Genomes & Developmental Control

Identification of lineage-specific zygotic transcripts in early Caenorhabditis elegans embryos

Scott M. Robertson1, Premnath Shetty1, Rueyling Lin*

Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

Received for publication 22 July 2004, revised 7 September 2004, accepted 9 September 2004

Available online 5 October 2004

Abstract

During Caenorhabditis elegans embryogenesis, a maternally supplied transcription factor, SKN-1, is required for the specification of the mesendodermal precursor, EMS, in the 4-cell stage embryo. When EMS divides, it gives rise to a mesoderm-restricted precursor, MS, and an endoderm-restricted precursor, E. To systematically identify genes that function as key regulators of MS and/or E-derived tissues, we identified, by microarray analyses, genes that are newly transcribed within a short developmental window (approximately 30 min) encompassing the generation and fate specification of the MS and E blastomeres. By comparing total cDNAs generated from individual, carefully staged embryos, we identified 275 genes up-regulated in 12-cell embryos compared to 4-cell embryos. Fifty of these 275 genes are down-regulated in 12-cell skn-1 mutant embryos and are designated skn-1-dependent zygotic (sdz) genes. The spatial and temporal expression patterns in C. elegans embryos of 10 randomly selected sdz genes were analyzed by a nuclear GFP reporter driven by the endogenous 5′ regulatory sequence of each gene. GFP expression, although absent at the 4-cell stage, was detected at the 12- to 16-cell stage for all 10 genes and was restricted to EMS-derived lineages for 7 of the 10. Among the seven lineage-specific genes, three genes are expressed equally in both MS and E lineages, two are expressed exclusively or predominantly in the MS lineage, and two are expressed exclusively in the E lineage. Depletion of skn-1 by RNAi abolishes the expression of all seven reporter transgenes in vivo, confirming that these genes are indeed skn-1 dependent. These results demonstrate the successful combination of single-staged embryo cDNAs, genetic mutants, and whole transcriptome microarray analysis to identify stage- and lineage-specific transcripts in early C. elegans embryos.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Microarray; C. elegans; Embryos; skn-1; Lineage specific; Zygotic

Introduction

During Caenorhabditis elegans embryogenesis, factors required for early blastomere divisions, the specification of body axes, and the specification of early blastomere fates are largely contributed maternally (Schnabel and Priess, 1997). Maternally supplied factors support the development of newly fertilized embryos before the initiation of transcription from the zygotic genome.

Over the past two decades, more than 40 maternally supplied genes have been identified that have essential functions in the specification of C. elegans early blastomere fate (for a review, see Bowerman, 1998; Gotta and Ahringer, 2001; Kuersten and Goodwin, 2003; Schnabel and Priess, 1997). Most of the C. elegans maternal factors involved in cell fate specification are either themselves transcription factors or play some role, directly or indirectly, in restricting the function of a transcription factor to specific blastomeres. These maternal transcription factors regulate, either through activation or repression, the expression of zygotic target genes that then control further downstream aspects of lineage specification. Key to understanding precursor specification in C. elegans embryos is to understand how...
this transition from maternal to zygotic control is regulated (Maduro and Rothman, 2002; Newman-Smith and Rothman, 1998). Keen interest has recently centered upon identification of these target genes. However, genetic screens designed to identify these genes have to date been remarkably unsuccessful (Newman-Smith and Rothman, 1998; Terns et al., 1997). Based upon the small number of targets that have been identified, we now know that this is primarily due to a surprising degree of genetic redundancy in early embryonic transcription.

For example, a maternally supplied transcription factor, SKN-1, is required for the specification of the mesendodermal precursor EMS in the 4-cell stage embryo (Bowerman et al., 1992). When EMS divides, it gives rise to a mesoderm-restricted precursor, MS, and an endoderm-restricted precursor, E. In embryos derived from skn-1 mutant mothers, neither the MS nor E blastomeres are specified properly (Bowerman et al., 1992). SKN-1 acts, probably directly, through two functionally redundant GATA factors, MED-1 and MED-2, to specify EMS-derived tissues (Maduro et al., 2001). Depletion by RNAi of med-1 or med-2 alone does not have any detectable phenotype, whereas depletion of both simultaneously phenocopies very closely the skn-1 mutant.

Whereas MED-1 and MED-2 function redundantly in specifying EMS-derived tissues, another pair of zygotically expressed GATA factors, END-1 and END-3, function redundantly in E-derived endoderm specification (Maduro and Rothman, 2002; Zhu et al., 1997). Reduction of function of either end-1 or end-3 alone does not reveal any phenotype, while a large fraction of animals with reduction of function of both genes lack intestine (Maduro and Rothman, 2002). Downstream of END-1 and END-3, yet another pair of functionally redundant GATA factors, ELT-2 and ELT-7, regulate intestinal development (Maduro and Rothman, 2002). Redundancy of zygotically expressed genes in early embryos is not restricted to either GATA factors or the EMS lineage: for example, the T-box genes tbx-8 and tbx-9 are redundantly required for morphogenesis of the posterior body in embryos before the process of embryogenesis (Andachi, 2004; Pocock et al., 2004), while tbx-37 and tbx-38 function redundantly in anterior pharyngeal precursor downstream of Notch-mediated signaling (Good et al., 2004).

Although the SKN-1 and MED-1/2 transcription factors regulate the formation of both MS- and E-derived tissue types, the correct differential specification of MS versus E fates requires Wnt signaling (Rocheleau et al., 1997; Thorpe et al., 1997). Wnt signaling from P2 to EMS in the wild-type 4-cell embryo induces the posterior daughter of EMS, E, to generate endoderm. In the absence of this signal, the E blastomere develops tissue types (mesoderm) and a lineage pattern similar to its anterior sister, MS. Wnt signaling results in reduced nuclear levels of the maternally supplied TCF/LEF transcription factor POP-1 in E versus MS (Lin et al., 1995; Rocheleau et al., 1997; Thorpe et al., 1997). POP-1 functions, at least in part, to repress the endoderm-containing genes end-1 and end-3 in the MS blastomere (Calvo et al., 2001; Maduro and Rothman, 2002; Zhu et al., 1997). In the absence of POP-1, end-1 and end-3 are expressed in both MS and E, and both blastomeres adopt an E-like fate. In the absence of Wnt signal, POP-1 levels remain high in both the MS and E nuclei, endoderm formation is repressed, and both blastomeres adopt an MS-like fate (Rocheleau et al., 1997; Thorpe et al., 1997). Correct lineage distinction requires the Wnt signal-induced reduction of nuclear POP-1 levels specifically in the E nucleus.

