 Heidelberg; Germany (B. Neumann); Institut Cochin, Département de Biologie Cellulaire, 22 rue Méchain, 75014 Paris, France (A.-M.A.); Institut für Klinische Genetik, Medizinische Fakultät Carl Gustav Carus, TU Dresden, Fetscherstrasse 74, 01307 Dresden, Germany (J.A.); Instituto Gulbenkian de Ciência, Rua da Quinta Grande n 6, 2780-901 Oeiras, Portugal (P.B.); Cavaleri Ottolenghi Scientific Institut (COSI) of Neurobiology, Azienda Ospedale San Luigi Gonzaga, Regione Gonzole 10-1004 Orbassano, Italy (E.C.); Mühlstrasse 36, 04317 Leipzig, Germany (B. Nitzsche); Brombeerweg 2, 76275 Ertlingen, Germany (R.H.); Ingeniweb, 2 cours du 14 Juillet, 78300 Poissy, France (M.R.); Rätälänkatu 20 A 11, 20810 Turku, Finland (J.F.); Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada (M.J.); University of California, San Diego, CMM-East 3080, 9500 Gilman Drive, La Jolla, California 92093, USA (K.O.); Swiss Institute for Experimental Cancer Research (ISREC), CH-1066 Lausanne, Switzerland (P.G.); MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK (A.C.)

**A key challenge of functional genomics today is to generate well-annotated data sets that can be interpreted across different platforms and technologies. Large-scale functional genomics data often fail to connect to standard experimental approaches of gene characterization in individual laboratories. Furthermore, a lack of universal annotation standards for phenotypic data sets makes it difficult to compare different screening approaches. Here we address this problem in a screen designed to identify all genes required for the first two rounds of cell division in the *Caenorhabditis elegans* embryo. We used RNA-mediated interference to target 98% of all genes predicted in the *C. elegans* genome in combination with differential interference contrast time-lapse microscopy. Through systematic annotation of the resulting movies, we developed a phenotypic profiling system, which shows high correlation with cellular processes and biochemical pathways, thus enabling us to predict new functions for previously uncharacterized genes.**

The concept of 'systems biology' promotes a broad approach to understanding complex biological processes by examining the interplay between molecular assemblies and networks, rather than focusing on individual molecules. Nevertheless, to make this approach possible it is necessary to know which molecules, among the thousands of predicted gene products, actually make up these assemblies and how they might contribute to functionality within the networks. Although achieving this goal is now facilitated by the advent of functional genomics, a key challenge of these approaches, beyond the efficient exploitation of sequence data, is to bridge the gap between large-scale analyses and more in-depth studies by individual laboratories.

The discovery of RNA-mediated interference (RNAi) as a method for targeted gene silencing introduces the best functional genomics tool available for metazoan organisms so far, providing a direct causal link between genes and cellular functions. The basic method harnesses an endogenous gene regulation pathway by which exogenous double-stranded RNA molecules (dsRNAs), after their introduction into cells, are processed to generate a pool of short interfering RNAs (siRNAs) that in turn direct the catalytic degradation of complementary 'target' mRNAs. We and others have previously established the systematic application of RNAi as a genome-scale screening method of identifying and characterizing genes with functions of interest<sup>1-6</sup>. For monitoring individual pathways, such screens can be run with readouts of relatively 'low content' such as single reporter assays, thus enabling a higher throughput. In comparison, higher-content assays, ideally integrating

both spatial and temporal information, enable more comprehensive analyses. However, the challenge arises to provide systematically annotated information that allows comparison to other screening methods, such as two-hybrid data and expression profiling<sup>7,8</sup>.

The primary goal of the present study was to identify all genes essential for mitotic cell division in a metazoan organism. For a mother cell to divide into two daughters correctly, several coordinated events must take place. These include duplication of the centrosome starting in late G1, replication of chromosomal DNA at S phase, assembly of a mitotic spindle at M phase, segregation of the replicated genome to opposite poles of the mitotic spindle at anaphase, and division of the mother cell into two daughter cells at cytokinesis. Previous work has established the early *C. elegans* embryo as an excellent *in vivo* system for detailed studies of metazoan cell division (reviewed in ref. 9). In particular this system allows time-lapse cytological analyses of the first embryonic cell divisions with excellent spatial and temporal resolution<sup>10</sup>. The early *C. elegans* embryo does not arrest in response to cell cycle checkpoint mechanisms<sup>11</sup>, which normally limit the analysis of disruption phenotypes in most other cellular systems. Furthermore, *C. elegans* is amenable to RNAi experimentation<sup>12</sup> and also offers a fully sequenced genome<sup>13</sup>, thereby making it uniquely suited for genome-scale RNAi screening.

## Genome-wide screening using a high-content assay

Our aim was to test all *C. elegans* genes for their role in the first two rounds of mitotic cell division after fertilization. To this end, we

Table 1 **Phenotypic and functional classification of early embryonic (DIC) phenotypes**

Phenotype class	Phenotype description	Total (unknown function)	Biochemical pathway	Functional class	Examples of genes
Sterility/impaired fertility in F0	Injected worm produces no/few embryos	42 (4)	Secretion	Production of oocytes	F12F6.6(sec-24.1) C39F7.4(rab-1) F26D10.3(hsp-1) F32D8.6(emo-1) T24B8.1(rpl-32) Y113G7A.3(sec-23) ZC518.2(sec-24.2) F43E2.8(hsp-4) W01A11.2 B0336.2(arf-1) C01F6.3(ccp-31A1)
Osmotic integrity	Embryo fills egg shell, lyses upon dissection or during recording	109 (15)	General cell maintenance	Membrane stability, egg shell formation	C07G2.3a(cct-5) D2045.1(atx-2) F53G12.1(rab-11.1) F57B10.10(dad-1),T25G3.2(chi-1) T20H4.3(prs-1) T23D8.4(eif-3.C)Y47G6A.12(sep-1) Y54E10BL.6(mek-2) Y56A3A.20(ccf-1) Y62E10A.15 Y71F9AL.13a(rpl-1) Y71F9B.4(snr-7) Y76A2B.1(pod-1) Y77E11A.13a(npp-20) Y116A8C.42(snr-1)ZK675.1(ptc-1)
Polar body extrusion	Unextruded or resorbed polar body(ies) leading to extra PNs in P <sub>0</sub> and/or extra karyomeres in AB/P <sub>1</sub>	12 (2)	Defective meiosis, kinesin-like proteins	Meiosis chromosome segregation	C06G3.2(klp-18) F52H2.7 F57B10.12(mei-2) M01E11.6(klp-15) Y54E10A.9a(vbh-1) C41G7.2(klp-16) R06A4.4a(imb-2)
Passage through meiosis	Male and female PNs not visible; embryo often fills egg shell completely	47 (4)	Cell cycle progression, protein destruction	Meiotic cell cycle progression	ZK945.2(pas-7) F25B5.4a(ubq-1) C47B2.4(pbs-2) C02F5.9(pbs-6) C30C11.2(rpn-3) C36B1.4(pas-4) F35G12.9(apc-11) K06A5.7(cdc-25.1) T05G5.3(cdk-1) ZK177.6(fzy-1)
Entry into interphase	Embryos spend longer than normal when entering first interphase	4 (0)	Cell cycle progression	Meiotic cell cycle progression	C09G4.3(dom-6) F38H4.9 F48E8.5 ZK520.4a(cul-2)
Cortical dynamics	Little/no cortical ruffling or pseudocleavage furrow, or excessive cortical activity at two-cell stage	19 (6)	Rho, transcription	Cortical structure	Y51H4A.3(rho-1) Y71F9AL.16(arx-1)
Pronuclear/nuclear appearance	PN and nuclei are small or missing altogether, often accompanied by spindle defects	28 (4)	Nuclear pore complex, import machinery	Reformation of nuclei after meiosis/mitosis	T01G9.4(npp-2) C26D10.1(ran-3) F28B3.8(imb-1) W04D2.1a(atn-1) F26B1.3(ima-2)
Centrosome attachment	Centrosomes detach from the male PN	3 (2)	–	Centrosome attachment	F23F12.2 F57B1.2 ZK546.1
Pronuclear migration	Lack of male pronuclear migration, female pronuclear migration variable, sometimes multiple female PNs, no or small spindle	38 (3)	Dynein–dynactin, microtubule	Microtubule function	F22B5.7(zyg-9) C47B2.3a(tba-2) F13D12.7(gpb-1) C36E8.5(tbb-2) K08E7.3(et-99) R151.9 T03F6.5(lis-1) T16G12.1 T21E12.4(dhc-3) Y54E2A.3(tac-1) Y108G3AL.1(cul-3) F49E11.1a(mbk-2) F56A3.4(spd-5)
Spindle assembly	Spindle bipolarity is not established	21 (6)	Aurora A, centrioles	Microtubule function	K07C11.2(air-1) F59E12.2(zyg-1) F10E9.8(sas-4)
Spindle elongation/integrity	Bipolar spindle shows clear elongation defect				F36A4.7(ama-1) F44A2.3 K11D9.1a(klp-7)
Sister chromatid separation (cross-eyed)	Daughter nuclei are deformed and stay close to central cortex, cytokinesis defects	64 (7)	Cohesion, kinetochore replication, histone	Chromosome function	C25D7.6(mcm-3) F18E2.3(scc-3) W03D2.4(pcn-1) B0035.8(his-48) B0207.4(air-2) C02F5.1(knl-1) C43E11.10(cdc-6) K12D12.2(npp-3) T06E6.2a(cyb-3) T27F2.3(bir-1) Y54E10A.15a(cdt-1) R06F6.1(cdl-1)
Nuclear appearance	PNs are normal but nuclei are completely missing or significantly smaller than normal; often accompanied by spindle and cytokinesis defects	5 (0)	Replication, histone	Chromosome function	F10C2.4 F55G1.11(his-60)Y39G10AR.14(mcm-4)
Chromosome segregation (karyomeres)	Karyomeres in AB and/or P <sub>1</sub> , often accompanied by weak/thin wobbly spindle	23 (2)	Cyclin, histone	Chromosome segregation	Y43F4B.6(klp-19) ZC168.4(cyb-1) K06A5.4 T23H2.1(npp-12)
Cytokinesis	Cytokinesis defects in the first and/or second stages of cell division	15 (1)	Actin/myosin	Cytokinesis	M03F4.2a(act-4) F11H8.4a(cyk-1) C56G7.1(mlc-4) K08E3.6(cyk-4) M03D4.1a(zen-4) Y18D10A.20(pfn-1)
Asymmetry of division	Symmetric (PAR-like) divisions or excessive posterior displacement (zyg-8 like phenotypes)	12 (1)	PAR	Polarity	W08F4.8 F09E5.1(pkc-3) M117.2(par-5) C38C10.4(gpr-2) F20G4.3(nmy-2) F22B7.13(gpr-1) F58B6.3a(par-2) T25C8.2(act-5) T26E3.3(par-6)
Pace of P-lineage	More than 5 min between AB and P <sub>1</sub> divisions	14 (3)	DNA replication	DNA damage checkpoint	F31E3.3(rfc-4) F33H2.5 R01H10.1(div-1) Y41C4A.10(elib-1) ZK637.8a(unc-32) ZK742.1a(imb-4)