In order to gain a more complete understanding of how MS and E fates are specified, we must identify those zygotic genes responsible for the development of these two blastomeres. The functional redundancy in early zygotic expression in C. elegans has confounded many attempts to identify, by genetic screens, zygotic targets of maternally supplied factors such as SKN-1 and POP-1. Therefore, in order to systematically identify zygotically expressed genes that function as key targets of maternal factors, a new experimental approach was required. We show here that by combining a whole transcriptome analysis of precisely staged early embryos with specific genetic mutants, we can identify genes that are zygotically expressed early either exclusively or predominantly in the EMS, MS, or E lineages. The identification of these genes should facilitate the analysis of both mesendoderm specification and the maternal-to-zygotic transition during C. elegans embryogenesis. Appropriate use of other genetic mutants and/or embryonic time points should allow the identification of additional zygotic transcripts that exhibit lineage specificity.

Materials and methods

Strains

N2 was used as the wild-type strain. Worms were grown and embryos were prepared at 20°C (Brenner, 1974). Genetic markers used in this study are LGIII: unc-119(ed3), tbx-11(tm278); LGIV: skn-1(zu67), dNT1; LGV: dNT1.

Embryo isolation

Gravid hermaphrodites were dissected in M9. The 2-cell embryos were collected and allowed to develop for approximately 15–20 min to 4-cell or 45–55 min to 12-cell stage embryos. Embryos were staged under the dissecting microscope by reference to diagnostic cell divisions. The 4-cell embryos were taken for cDNA preparation at approximately the mid cell-cycle stage, corresponding to the time when EMS extends its process toward ABa. The 12-cell embryos were taken for cDNA preparation approximately 5 min after the four AB descendants completed their division to 8 AB descendants.
Embryos of the desired stage were transferred individually to a drop of hypochloride solution (0.5%) on a siliconized slide for 1 min, followed by three sequential washes each in a drop of egg salt buffer (Edgar, 1995). Embryos were kept in suspension by being blown in and out of a glass pipette to prevent them from sticking to the glass slide. After washing, individual embryos were placed in 0.5 μl of cDNA first-strand buffer (see below) on the underside of the lid of a 500-μl PCR tube. A small piece of coverslip (cut to approximately 1.5 × 1.5 mm) was placed over the embryo using fine-tipped forceps and pressed down to squash the embryo (observed under the dissecting scope). The PCR tube was capped, spun briefly to bring embryo and first strand buffer to the bottom, and immediately frozen on dry ice.

RT-PCR

Upon thawing, 4 μl of first-strand buffer [50 mM Tris–HCl, pH 8.3; 75 mM KCl; 3 mM MgCl2; 0.5% (vol/vol) NP-40; 1 mM DTT; 100 μg/ml acetylated BSA; 20–40 U (1 μl)/100 μl RNA Guard (Amersham Biosciences); 30 U (1 μl)/100 μl Prime RNase Inhibitor (Eppendorf-5Prime, Boulder, CO); 0.25 mM each dNTP; 15 mM SR-T24 oligo (5′-GTAGAATCGGAGATATTCT24)] was added to each PCR tube containing a squashed embryo. After heating to 65°C for 90 s, the tube was cooled to 45°C and reverse transcription initiated with the addition of 100 U (0.5 μl) Superscript III (Invitrogen). After 60 min, the reaction was terminated by heating to 70°C for 15 min, 2.5 U (0.5 μl) RNase H (USB) was added and RNA digested at 37°C for 20 min. The cDNA strands were polyA tailed by the addition of 5 μl of 2× TdT buffer (Invitrogen; final reaction concentrations 100 mM potassium cacodylate, pH 7.2; 2 mM CoCl2, 0.2 mM DTT) containing dATP to a final concentration of 750 μM, and terminal deoxynucleotidyl transferase (TdT, Roche) to a final 1 U/μl. Supernatant III (Invitrogen). After 60 min, the reaction was stopped by heating to 70°C for 10 min. Forty microliters of PCR buffer was added to each tube (final concentrations: 10 mM Tris–HCl, pH 8.3; 50 mM KCl; 2.5 mM MgCl2; 100 μg/ml BSA, 0.05% (vol/vol) Triton X-100; 0.75 mM each dNTP, 5 μM SR-T24 primer, and 2.5 U Taq DNA polymerase (Promega)/50 μl reaction) followed by 50 cycles of PCR amplification (25 cycles of 30 s at 94°C, 30 s at 50°C, and 5 min at 72°C, then 25 cycles of 30 s at 94°C, 30 s at 50°C, and 2.5 min at 72°C). This primary RT-PCR amplification product could then be reamplified (primary product diluted 1:100; PCR of 25 cycles of 30 s at 94°C, 30 s at 50°C, and 3 min at 72°C) as needed to generate more total cDNA.

Pseudo northern

Following PCR reamplification, 10 μl (approximately 2 μg) from multiple single embryo samples were electrophoresed side-by-side through a 1.5% agarose gel (4–5 V/cm; 20 min) and transferred to charged nylon membrane (Zetaprobe GT, Biorad). Membranes were hybridized overnight at 65°C with random-primer 32P-labeled cDNA fragments (Ready-to-Go Labeling, Amersham Biosciences) including the 3′ region of the genes (see below) and washed (Church and Gilbert, 1984) prior to exposure to autoradiographic film (Kodak). DNA fragments used to analyze individual 4- and 12-cell embryo cDNAs, all derived from cDNA, are tba-1, 1037 bp including 136 bp 3′ UTR; cdk-1, 859 bp including 55 bp 3′ UTR; med-1, 809 bp including 280 bp 3′ UTR; end-1, 836 bp including 165 bp 3′ UTR; end-3, 870 bp including 141 bp 3′ UTR; and pes-10, 1475 bp including 86 bp 3′ UTR. Approximately half of the skn-1 12-cell embryos expressed pes-10, whereas pes-10 expression in the others was greatly reduced or undetectable. The three skn-1 12-cell cDNAs selected for these studies were pes-10 negative.

Microarray hybridization

Microarray hybridizations were carried out at the Stanford Microarray Facility using a custom-designed PCR fragment array (total spot number = 19,213) that covers approximately 94% (n = 17,815) of predicted worm genes (Kim et al., 2001; Reinke et al., 2000). Individual wild-type 4-cell and skn-1 12-cell total cDNAs were Cy3 labeled (n = 3 in each group) and hybridized versus three Cy5-labeled individual wild-type 12-cell total cDNAs. Six hybridizations in total were utilized to generate the data set presented here (4-cell wild-type versus 12-cell wild-type in triplicate, and 12-cell skn-1 versus 12-cell wild-type in triplicate).