Table 1 – continued

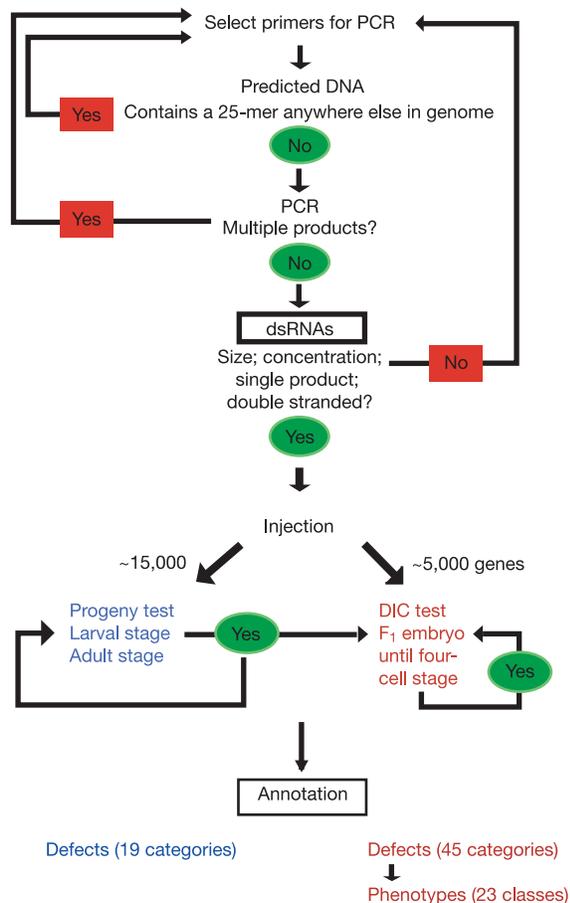
Phenotype class	Phenotype description	Total (unknown function)	Biochemical pathway	Functional class	Examples of genes
General pace of development	More than 30 min from PN meeting to furrow initiation in AB	49 (11)	ATP generation	Mitochondrial function	Y69A2AR.18a F58F12.1 W10D9.5 T06D8.8(rpn-9) R53.4 T27F7.1 F44G4.2 R06F6.2 F01F1.12a F23B12.5 R11A8.6(irs-1) C37H5.8(hsp-6) T23D8.1(mom-5) C06H2.1 C15H9.10 C34E10.6(atp-2) C53A5.1 C56C10.8 D2030.4 D2085.3 E04A4.5 F23H11.5 F29B9.11 F35G12.2 F35G12.10(asb-1) F36A2.7 F42A6.6 F54H12.1a(aco-2) F56D2.1 F59C6.5 H06H21.3 H28O16.1 K07A12.5 T05H4.12 T07C4.7(mev-1) T09B4.9 T20G5.2 T20H4.5 T22B11.5 W10D5.2 Y37E3.9 K04G7.4a C25H3.9a T05H10.6a Y82E9BR.3a T27E9.1a T10E9.7a W02D9.1(pri-2) C41G7.3
Severe pleiotropic defects	Often multiple female PN's, aberrant cytoplasmic texture, drop in overall pace of development, osmotic sensitivity	108 (7)	Translation machinery, ribosomal proteins, tRNA synthase, translation initiation factor	Protein synthesis	F28C6.7a(rpl-26)T05C12.7(oct-1) C01H6.5a(nhr-23) ZC434.5(ers-2) Y41E3.4(qrs-5) F22B5.2(eif-3.G) F22B5.9(fr-2) F56E10.4(rps-27) C47E12.1(srs-2) F28H1.3(ars-2) C37H5.6a C49H3.11(rps-2) W02F12.5 B0041.4(rpl-4) B0250.7 B0336.10(rpl-23) B0348.6a(ife-3) B0393.1(rps-0) B0393.8 B0412.4(rps-29) B0412.5 B0464.1(drs-1) B0511.10(eif-3.E) C04F12.4(rpl-14) C09D4.5(rpl-19) C09H10.2 C14B9.7(rpl-21) C16A3.9(rps-13) C52E4.3(snr-4) C53H9.1(rpl-27) F22D6.3(nrs-1) F26F4.10(rrt-1) R08D7.3(eif-3.D) T05G5.10(iff-1) Y87G2A.5(vrs-2) ZC434.2(rps-7) ZK652.4(rpl-35) ZK792.2(inx-8) H06I04.4a(ubl-1) K01C8.10(oct-4)
Integrity of membrane-bounded organelles	Sparse or enlarged yolk granules	13 (2)	Secretion/endocytosis	Organelle maintenance	W03C9.3(rab-7) C10C6.6 Y53C12A.1(wee-1.3) C03C10.1(kin-19) C30F8.2 C45G9.5 F18C12.2a(rme-8) H15N14.2(nsf-1) T07C4.8(ced-9) T20G5.1 W08E3.1(snr-2) Y37D8A.10 T23G7.4(sec-5)
Egg size	Egg is larger or smaller than normal	5 (2)	Importins	–	F32E10.4(ima-3) C53D5.6(imb-3) F21H12.2 R119.4(pqn-59) C27D9.1
Aberrant cytoplasmic structures	Areas devoid of yolk granules throughout the embryo	9 (6)	–	–	C56C10.3 K04D7.1 C07G1.5(pqn-9) C09G12.9 CD4.4 E01B7.1 F23C8.6 T17E9.2a T20B12.1
Complex phenotype	Complex combination of defects that does not match other class definitions	21 (0)	–	–	C08B11.1(zyg-11) C14B9.4a(plk-1) C18E9.6 ZC404.8(spn-4) C39B5.2 C47G2.5 F07A5.1a(inx-14) F25H2.4 F26H9.6(rab-5) F43C1.2a(mpk-1) F45H11.2(ned-8) H38K22.2a M7.1(let-70) Y18D10A.17 Y37B11A.3 Y56A3A.18 Y71H2B.3 Y75B8A.30(pph-4.1) Y82E9BR.15(elc-1) F14B8.1a(nhx-4) B0334.1a(ooc-3)
All early embryonic phenotypes	Phenotypes scored in DIC test	661 (94)	Cell division/general metabolism	–	–

Representative examples are shown for genes in 23 phenotype classes, and their association with specific biochemical pathways and functions. AB and P<sub>1</sub>, anterior and posterior blastomere, respectively, of the two-cell embryo; PAR genes, genes associated with polarity of the embryo; PN, pronucleus.

combined genome-wide RNAi screening with a time-lapse microscopy readout. As a basic method we injected dsRNA into young adult worms and examined the phenotype of the resulting F<sub>1</sub> progeny<sup>2</sup>. To design a genome-wide set of inhibitory dsRNAs, each targeting only one gene, an earlier algorithm used to generate dsRNAs for chromosome III (ref. 2) was updated to optimize further the targeting specificity of dsRNAs by taking into account the fact that the average length of siRNAs in *C. elegans* is 25 base pairs (bp) (ref. 14). Thus, whenever possible, dsRNAs were designed to avoid contiguous siRNA-sized (25 bp) stretches of complementarity to off-target transcriptome sequences. We were able to design dsRNAs to 16,465 (84.2%) genes with this stringency (according to Wormbase 100; see Methods). For an additional 14% of genes, dsRNAs were designed by allowing one or more 25-mers to be present in more than one gene. By this approach, 20,326 dsRNAs targeting 19,075 (98%) of the genes in *C. elegans* were designed, produced and screened. Continuous changes in gene annotation have meant that gene coverage by the designed dsRNAs was regularly updated during the course of the project.