All cDNA samples were aminoallyl-dUTP derived and fluorescently labeled with either Cy3 or Cy5 as follows: 1 μl of 1:100 diluted primary PCR product was added to 49 μl polymerase buffer containing 5 μM SR-T24 primer, 0.2 mM each dCTP, dGTP, dATP, 0.1 mM dTTP, 0.2 mM aminoallyl-dUTP (Sigma-Aldrich, St. Louis, MO), and 2.5 U Taq polymerase, and the mixture amplified through 30 cycles (30 s at 94°C, 30 s at 42°C, 2 min at 72°C). Aminoallyl-derived cDNA (typically 5 μg/reaction) was purified using a ProbeQuant G-50 microspin column (Amersham Biosciences). Several reactions were pooled and concentrated to approximately 8–10 μl by centrifugation under vacuum. Average yield of aminoallyl cDNA prior to Cy3/Cy5 labeling was 20–25 μg. Eight microliters of aminoallyl cDNA was added to 8 μl of 0.1 M NaHCO3, mixed and kept on ice. The contents of one vial of Cy3 or Cy5 monofunctional reactive dye (Amersham Biosciences) were dissolved in 15 μl dimethyl sulfoxide (DMSO), 5 μl of the appropriate dye transferred to the tube containing aminoallyl cDNA and incubated at room temperature in the dark for 30 min. Cy3- or Cy5-labeled cDNA was purified using ProbeQuant G-50 microspin columns, and the eluate was concentrated to 7–10 μl by vacuum centrifugation. Yield and labeling efficiency were measured by absorbance
at 260 nm (yield) and 550 nm (Cy3 labeling efficiency) or
649 nm (Cy5 labeling efficiency). Yields of Cy3- or Cy5-
labeled cDNA averaged 19 μg, and dye incorporation
averaged 1 dye molecule per 50 nucleotides. Approx-
imately 20 μg each of the appropriate Cy3- and Cy5-
labeled cDNAs was hybridized to the microarray. Under
these conditions, many array spots, including all high
abundance maternal mRNAs, were oversaturated, whereas
the signals from a number of known low abundance
transcripts were rendered detectable.

Normalization of microarray data

Our hybridization conditions resulted in oversaturation
of highly abundant housekeeping genes. As these genes
are commonly used to normalize the raw data values in
both channels of a microarray hybridization, we sought
another approach. Normalization, however, was also
complicated by the following two factors: one, a large
number of maternal transcripts are actively being
degraded as development proceeds (class II; Seydoux
and Fire, 1994) and therefore would be elevated in the 4-
cell total cDNAs compared to the 12-cell cDNAs; and
two, newly transcribed zygotic genes would be elevated
in the 12-cell cDNAs versus 4-cell cDNAs. As there
would be an unknown but presumably relatively large
number of real differences between the 4- and 12-cell
cDNAs, we attempted a normalization based upon a
comparison of as large a group of genes as possible
expressed at a user-defined low-to-moderate level.
Because raw values between 30,000 and 40,000 were
considered near saturation, we selected a window of raw
values between 2000 and 5000 for channel 2 (Cy5-
labeled wild-type 12-cell embryo cDNA in all six
hybridizations), which encompassed approximately 1500
genes. An average channel 2 raw data value was
calculated for this gene set, divided by the average
channel 1 raw data value for the same genes, and then
the entire channel 2 raw data set divided by this value to
generate the normalized channel 2 values. Comparisons
were then made between the raw channel 1 values (wild-
type 4-cell or skn-1 12-cell) and the normalized channel
2 values (wild-type 12-cell) for each hybridization to
define both the zygotic and sdz gene sets. Because of
concerns regarding real systematic differences between 4-
and 12-cell samples, we also performed this calculation
in the opposite direction: raw values between 2000 and
5000 in channel 1 (approximately 5800 genes) were
averaged and divided by the average raw value for the
same genes in channel two, and then the entire channel 2
raw data set multiplied by this value. When either
normalization process is applied, exactly the same 50
sdz genes are identified (see Results). The results
presented are based upon the latter normalization method.
Further details concerning the microarray hybridization and
normalization are available upon request.

Plasmid construction

Most GFP reporters were generated by fusing the
genomic sequence upstream of the corresponding predicted
open-reading frames to pRL1075, a derivative of pAP.10
(Gaudet and Mango, 2002), pRL1075 was generated by
replacing the 0.5-kb Stul-Asp718 fragment with a gateway
destination vector converting cassette B (Invitrogen). The
5′ genomic sequence was PCR amplified and recombined
into pRL1075 according to the manufacturer’s instructions.
This placed the presumptive promoter sequence 5′ of
GFP::histone H2B, which generated a nuclear GFP signal
in cells expressing the reporter gene. The 5′ regulatory
sequence analyzed for each gene was chosen based on the
predicted gene structure and proximity to its nearest 5′
neighboring gene. For genes whose nearest 5′ neighbor
was less than 2 kb away, we used the entire sequence
between the predicted coding sequences of those two
genes to generate the reporter constructs. For genes with
nearest 5′ neighbors more than 2 kb away, an arbitrary
length of 5′ genomic sequence (2–6 kb) was chosen to
generate the promoter fusions. The reporter construct for
sdz-1 was a translational fusion, which contains approx-
imately 600 bp downstream and approximately 1.2 kb
upstream of the predicted AUG. None of the 10 genes
tested contains introns larger than 300 bp. Sizes of
genomic DNA fragments tested as presumptive 5′ regula-
atory sequences for each gene are sdz-1: 1832 bp; sdz-4:
999 bp; sdz-23: 1236 bp; sdz-26: 2299 bp; sdz-31: 1373
bp; sdz-33: 651 bp; sdz-36: 1545 bp; sdz-38: 5834 bp; nit-
l-1: 767 bp; and thx-35: 734 bp.

Transgenic animals

Transgenic animals were generated by injection of gfp
constructs with the unc-119 rescuing plasmid pDP#MM016B
into the unc-119(ed3) mutant strain as described (Maduro
and Pilgrim, 1995). For each construct, at least three independent
transgenic lines were analyzed and the expression patterns
were consistent between the three lines. The following
transgenes are maintained as extrachromosomal arrays in
the respective strains: P
sdz-1 gfp::H2B (pRL1373) in
tX594(teEx223); P
sdz-4 gfp::H2B (pRL1376) in
tX598(teEx227); P
sdz-23 gfp::H2B (pRL1370) in
tX590(teEx219); P
sdz-26 gfp::H2B (pRL1350) in
tX593(teEx222); P
sdz-31 gfp::H2B (pRL1371) in
tX591(teEx220); P
sdz-33 gfp::H2B (pRL1375) in
tX596(teEx225); P
sdz-36 gfp::H2B (pRL1379) in
tX599(teEx228); P
sdz-38 gfp::H2B (pRL1374) in
tX595(teEx224); P
nit-1 gfp::H2B (pRL1372) in
X592(teEx221); and P
thx-35 gfp::H2B (pRL1377) in
X597(teEx226). In addition, an integrated line TX
585(tels18) carrying pRL1370 was generated by micro-
particle bombardment. The transmission frequency of each
transgene in its respective strain is as follows: 62% (TX594),
51% (TX598), 73% (TX590), 100% (TX585), 71% (TX593),
65% (TX591), 54% (TX596), 65% (TX592), 56% (TX599), 66% (TX597), and 51% (TX595).

Microscopy

Imaging of live embryos was performed using an Axioplan microscope equipped with epifluorescence and differential interference contrast (DIC) optics, a MicroMax-512EBFT CCD camera, Ludl Electronic Products filter wheels, shutters, and z-axis controller (Rogers et al., 2002). All embryo images were collected in grayscale with a custom software package (os4d 1.0, freely available upon request to jwaddle@mail.smu.edu). Embryos shown in Figs. 5 and 6 were collected at the 2-cell stage, allowed to develop further, and imaged at approximately the 16- to 24-cell stage.

RNA interference

The dsRNA for skn-1 was synthesized from a full-length cDNA (a gift from Keith Blackwell) whereas the dsRNAs for the two zygotic gene set members, F38C2.5 and Y57G11C.25, were synthesized from the corresponding genomic fragments in the Ahringer RNAi library (Fraser et al., 2000; Kamath et al., 2003). RNAi was performed by injection into the transgenic animals (Rogers et al., 2002). Early embryos were collected by cutting open RNAi animals approximately 24–36 h after injection. The skn-1(RNAi) animals produced 100% (n > 1000) dead embryos with a phenotype indistinguishable from skn-1(zu67) embryos (Bowerman et al., 1992). Concentrations of dsRNAs used were 2.5 mg/ml for skn-1, 6 mg/ml for F38C2.5, and Y57G11C.25.

Results

Overview

In order to systematically identify zygotically expressed genes that function as key targets of maternal factors, a novel experimental approach was deemed necessary. We have taken a molecular approach that should allow the identification of zygotically expressed genes during any defined developmental window and in specific lineages. We chose to first test this approach by focusing on the mesendodermal precursor EMS and its immediate descendents, MS and E, for two reasons: first, several key transcription factors, both maternal and zygotic, that regulate the fates of EMS, MS, and/or E have been identified; and second, well-characterized genetic mutations that affect the fate specification of each of these blastomeres are available (Fig. 1).

We first identify by whole transcriptome analyses those zygotic genes that are newly transcribed over a developmental time span during which the MS and E blastomeres are generated and their fates specified, that is, the 4- to 12-cell stage. We then compare the genes transcribed within this window in wild-type and skn-1 mutant embryos to identify a subset of genes whose expression is dependent upon the SKN-1 transcription factor. Genetic studies have determined that SKN-1 plays a critical role in EMS blastomere specification, and therefore genes that are down-regulated in skn-1 mutants are likely to be expressed predominantly or exclusively in the EMS, MS, or E blastomeres. We then validate the microarray data by employing an expression-based assay within the intact embryo for randomly selected candidate genes.
Single-embryo cDNAs

Total cDNA was prepared and amplified from individual carefully staged *C. elegans* embryos using the protocol of Brady (Brady and Iscove, 1983; Robertson et al., 2000), as modified by Iscove et al. (2002) and us (see Materials and methods) to optimize microarray hybridization. The 4-cell, 8-cell, and 12-cell total cDNAs were prepared from individual wild-type N2 embryos, along with 12-cell cDNAs from individual *skn-1(zu67)* mutant embryos. For each wild-type stage, at least 40 individual cDNA samples were prepared and analyzed, while at least 20 12-cell *skn-1(zu67)* mutant embryos were analyzed.

The quality of these total cDNAs was assayed by pseudo-Northern blot analyses, utilizing probes derived from ‘housekeeping’ genes as well as the small number of zygotic genes known to be preferentially expressed in either the 4- or 12-cell stage embryos. Alpha tubulin, *tba-1*, and the *cdc-2* homolog, *cdk-1*, have been shown to be ubiquitously detected in early embryos (Seydoux and Fire, 1994) and serve as controls for the quality of RT-PCR. As expected, all our cDNA samples expressed very high levels of *tba-1* and *cdk-1* (Fig. 2).

Transcripts from the *med-1* gene, encoding a GATA transcription factor, are detected at a high level in wild-type 4-cell samples, but a low level in 8- or 12-cell samples (Fig. 2 and data not shown). This is consistent with genetic and molecular characterization that the *med-1* gene is a direct target of the SKN-1 transcription factor (Maduro et al., 2001). On the contrary, transcripts from two other GATA factor encoding genes, *end-1* and *end-3*, are detected only in our 8-cell (not shown) and 12-cell wild-type samples and not the 4-cell wild-type samples (Fig. 2). This is also in agreement with in situ hybridization results indicating that the expression of *end-1* and *end-3* is restricted to the E blastomere and its descendants starting at the 8-cell stage (Maduro and Rothman, 2002; Zhu et al., 1997).

The 12-cell stage total cDNAs were also prepared from individual *skn-1(zu67)* mutant embryos. In *skn-1* mutants, EMS is not properly specified. The *skn-1* mutant embryos are defective in MS-derived (100%) and E-derived tissues (60%) (Bowerman et al., 1992). We detected very low or no *end-1* or *end-3* transcripts in the *skn-1* 12-cell cDNAs (*n* = 20, Fig. 2).