The primary phenotypic analysis was performed with time-lapse differential interference contrast (DIC) microscopy of F<sub>1</sub> embryos from about 20 min after fertilization until the four-cell stage, 35 min later ('DIC test'; see also ref. 2). In addition, F<sub>1</sub> broods from individual injected parents were examined to determine the viability

of the embryo as well as gross phenotypic defects at larval and adult stages ('progeny test') (Fig. 1). We first defined a set of 2,093 genes that, on the basis of their expression pattern and various bioinformatics criteria, were expected to have a higher likelihood of exhibiting functions in early embryogenesis. Each of these genes was tested individually by both the DIC and progeny tests, as described above. Results from this initial 2,093-gene set and from our earlier chromosome III study<sup>2</sup> showed that virtually all genes exhibiting early embryonic phenotypes detectable by the DIC test also show embryonic lethality in the progeny test. Therefore, dsRNAs targeting the remaining genes, which were expected to yield a significantly lower hit rate, were initially injected in pairs, and analysed by the less laborious progeny test. All dsRNAs from embryonic lethal pairs were then reinjected individually and analysed with both tests. Each incidence of a detectable phenotype was confirmed by at least two independent experiments (see Methods). During the course of the study, more than 40,000 time-lapse DIC recordings from over 19,000 individual experiments were acquired and annotated. Necessary for the interpretability of the final data set was the strict standardization of all procedures throughout this large-scale endeavour, from data collection to data annotation and analysis. For this purpose, data collection was automated by a customized laboratory information management system developed specifically for this project. The resulting data set is available online at <http://www.worm.mpi-cbg.de/phenobank2/>.

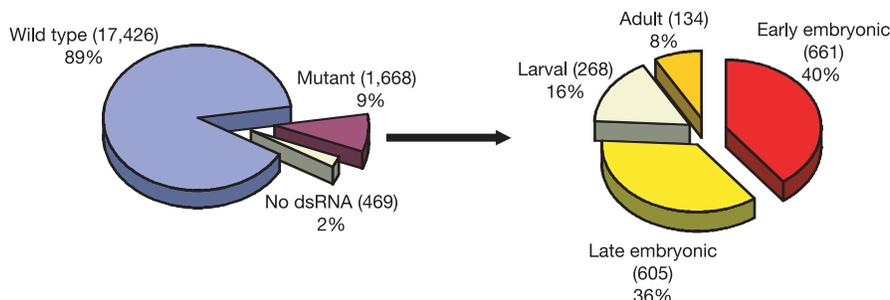


**Figure 1** Flowchart for the screening process.

**661 genes required for early embryogenesis**

Of the 20,326 dsRNAs tested so far, 1,995 dsRNAs targeting 1,668 genes, corresponding to 8.5% of all predicted genes in *C. elegans*, were found to elicit detectable RNAi-induced phenotypic alterations in at least two distinct experiments (Fig. 2).

We focused our attention particularly on the 661 genes that reproducibly exhibited altered phenotypes in the early embryo (Fig. 2 and Table 1, and Supplementary Table 1). How comprehensive was our screen in detecting those loss-of-function DIC phenotypes? Because the interpretation of negative results in RNAi screens is always problematic, we used two different methods to address this



**Figure 2** Distribution of phenotypes identified in the *C. elegans* genome-wide RNAi screen. A total of 19,075 genes were targeted by dsRNAs. Progeny tests were performed to assess late embryonic, larval and adult phenotypes. Early embryonic phenotypes were

assessed by filming the first two rounds of cell divisions with time-lapse DIC microscopy. Each incidence of a phenotype was confirmed by two independent experiments.

issue. First, we compared the results from our RNAi screen with previously published conventional genetic analyses (Supplementary Table 2). Of the 65 genes previously shown by mutagenesis to be required during the first cell division, we obtained corresponding RNAi phenotypes for 62 or 95% of these genes. Retesting the three missed genes with the same batch of dsRNAs increased this success rate to 64 (98%) of the of 65 genes, thus ascribing most of the initial failures to technical variability. These data would suggest that over 90% of genes required for the first cell division were detectable by RNAi under our screening conditions.

However, mutagenesis-based screens are typically limited in their coverage because of their tendency to focus on genes that yield robust and easily identifiable mutant phenotypes. We therefore also compared our data with two other RNAi-based screens<sup>7,15</sup> that similarly documented early embryonic phenotypes by using DIC recordings in *C. elegans* (Supplementary Table 3). We note that a significant proportion of those genes found by others to be DIC positive but annotated by us as DIC negative did, in fact, exhibit detectable DIC defects, although with insufficient reproducibility to match our criteria (see online database for detailed documentation). Furthermore, 75% of these ‘missed DIC genes’ also displayed altered phenotypes in our progeny tests. These analyses suggest a DIC detection accuracy of between 75% and 90% for our screen.

**From phenotypic annotation to gene function**

The biggest challenge of this screen was to convert the data from more than 40,000 recordings into standardized, searchable text annotations. The complexity of the DIC recordings rendered automated analyses unfeasible. Thus, to minimize the inherent variability of visual evaluations, all data sets were annotated by at least two scientists, and importantly, one person (B.S.) finalized the analysis of all recordings, giving the annotation a unity lacking when individual genes are analysed in separate, individual laboratories. We based our annotation of DIC data on a descriptive classification of all deviations from the wild type. Every DIC recording was scored for 45 distinct defect categories (Table 2). Each experiment comprised five DIC recordings of single embryos originating from five different injected worms. Reproducibility was documented as the percentage of DIC recordings exhibiting a particular defect within a given experiment (see Methods). With this strategy we achieved a high resolution of phenotypic annotation, by which combinations of defects defined gene-specific phenotypic signatures<sup>7</sup>. After identifying patterns of highly reproducible ‘core’ defects, we manually grouped the 661 DIC-positive genes according to their phenotypic signatures, thereby defining 23 phenotype classes (Figs 3 and 4 and Table 1). The defect pattern for each gene can also be presented by assigning a single-digit score of reproducibility to each of the 45 defect categories. These scores are graphically represented as distinct