C. elegans microarray hybridization

Microarray hybridizations were carried out at the Stanford Microarray Facility using a custom-designed array that covers almost all known and predicted worm transcripts (Kim et al., 2001). Three individual wild-type 4-cell and *skn-1* 12-cell cDNAs were labeled with Cy3 (green) and hybridized versus three individual Cy5-labeled (red) wild-type 12-cell samples. Six hybridizations in total were used to generate our data set: wild-type 4-cell versus wild-type 12-cell (Figs. 3 and 4, hybridizations 1–3) and *skn-1* 12-cell versus wild-type 12-cell (Figs. 3 and 4, hybridizations 1–3).
hybridizations 4–6). Hybridization results were normalized as described in Materials and methods and validated by a panel of control and stage-specific genes (Fig. 3A). These samples, although generated from extremely small amounts of starting material, yielded robust, reproducible, and verified signals following microarray hybridization. ‘Housekeeping’ genes such as actins, tubulins, and cdk-1 are highly expressed in all samples and show little fold-change in expression levels between wild-type 4-cell, wild-type 12-cell, or skn-1 12-cell samples (appear black or near-black in Figs. 3 and 4). Genes, such as end-1, end-3, and pes-10, known to be newly expressed at the 8- to 12-cell stage (Maduro and Rothman, 2002; Seydoux and Fire, 1994; Zhu et al., 1997; and Fig. 2), give microarray hybridization results consistent with this developmental time course. The end-1, end-3, and pes-10 genes exhibit approximately 6-, 7-, and 20-fold elevated expression, respectively, in 12-cell versus 4-cell embryos (Figs. 3 and 4). Our pseudo-Northern results, along with other lines of evidence, show that expression of these three genes is SKN-1 dependent (Maduro and Rothman, 2002; Maduro et al., 2001). Accordingly, the microarray data show end-1, end-3, and pes-10 expressions to be 7-, 12-, and 19-fold reduced, respectively, in 12-cell skn-1 embryos versus 12-cell wild-type embryos. The med-1 expression begins in the EMS blastomere at the 4-cell stage and then decreases (Maduro et al., 2001), as confirmed by our pseudo-Northern results (Fig. 2 and data not shown). The microarray results show med-1 expression elevated over four- and sevenfold in the 4-cell embryo samples compared to the 12-cell embryo samples in two out of three hybridizations. The raw channels 1 and 2 values for med-1 were much lower in the third hybridization and therefore the anomalous result in this case is likely caused by a PCR or spotting artifact on the microarray slide. These results validate the sensitivity and accuracy of the microarray hybridization results and further demonstrate that conditions have been optimized such that we are able to correctly identify early zygotic transcription factors that exhibit blastomere-restricted expression. Examination of a total of 22 additional maternal and zygotic genes whose expressions in wild-type early embryos have been investigated by in situ hybridization or immunofluorescence further validates our normalization (Fig. 3B). All maternally expressed genes examined (n = 16) showed either lower transcript levels at the 12-cell than the 4-cell stage (green in Fig. 3B) or were unchanged (i.e., less than twofold difference; black or near-black in Fig. 3B). We examined six known zygotic genes that have been shown by in situ hybridization to begin to express at the 8- to 12-cell stage in wild-type embryos, vet-1, vet-2, vet-3, vet-4, vet-6, and

Fig. 3. Validation of microarray analysis by known maternal and zygotic transcripts. (A) The expression levels of control genes as analyzed in the microarray hybridizations. In all six hybridizations, total cDNA from wild-type 12-cell embryos was Cy5-labeled (red) and compared to Cy3-labeled cDNA (green) as indicated. Genes with elevated expression levels in the wild-type 12-cell embryo versus either the wild-type 4-cell or skn-1 12-cell will appear shades of red, with all elevations above fivefold appearing as bright red (see scale bar). Genes whose expression levels vary little between the samples will appear black or near-black. Those genes whose expression levels are higher in the wild-type 4-cell embryo will appear shades of green. (B) The microarray expression levels of an additional 23 genes whose expression levels have been analyzed by in situ hybridization. I (Type I maternal genes): genes that are maternally supplied and maintained at a high level in early embryos. Expectation: black or near-black. II (Type II maternal genes): genes that are maternally supplied and degraded in somatic blastomeres but maintained in the germline lineage in early embryos. Expectation: green. zyg (zygotic genes): genes that are expressed only zygotically in early embryos. Expectation: red. Note however that med-1 is zygotically expressed at the 4-cell stage and then transcript levels are down-regulated. All other zyg class genes are elevated in expression at the 12-cell stage versus the 4-cell stage.
nhr-2 (Seydoux and Fire, 1994; Seydoux et al., 1996). Four of these genes are strongly up-regulated in our 12-cell samples, with vet-1, vet-3, vet-4, and vet-6 up-regulated over 11-, 20-, 18-, and 12-fold, respectively. The microarray genomic PCR fragment for vet-2 is derived from the far 5′ region of a large gene. Because the total single embryo cDNAs are underrepresented in 5′ sequences, this probably accounts for the decrease in overall signal in our microarray analysis, leading to vet-2 appearing only mildly up-regulated (over threefold) in the 12-cell sample. The nhr-2 gene is up-regulated 2.3-, 1.8-, and 1.6-fold in the three microarray hybridizations. The normalized channel 2 (12-cell) signals for this gene are quite high (approximately 5200, 7140, and 4750), and the low fold values derive from the high channel 1 values to a maternal contribution to the 4-cell nhr-2 transcript level (Sluder et al., 1997).

SKN-1-dependent early zygotic transcripts (sdz genes)

The microarray hybridization results were first analyzed to identify genes that are transcriptionally activated during the 30- to 40-min time period from the 4-cell to 12-cell stage embryo (these will be referred to as the 12-cell zygotic gene set). The microarray data set was filtered to identify genes whose expression at 4-cell was either low or background (channel one raw V 3500) and whose expression at 12-cell was elevated 2.5-fold or more in at least two of the three hybridizations. This analysis defined 275 12-cell zygotic genes. This filtering excluded known zygotic genes, for example vet-1 and vet-4, whose expression levels were already high at the 4-cell stage. Fifty-three of the 275 12-cell zygotic genes (approximately 20%) are elevated at least 10-fold compared to the 4-cell samples.