Table 2 Defect categories and associated scoring criteria

No.	Defect category	Scoring criteria
1	Egg shape/size	Eggs small (less than 70% of wild type), large (more than 130% of wild type) or irregularly shaped
2	Osmotic sensitivity	Swelling of the embryo to fill the egg shell and/or irregular granule movements
3	P <sub>0</sub> cortical ruffling	Excessive or no early cortical ruffling (wild type: wave of membrane ruffling from posterior to anterior resulting in pseudo-cleavage furrow)
4	P <sub>0</sub> cytoplasmic flows	Lack of granular flows towards the male PN (wild type: directional flow of yolk granules towards the male PN)
5	P <sub>0</sub> pseudo-cleavage furrow	No or excessive furrowing (wild type: covering 20–30% the width of the embryo)
6	P <sub>0</sub> pronuclei – size/shape	Size more than 30% smaller or larger than wild type (wild type: diameter approx. 25% the width of the embryo at onset of PN migration), or irregular shape of PNs
7	P <sub>0</sub> pronuclei – number	Numbers of PNs above or below 2 (wild type: one female, one male PN)
8	P <sub>0</sub> pronuclear migration (female)	No or little migration of female PN towards the male PN (wild type: directional movement of the female towards the male PN)
9	P <sub>0</sub> pronuclear migration (male)	No migration of male PN from posterior cortex towards centre of the embryo
10	P <sub>0</sub> pronuclear meeting	Position near the cortex or centrally (wild type: near the centre of the posterior half of the embryo)
11	P <sub>0</sub> pronuclear centration	Lack of centration
12	P <sub>0</sub> pronuclear rotation	Rotation takes place after PN envelope breakdown (wild type: approx. 2–3 min before Pronuclear Envelope Breakdown)
13	P <sub>0</sub> pronuclear envelope breakdown	Lag time between PN meeting and PN envelope breakdown exceeds 8 min (wild type: 4–5 min)
14	P <sub>0</sub> spindle assembly – bipolarity	Lack of two visible poles, that is, two regularly shaped sites of yolk granule exclusion
15	P <sub>0</sub> spindle integrity	Irregular length (wild type: 25–30% the length of the embryo during metaphase), or thickness (wild type: 20–25% the width of the embryo at metaphase), or lack of rigidity
16	P <sub>0</sub> spindle elongation	Spindle is shorter or longer than 50–60% the length of the embryo at telophase
17	P <sub>0</sub> spindle rocking	No or excessive spindle rocking
18	P <sub>0</sub> spindle positioning – asymmetry	Aberrant spindle positioning at telophase (wild type: along the longitudinal axis with the posterior pole shifted posteriorly approx. 10–15%)
19	P <sub>0</sub> spindle poles	Irregular shape, in particular lack of flattening of posterior pole in telophase
20	P <sub>0</sub> cytokinesis – furrow specification	Fewer or more than two sites of furrowing and/or aberrant positioning (wild type: two sites, intersecting the long axis by a ratio of approx. 3:2)
21	P <sub>0</sub> cytokinesis – furrow ingression	Little or no ingression, or uneven ingression from the two sites
22	P <sub>0</sub> cytokinesis – completion/stability	Regression of the furrow
23	P <sub>1</sub> /AB nuclear separation – cross-eyed	Reforming daughter nuclei stay closely attached to the central cortex
24	AB nuclear migration	Time for centration of AB nucleus exceeds 7–8 min, AB nucleus migrates towards the cortex before centration (wild type: AB nucleus usually centres directly after cytokinesis)
25	P <sub>1</sub> /AB cortical activity	Excessive membrane ruffling and blebbing
26	P <sub>1</sub> /AB nuclei – number	Numbers of nuclei in daughter cells below or above 1
27	P <sub>1</sub> /AB nuclei – size/shape	Size >30% smaller or larger than WT (diameter approx. 25% of AB blastomere), or irregular shape of nuclei
28	P <sub>1</sub> nuclear migration/rotation	Lack of migration of P1 nucleus towards posterior cortex, lack of rotation of P1 spindle before nuclear envelope breakdown
29	P <sub>1</sub> /AB spindle assembly	Aberrant bipolarity or length or thickness
30	AB spindle orientation	Rotation of AB spindle (wild type: AB spindle keeps orientation, whereas P1 spindle rotates)
31	P <sub>1</sub> /AB asynchrony of divisions	Delay between AB and P1 cleavage furrow initiation is either shorter than 2 min or exceeds 5 min (wild type: approx. 2–3 min)
32	P <sub>1</sub> /AB cytokinesis	Aberrant furrow initiation or ingression or completion
33	Four-cell stage cross-eyed	Reforming daughter nuclei stay closely attached to the central cortex
34	Four-cell stage nuclei – size/shape	Irregular size and/or shape of nuclei in daughter cells
35	Four-cell stage configuration	PAR-like configurations of blastomeres
36	Tetrapolar spindle	Four poles, visible by exclusion of yolk granules
37	Yolk granules – density	Reduction more than 30%
38	Yolk granules – size	Aberrant size of individual or all granules
39	Areas devoid of yolk granules	Aberrant cytoplasmic structures excluding yolk granules
40	Polar bodies	Aberrant number (fewer or more than two) or size (matches or exceeds size of early PNs) or internalization of polar bodies
41	Unclear – multinucleate	Aberrant numbers of nuclei
42	Overall pace of events	Time span between pronuclear meeting and initiation of AB cleavage furrow exceeded 30 min (wild type: 18–22 min)
43	Number/age of embryos	Absence or reduction of one-cell and two-cell stage embryos, suggesting reduced fertility of the parent worm
44	Meiotic arrest	No visible PNs, few or no cytoplasmic events, embryo often fills egg shell
45	Other	Any other observation deferring from wild-type events
46	Inadequate test	Technical inadequacy, (focusing, coverage of recordings, etc.)

AB and P<sub>1</sub>, anterior and posterior blastomere; PAR genes, genes involved in polarity of the embryo; PN, pronucleus.

colours in Fig. 3. This type of digital representation will help to facilitate further classification efforts, for example through the application of clustering algorithms.

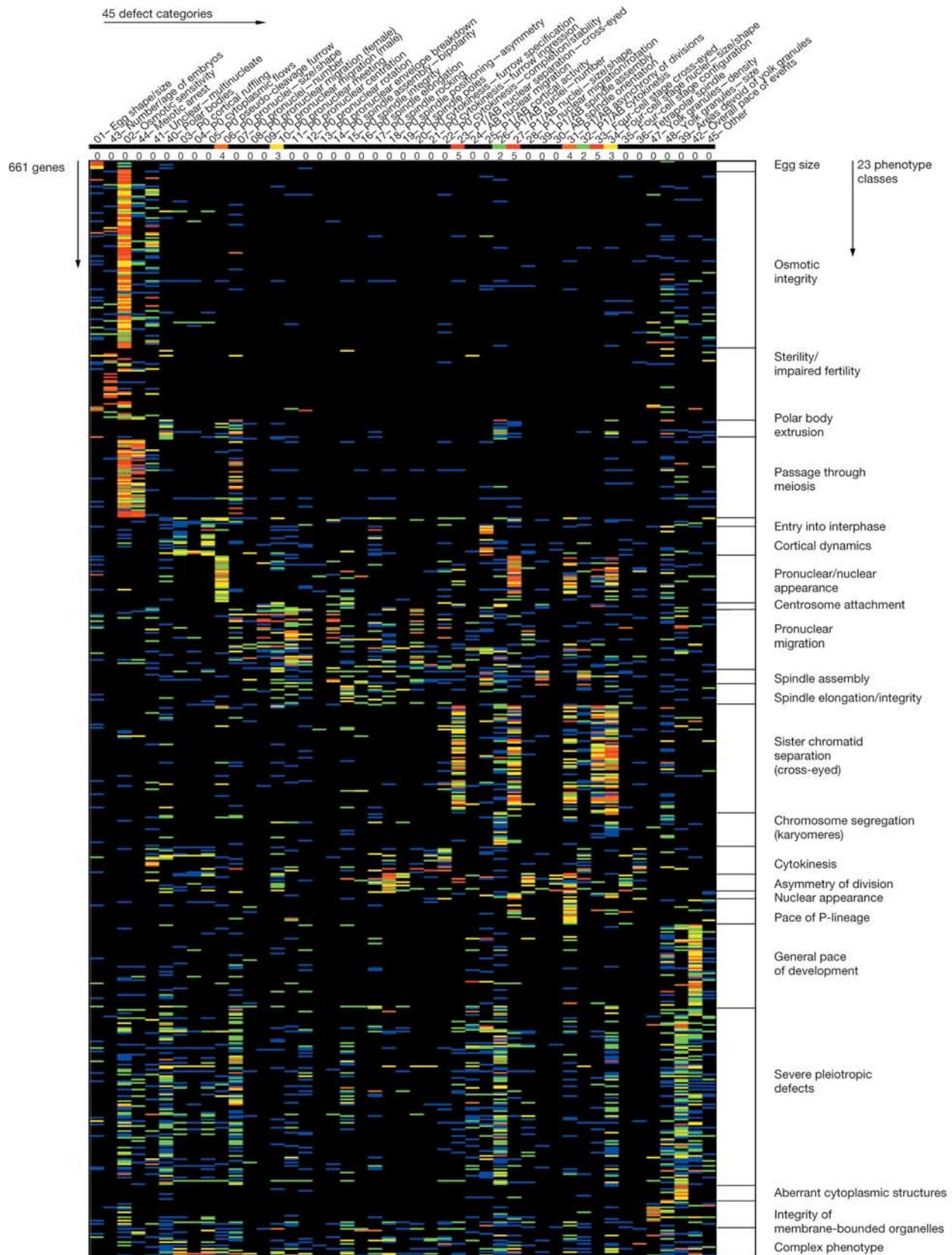
To extract new insights into gene function from this data set, we first determined whether phenotype classes correspond to specific biochemical processes by searching each class for genes of known

function. Using this approach, we found that 16 of the 23 defined phenotype classes indeed implicated specific biochemical pathways (Table 1). For example, 25 of 49 genes in the ‘pace of development’ class are known components required for ATP metabolism<sup>2</sup>. Similarly, 17 of 28 genes in the ‘pronuclear/nuclear appearance’ class are known to be involved in nuclear transport or nuclear pore

Table 3 Components of protein complexes give rise to distinct common phenotypes

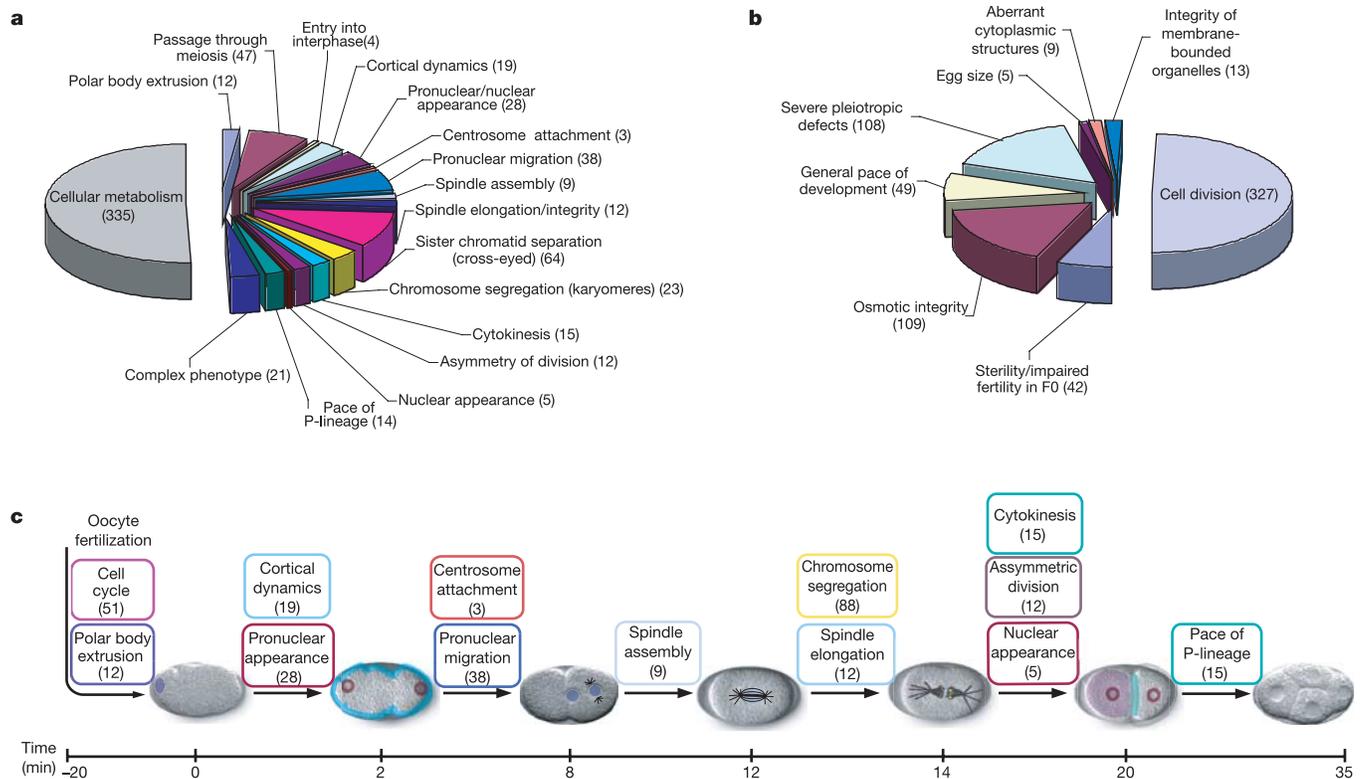
Protein complex	Total no. of genes	Genes scored		Assigned phenotype class
		No.	%	
Anaphase-promoting complex	9	6	67	100% passage through meiosis
Minichromosome maintenance	6	6	100	83% sister chromatid separation
Nuclear pore protein	20	14	70	79% pronuclear/nuclear appearance
Ribosome protein small subunit	26	24	92	79% severe pleiotropic defects
Ribosome protein large subunit	42	38	90	
α proteasome subunit 20S particle	7	7	100	93% passage through meiosis
β proteasome subunit 20S particle	7	7	100	
Non-ATPase subunit of 19S particle	11	9	82	80% passage through meiosis
ATPase subunit 19S particle	6	6	100	
F <sub>1</sub> -ATPase	10	9	90	55% general pace of development
Dynein–dynactin	12	11	92	100% pronuclear migration