Among the 12-cell zygotic genes, we further defined a subset of genes that were down-regulated at least threefold in all three cDNA samples generated from skn-1 12-cell embryos (Fig. 4). These genes (n = 50) are designated skn-1-dependent zygotic (sdz) genes. Nine of the sdz genes identified in this manner, including end-1 and end-3, have previously been described and we will refer to them by their accepted gene names. In addition, three of the sdz genes, based upon sequence analysis, belong to two different functional gene classes. DC2.7 and R02C2.2 encode kinases and will be referred to as kin-33 and kin-34, whereas ZK1058.6 encodes a nitrilase and will be referred to as nit-1. The other 38 genes are numbered sequentially sdz-1 through sdz-38 (Fig. 4). These gene name designations have been accepted by the CGC Genetic Map and Nomenclature Curator (http://biosci.umn.edu/CGC/Nomenclature/nomenguid.htm). It was our expectation that this group of sdz genes would contain both SKN-1 regulated EMS transcripts (i.e., both direct and indirect targets) as well as MS- and/or E-restricted transcripts.

Fig. 4. sdz genes identified in microarray analyses. Left: microarray hybridization pattern of the 50 genes that are zygotically expressed (appear red in hybridizations 1–3) and whose expression is down-regulated in skn-1 samples at least threefold in all three arrays (appear red in hybridizations 4–6). This defined the sdz gene set. Right: average fold differences among the three arrays for each sdz gene. First column indicates fold elevation in wild-type 12-cell embryos versus wild-type 4-cell embryos. Second column indicates fold reduction in expression in skn-1 mutant 12-cell embryos versus wild-type 12-cell embryos. The sdz gene set is listed alphabetically according to clone name (in brackets). Four control genes are included at the top.
Classes of early zygotic genes

Sequence analysis of the 12-cell zygotic and sdz gene sets shows that the vast majority derives from a relatively small number of functional gene classes (Table 1).

Transcription factors

Central to the maternal-to-zygotic transition is the conversion from using maternally supplied transcription factors to those encoded from the zygotic genome to regulate further development (Maduro and Rothman, 2002; Newman-Smith and Rothman, 1998). Approximately 12% and 14% of 12-cell zygotic genes and sdz genes, respectively, encode proteins predicted to be transcription factors or to have a DNA binding domain. Two of these sdz genes, tbx-35, which encodes a T-box protein, and sdz-38, which encodes a C2H2 zinc finger containing protein, were subjected to further expression analysis (next section).

Signaling proteins and membrane proteins

Although C. elegans has an invariant lineage (Sulston and Horvitz, 1977; Sulston et al., 1983), suggestive of a very cell-autonomous developmental program, studies over the past 20 years have demonstrated that cell–cell interactions play an essential role throughout its development (Greenwald, 1997; Schnabel and Priess, 1997). Approximately 12% of 12-cell zygotic genes and 18% of sdz genes encode proteins with predicted N-terminal signal peptides and/or transmembrane domains. Additional expression analysis for one of these genes, sdz-31, encoding a predicted membrane protein, is presented in the next section.

Proteins involved in protein degradation

The largest class of both 12-cell zygotic and sdz genes encode proteins implicated in protein degradation, including F-box proteins, BTB-domain containing proteins, and ring finger containing proteins. These three classes of genes represent 34% and 22% of the 12-cell zygotic genes and sdz genes, respectively.

Proteins of unknown function

Approximately 29% and 30% of the 12-cell zygotic genes and sdz genes, respectively, encode proteins with no apparent domains or domains of unknown functions.

Only 13% of the 12-cell zygotic gene set and 16% of the sdz gene set fall outside of these four classes.

Potential redundancy

Sequence analysis of the 12-cell zygotic and sdz gene sets identifies several cases of potential genetic redundancy, consistent with the initial premise that many key zygotic regulators function redundantly. For example, there are approximately 20 genes in the C. elegans genome predicted to encode CCCH type zinc fingers. At least seven CCCH proteins have been shown to play key regulatory roles in either oocyte development or early embryogenesis (Detwiler et al., 2001; Guedes and Priess, 1997; Mello et al., 1992; Schubert et al., 2000; Tabara et al., 1999). All seven of these genes are expressed maternally and at least two pairs (MEX-5/MEX-6 and OMA-1/OMA-2) have been shown to have redundant functions (Detwiler et al., 2001; Schubert et al., 2000). Microarray fragments for 2 of the 20 CCCH genes derive from the very 5’ end of those genes and therefore may not be suitable targets for the single embryo total cDNAs. Of the 18 genes predicted to encode CCCH proteins that would be hybridized with our cDNAs, only two, F38C2.5 and Y57G11C.25, were identified among the 12-cell zygotic gene set. In order to test whether this gene pair demonstrated functional redundancy, we examined the role of these two CCCH proteins in embryo development. We could show that while depletion of either F38C2.5 or Y57G11C.25 singly by RNAi resulted in no observable phenotype (0%, n > 500), depletion of both genes simultaneously by double RNAi resulted in 100% embryonic lethality (n > 500). Embryos derived from double RNAi animals appear to produce all tissue types but fail to undergo proper morphogenesis and elongation. The detailed characterization of this double RNAi phenotype will be reported elsewhere. These results do, however, support our original contention that this approach could identify the important class of functionally redundant early zygotic genes. We have tested the following combinations of depletion and/or knockout for genes with either the same spatial expression pattern in embryos (see next section) or belonging to the same gene family and so far have not observed any abnormal phenotype in any combination: sdz-23(RNAi);sdz-26(RNAi), sdz-1(RNAi);sdz-31(RNAi);nit-1(RNAi), sdz-31(RNAi);nit-1(RNAi), tbx-35(RNAi);sdz-38(RNAi); and tbx-35(RNAi);tbx-11(tm278).
In vivo expression patterns of selected sdz genes

To validate the effectiveness of this approach in identifying lineage-specific zygotic transcripts, we randomly selected 10 sdz genes to examine their temporal and spatial expression patterns in early embryos. Following promoter fusion to a GFP::histone H2B reporter and the generation of transgenic lines, expression in the early embryo can be followed in a precise spatiotemporal manner by nuclear GFP fluorescence. All 10 selected genes demonstrated nuclear GFP fluorescence in 12-cell stage embryos, whereas no nuclear fluorescence was detected at the 4-cell stage. Seven of these 10 genes exhibit an expression pattern that is either restricted to, or preferentially in, the EMS, MS, or E lineages prior to the 28-cell stage (see below). The other three gene expression profiles (sdz-4, sdz-33, and sdz-36) appeared to be ubiquitous starting at the 12-cell stage. We do not yet know whether this represents their actual in vivo expression patterns or whether our reporter constructs for these particular genes did not contain all sequences required for lineage-specific expression.

Back-hybridizations of 15 genes to the 4- and 12-cell single embryo cDNAs, including the 7 EMS, MS, or E-restricted sdz genes and one 12-cell zygotic gene shown to be ubiquitously expressed by the GFP reporter assay (T10C6.8), confirms that all 15 are elevated in transcript levels by the 12-cell stage (data not shown).

Fig. 5 shows examples for each lineage-restricted sdz gene expression pattern. (1) EMS-restricted: sdz-1, sdz-31, and nit-1 are EMS-restricted, that is, expressed specifically and equally in both the MS and E lineages. While sdz-1 and sdz-31 continued to be detected exclusively in the EMS lineage throughout early embryogenesis, nit-1 expression shifted to being primarily in the E descendants after the 300-cell (approximately 4E) stage. (2) E-restricted: sdz-23 and sdz-26, like end-1 and end-3 (not shown), were first detected in the E blastomere at the 8-C stage and remained E lineage-restricted throughout embryogenesis. (3) MS-restricted: tbx-35 and sdz-38 fluorescence were first detected at the 14-cell stage. The tbx-35 (encoding a T-box transcription factor) expression was detected exclusively in MS descendants, whereas sdz-38 (encoding a C2H2 zinc-finger DNA binding protein) was detected primarily in MS descendants and very weakly in the E descendants. This represents the first report of any genes exhibiting an MS-restricted expression pattern. The promoter fusion assay only allows us to estimate the start of expression and the spatial expression pattern for each gene and does not necessarily accurately reflect the duration of the expression within each lineage.

We further examined whether the expression of these 10 randomly selected sdz genes in vivo is dependent on the SKN-1 transcription factor by depleting skn-1 in transgenic animals. RNA interference was performed by injection of double stranded skn-1 RNA into the adult gonads and examining the GFP pattern in embryos generated within the period 24–36 h postinjection. One hundred percent of the RNAi embryos generated within this period are lethal with a characteristic skn-1 mutant phenotype. For the three genes that exhibit ubiquitous expression at the 12-cell stage, GFP nuclear fluorescence was unchanged in response to skn-1(RNAi). In all seven cases where expression is limited to the EMS, MS, or E lineage, GFP nuclear fluorescence is eliminated in skn-1(RNAi) embryos (≥98%, n > 200 for each strain tested, Fig. 6). This confirms in vivo the microarray results that these genes are indeed skn-1 dependent.
Discussion

We report here the identification of stage- and lineage-specific transcripts in early *C. elegans* embryos by whole transcriptome analysis using total cDNAs prepared from individual wild-type and mutant embryos. Using this approach, we identified 275 genes up-regulated during embryonic development from the 4-cell to the 12-cell stage. This approximately 30-min developmental window includes the fate specification of the mesendodermal precursor EMS, as well as its two daughters, the endodermal precursor E and the mesodermal precursor MS. By comparing total cDNA prepared from individual wild-type 12-cell embryos to cDNA prepared from individual *skn-1* mutant 12-cell embryos, we defined 50 of these genes as *skn-1*-dependent zygotic (*sdz*) transcripts within this developmental window. We propose that this gene set includes targets, both direct and indirect, of the maternally supplied transcription factor SKN-1 that function in specification or early development of these lineages. Both *end-1* and *end-3*, expressed specifically within the E lineage and to date the only known putative direct SKN-1 targets, are included within the *sdz* gene set. By promoter fusion analysis of 10 randomly selected *sdz* genes, we showed that all exhibit elevated expression by the 12-cell stage and that seven are expressed specifically in those lineages (namely, EMS, E, or MS) previously shown to require *skn-1* activity. We do not yet know whether the three ubiquitously expressed 12-cell zygotic genes included among the *sdz* gene set represent false-positives or lack sufficient regulatory sequence in the GFP reporter constructs. We are currently testing larger promoter fragments to address this question. We are currently testing larger promoter fragments to address this question. We are currently testing larger promoter fragments to address this question. We are currently testing larger promoter fragments to address this question.

**Single embryo cDNA**

One strength of this particular approach is that total cDNAs are prepared from individual carefully staged embryos. This circumvents pooling of embryos. There is at present no good way to synchronize *C. elegans* embryos, and it would therefore be difficult to pool even a small number of embryos of identical stages. As we have noticed some variation in the expression of control genes from embryo to embryo, a short developmental window, such as that analyzed here, would be blurred considerably by pooling. This variation in control gene expression is more significant in cDNAs prepared from various 12-cell mutant embryos, including *skn-1(zu67)*, *pop-1(zu189)*, and *mom-2(or9)* (data not shown) than wild-type 12-cell embryos. We believe this sample-to-sample variation to be primarily due to genetic penetrance effects, compounded by the combined effects of actual stage-specific variation among identically aged embryos and an inherent imprecision in the staging of selected embryos. Analysis of individual embryos allows precise staging based upon a standardized set of molecular parameters, allowing full advantage to be taken of actual embryo-to-embryo developmen-
opmental variation. By using three individual 4-cell wild-type, three individual 12-cell wild-type, and three individual 12-cell skn-1 embryos as sources for cDNA and by focusing on genes whose expression patterns are replicated in independent hybridizations, we attempted to define a gene set that accurately represents the most significant developmentally regulated zygotic transcripts.

Sufficient cDNA for labeling is prepared from single 4- or 12-cell embryos by extensive PCR amplification. We maintain, and it has previously been reported (Iscove et al., 2002), that cDNA amplified in the manner described here maintains relative representation of mRNA species, including low abundance transcripts. Having analyzed almost 40 genes whose expression profiles in the early embryo have previously been characterized, the amplified stage-specific cDNAs recapitulate accurately the in vivo situation. We have also optimized probe generation and hybridization conditions such that changes in very low abundance transcript levels can be detected. These conditions do lead to oversaturation of many high abundance microarray spots. We reasoned, however, that SKN-1 targets involved in early endodermal and/or mesodermal specification would generate low-abundance transcripts (e.g., transcription factors, signaling molecules, and the like), as opposed to later high-abundance lineage transcripts, for example, those encoding structural proteins. In fact, by using these conditions, we were able to correctly identify changes in very low abundance transcript profiles (e.g., med-1, T24D3.1 and tbx-35, ZK177.10) that other more standard microarray hybridizations missed (Baugh et al., 2003).