Analyses were performed of the total number of genes in a complex, the number of genes associated with a phenotype in the DIC assay, and the predominantly assigned phenotype class.



**Figure 3** Systematic analysis of DIC time-lapse recordings monitoring the first two rounds of cell division in *C. elegans* embryos yields defect patterns or phenotypes that group into 23 different classes. A graphical representation of grouped defect patterns for all genes with early embryonic (DIC) phenotypes is shown. On the x-axis are 45 defect categories (see Table 2); on the y-axis are the defect patterns of one representative

experiment for each of the 661 genes showing reproducible phenotypes. The colour code resembles the reproducibility of the defect (scores out of 5): red, 5; orange, 4; yellow, 3; green, 2; blue, 1; black, 0. One example gene with associated scores for defects is shown at the top.



**Figure 4** Genes in phenotype classes are associated with distinctive biochemical pathways and key events in the first cell divisions of *C. elegans* embryos. **a**, Phenotypic classes of genes required for the process of cell division. **b**, Phenotypic classes of genes required for cellular metabolism. **c**, Phenotypic classes marking key biological events in the first round of cell division. Numbers of genes associated with each class, according to

the present study, are given in parentheses. The classes ‘passage through meiosis’ and ‘entry into interphase’ are summarized as ‘cell cycle’, and the classes ‘sister chromatid separation’ and ‘chromosome segregation’ are summarized as ‘chromosome segregation’.

assembly<sup>16</sup>, and 16 of 38 genes in the ‘pronuclear migration’ class are components required for microtubule function, including members of the dynein–dynactin complex<sup>17</sup>, tubulins and their cofactors. To confirm the accuracy of the phenotypic classification, we also examined how many genes that are thought to be part of the same protein complex gave rise to the same phenotype (Table 3). This revealed that 11 of 12 identified genes of the dynein–dynactin complex were indeed contained within the pronuclear migration class, 79% (54 of 68) of known ribosomal components were classified as ‘severe pleiotropic defects’<sup>2</sup>, and 93% (13 of 14) of known proteasome subunits were contained within the ‘passage through meiosis’ class<sup>18</sup>.

Thus, the strength of our combined phenotypic and bioinformatics analysis gives us considerable predictive power with respect to the function of previously uncharacterized genes. For example, bioinformatic analyses of genes in the respective phenotype classes allows us to propose 7 novel genes for chromosome function, 4 for nuclear envelope assembly and 4 for cell cycle progression, as well as 11 previously undescribed genes required for ATP metabolism and 7 for protein synthesis (Table 1). On the basis of this analysis, we found that 14% (94 of 661) of genes shown in this study to be required for the first cell divisions had previously not been assigned a function in any organism. Of these, about half had homologues in human, mouse, fly, yeast or *Arabidopsis* (see Supplementary Table 1). It should be noted that although DIC optics allowed us a very broad and detailed view of cellular events, more subtle defects in certain processes such as chromosome segregation cannot be seen with this technique. Nevertheless, one expects that such defects would most probably have caused detectable levels of embryonic lethality scored in our progeny testing. Time-lapse analyses of

fluorescently tagged markers in living embryos are now underway to provide further insights into these processes.

### Conclusion

Our data suggest that between 650 and 800 individual genes are required for early embryogenesis under laboratory conditions. Roughly half of these genes are involved in cell division processes such as chromosome segregation or cytokinesis, and the other half are needed for cell maintenance processes such as translation and mitochondrial function (Fig. 4 and Table 1).

In view of gene deletion studies that have identified more than 1,100 genes as being essential for *Saccharomyces cerevisiae*<sup>19</sup>, our present tally for *C. elegans* early embryos might appear surprisingly low. One possible explanation could be that the use of RNAi inherently underestimates the number of genes required for this system. However, our results indicate that the RNAi approach used here proved at least equal to the conventional ethylmethane sulphonate mutagenesis-based studies in identifying loss-of-function phenotypes.

Another possibility is that, unlike the embryo growing within the protective environment of the egg shell, somatic divisions such as those in *S. cerevisiae* or adult metazoan cells might be more complex because they must respond to external stimuli and insults arising from their more complex surroundings. Such a response requires robust signal transduction processes regulating all aspects of cell division. Early embryonic divisions follow a programmed course of division that might prove much simpler. A third possibility is that *C. elegans*, as a member of the nematode family with a particularly short development time, has at some stage simplified its cell division, perhaps owing to evolutionary pressure. Comparison

with RNAi screens in other organisms should help in distinguishing between these different ideas.

The present study illustrates the power of combining genome-scale RNAi with high-content assays to bridge the gap between systems-level biology and traditional single-gene-based analyses of function. Beyond attributing genes to various cell division processes, we were able to put our analysis into context by documenting a wider range of basic cellular functions, therefore creating a broader data set that permits exploitation by researchers from diverse fields. In particular, the ability to represent complex phenotypic data as digital signatures will be essential for integrating and comparing new data sets from disparate experimental sources. □

## Methods

### Generation of oligonucleotide primer pairs

Polymerase chain reaction primer sequences were designed by using the customized algorithm described in ref. 2, updated to avoid contiguous siRNA-sized (25-bp) stretches of complementarity to off-target transcriptome sequences. All primer sequences are available in the online database (<http://www.worm.mpi-cbg.de/phenobank2/>).

### Generation of dsRNA

Synthesis of dsRNAs *in vitro* was performed as described in ref. 2.

### Experimental set-up and phenotype readout criteria

dsRNAs were injected bilaterally into gonads of N2 hermaphrodites, maintained in accordance with standard procedures<sup>20</sup>. The animals were left at 20 °C for 24 h. Note that, in contrast to our chromosome III study<sup>2</sup>, pairs of dsRNAs were injected only for progeny test experiments.

For the progeny test, three injected animals were transferred to three individual wells of a six-well plate 24 h after injection of the dsRNA, and left at 18 °C. Three days later, the wells were checked for the presence and developmental stage of F<sub>1</sub> larvae (L1 to L4). Three days after that, the plate was inspected again for the presence and overall body morphology of F<sub>1</sub> adults and F<sub>2</sub> progeny.

Each animal generates well over 50 wild-type-looking F<sub>1</sub> progeny (wild type). In general, dsRNAs that caused defects in less than 10% of the progeny were not considered to be associated with a significant phenotype. Animals that produced fewer than 15 progeny were deemed to be sterile or to have fertility problems. Embryonic lethality was determined by counting all F<sub>1</sub> progeny at the stage of eggs and larvae (3 days after being shifted to 18 °C) and subsequently at the adult stage (6 days after being shifted to 18 °C). The number of unhatched eggs was then calculated and expressed as a percentage of the overall F<sub>1</sub> progeny.

Larval phenotypes fell into one of two groups: those in which the progeny at the final check resembled L1 or L2 animals, and those in which the progeny at this check resembled L3 or L4 animals. Other dsRNAs gave rise to a clear morphological defect in larvae or adults, such as Dpy or Unc; some were associated with slower development.

For the time-lapse DIC microscopy assay, all time-lapse experiments were performed under tight temperature control of the room at 21 ± 2 °C. Embryos were removed from the injected animals and recorded as described in ref. 2. A minimum of five embryos from five different worms were filmed, three of them from shortly after fertilization until the four-cell stage, and the remaining two for 5–10 min.

All recordings of early embryos were scored for 45 different criteria (see Table 2). Additional deviations from the wild type were noted when apparent (category 'other'). This resulted in defect patterns for individual dsRNAs, summarizing all observed defects with their respective reproducibility. All phenotypes were reproduced in at least two independent experiments. Whenever a phenotype was observed in only one or two of five recordings, but could be reproduced with a similar score in an independent experiment, it was termed a phenotype of low reproducibility. In comparison, phenotypes we describe to be of high reproducibility were scoring in four or five of five recordings in each experiment. Note that in most cases we chose to perform confirmation rounds with the same dsRNA, because the data from our chromosome III study<sup>2</sup> showed that a new 'second-best' dsRNA design often resulted in a weaker manifestation of the observed phenotype.

Similar defect patterns were manually grouped into 23 phenotype classes. To illustrate this clustering of phenotypic signatures graphically, one representative experiment was chosen for each of the 661 genes associated with a DIC phenotype, and each of the 45 defect categories was assigned a single-digit score according to its reproducibility in this experiment. These scores were then represented graphically by distinct colours (Fig. 3).

Minor deviations that were also occasionally observed in the wild type were scored at the defect level but were not taken into account for the phenotypic classification. These included slight irregularities in size for subpopulations of yolk granules, incomplete

rotation of the pronuclei in P<sub>0</sub>, and a slight delay for nuclear centration in the AB blastomere.

### Bioinformatic analysis

All statistics on genome coverage presented in this study are based on mapping to version 100 of the Wormbase genome annotation, which is the first 'frozen' genome annotation for *C. elegans*.

To identify orthologues in other species, Blast search of the relevant *C. elegans* protein was performed with the Blastp sequence-analysis program<sup>21</sup>. The sequence was searched against a collection of model organism genome databases, human, mouse, rat, *Drosophila* and yeast at the National Center of Biotechnology Information. A sequence was identified as an orthologue if it matched with an expected value of less than 10<sup>-5</sup> and was itself the reciprocal blast hit of the *C. elegans* gene. If the sequence match was less than 10<sup>-5</sup> but not a reciprocal hit, the sequence was identified as a homologue. Functional information on individual genes was retrieved from GeneOntology and Locus link.

Received 27 August 2004; accepted 10 January 2005; doi:10.1038/nature03353.

- Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
- Gönczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Clemens, J. C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA* **97**, 6499–6503 (2000).
- Kiger, A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27.1–27.15 (2003).
- Maeda, I. *et al.* Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**, 171–176 (2001).
- Piano, F. *et al.* Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* **12**, 1959–1964 (2002).
- Walhout, A. J. *et al.* Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**, 1952–1958 (2002).
- Schneider, S. Q. & Bowerman, B. Cell polarity and the cytoskeleton in the *Caenorhabditis elegans* zygote. *Annu. Rev. Genet.* **37**, 221–249 (2003).
- Gönczy, P. *et al.* Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *J. Cell Biol.* **144**, 927–946 (1999).
- Strome, S. & Wood, W. B. Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* **35**, 15–25 (1983).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Parrish, S. *et al.* Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol. Cell* **6**, 1077–1087 (2000).
- Zipperlen, P. *et al.* Roles for 147 embryonic lethal genes on *C. elegans* chromosome I identified by RNA interference and video microscopy. *EMBO J.* **20**, 3984–3992 (2001).
- Askjaer, P. *et al.* Ran GTPase cycle and importins alpha and beta are essential for spindle formation and nuclear envelope assembly in living *Caenorhabditis elegans* embryos. *Mol. Biol. Cell* **13**, 4355–4370 (2002).
- Gönczy, P. *et al.* Cytoplasmic dynein is required for distinct aspects of MTOC positioning, including centrosome separation, in the one cell stage *Caenorhabditis elegans* embryo. *J. Cell Biol.* **147**, 135–150 (1999).
- Davy, A. *et al.* A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep.* **2**, 821–828 (2001).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- 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** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Srayko, C. Cowan, L. Pelletier and W. B. Wood for critical reading of the manuscript, and M. Volkmer and S. Schloissnig for their help on the completion of the online Phenobank database. This work was supported in part by the German government through a DLR grant. A.C. was supported by the MRC, UK.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.S. (soennichsen@cenix-bioscience.com).