We compared our data to that generated from a previously reported microarray analysis using Affymetrix arrays (Baugh et al., 2003). Among the 275 12-cell zygotic genes that we identified and the 10 genes that we assayed in vivo using promoter GFP reporters, only 99 and 5, respectively, were reported as being detected consistently at a similar stage in the early C. elegans embryo. There are several factors that could explain these differences. First, as we have pointed out, our hybridizations have been optimized for the detection of low-abundance transcripts, and very high abundance transcripts have been excluded from our analysis. Second, we have excluded those genes with readily detectable transcript levels at the 4-cell stage. Third, our analysis has been conducted at the level of single embryos, and as no pooling has been required, we may have been more successful at identifying transcripts up-regulated within the very short (approximately 30 min) developmental window from the 4-cell to the 12-cell embryo. Finally, as we have noted, our single embryo cDNAs are enriched in 3' sequence relative to 5' end sequence. Consequently, hybridization signals are weak or undetectable if the genomic PCR fragment spotted on the slide derives from the far 5' end of a large gene.

We observed no significant variation in zygotic transcript representation across the C. elegans genome, including the X chromosome, with the exception of an approximately two- to threefold elevation in sequences deriving from the second chromosome. The significance of this result, if any, eludes us at the moment.

Classes of early zygotic genes

Consistent with our intended focus on relatively low abundance transcripts, genes implicated in either cell-to-cell signaling or transcriptional regulation comprise over one third of the 12-cell zygotic and sdz gene sets (Table 1). For the sdz gene set, 18% encode proteins containing signal peptides and/or transmembrane domains, 10% encode predicted transcription factors, and 6% encode predicted DNA or RNA binding proteins. The sdz gene set, which includes end-1 and end-3, is likely to include other key lineage regulators and will be the initial focus of further functional analysis. In early embryos, many signaling pathways regulating blastomere identity have been identified through maternal-effect lethal screens. However, the zygotic components of these signaling pathways remain completely unknown. Identification of lineage-specific secreted or membrane proteins are likely to reveal additional players in established or previously unidentified signaling pathways. Intriguingly, sdz-31, a predicted membrane protein, is expressed exclusively in the EMS lineage (Fig. 5).

The two putative transcription factors tbx-35 (T-box) and sdz-38 (C2H2 zinc finger) subjected to further expression analysis were found to be expressed exclusively in the MS lineage. This is highly significant, as this constitutes the first report of MS-restricted gene expression in the early C. elegans embryo. Given that both genes encode transcription factors, it also suggests that this approach will succeed in identifying factors, like the GATA factors END-1 and END-3 for the E/endoderm lineage, that control specification and development of other lineages. In this particular case, the results open a new avenue for analysis of early mesoderm specification in C. elegans.

One striking feature of the zygotic gene set identified here is the proportion of genes encoding proteins predicted to function in protein degradation. While a large number of maternally expressed regulatory proteins are degraded at around the 12-cell embryo stage, the machinery and mechanisms involved in their degradation are mostly unknown. As the degradation of key maternally expressed regulators coincides with the onset of early zygotic expression, it is an intriguing possibility that degradation of many maternal proteins is regulated by the synthesis of certain zygotic components. It has become increasingly apparent that the correct spatiotemporal regulation of maternal protein degradation is important for both the earliest events in the 1- and 2-cell embryo immediately following fertilization as well as for later events in early embryogenesis (DeRenzo et al., 2003; Lin, 2003; Pelletier et al., 2003). Our identification of zygotically expressed genes potentially involved in protein degradation that are
up-regulated from the 4- to 12-cell stage embryo should help our understanding of this developmentally critical process. It is also noteworthy that approximately 20% of the sdz gene set is also implicated in protein degradation, suggesting that this process probably plays an integral role in lineage specification by this maternal transcription factor and probably other maternal factors as well.

Our analyses also uncovered a large number of early zygotic and sdz genes that either do not contain any identifiable domain or contain domains of unknown function (22% and 12%, respectively, for the sdz gene set). Although affording no indication as to function, this result does suggest that this very focused molecular approach has identified a number of novel genes with probable roles in the early embryo.

In addition, our analysis uncovered genes that would not otherwise be suspected of being lineage specific in the very early embryo. For example, nit-1, encoding a putative nitrilase (carbon–nitrogen hydrolase) involved in the reduction of organic nitrogen compounds, would not be suspected of playing an important role in early embryonic development. However, nit-1 expression begins at the 8- to 12-cell stage and is found exclusively in the EMS lineage up until approximately the 300-cell stage, when its expression becomes progressively restricted to the E lineage. It is intriguing that a gene of this class is expressed so early and in such a lineage-specific manner, which might suggest a separate early function. This unbiased sequence-based approach should allow the identification of genes based solely on their stage- and lineage-specific expression patterns, in spite of what might at first glance appear an unlikely or unexpected molecular identity.

Redundancy

As previously pointed out, one major problem hindering a genetic analysis of early embryo development is the high degree of functional genetic redundancy. Our microarray data support that assessment. Several genes and their homologs appear to have similar expression profiles in the wild-type 12-cell samples and both are coordinately either down-regulated or unchanged in the skn-1 mutant embryos. The potential exists in any hybridization-based screen that genes that are closely related at the sequence level will be detected by cross-hybridization, which would generate false positives. However, in only a small percentage of cases within the 12-cell zygotic or sdz gene sets are closest sequence homologs detected. Although cross-hybridization cannot be discounted in all these cases, it does suggest that this protocol detects genes primarily based upon expression in the 12-cell stage embryo. By way of demonstration of this principle, we focused on the only two out of 20 CCCH genes known to reside in the C. elegans genome that were among the 275 12-cell zygotic gene set. These two CCCH genes, F38C2.5 and Y57G11C.25, are closest sequence homologs within the C. elegans genome. While depletion of either gene transcript alone does not reveal any phenotype, depletion of both genes simultaneously by RNAi results in 100% embryonic lethality. Although a functional redundancy may have been surmised in this case based upon sequence relationship, it is our contention that this and similar analyses will be able to identify other functionally redundant genes which, although belonging to the same gene family, are not closest sequence homologs or conceivably may even be genes from different gene families. Efforts are ongoing in the lab to systematically analyze additional potential functional redundancies.

Limitations and ways to improve

The DNA microarray used for these studies encompasses 17,815 genes (approximately 94% of the transcribed genome), with each gene represented by an approximately 1 kb PCR fragment derived from genomic DNA with primers selected to maximize exon coverage and minimize cross hybridization. Therefore, PCR fragments are fairly randomly distributed, with no biasing towards one end of the gene or the other (Kim et al., 2001; Reinke et al., 2000). The single embryo cDNAs were generated with an oligo-dT-based primer and, although averaging approximately 1.0–1.5 kb in length, are nonetheless biased towards the 3′ end of the gene. Although C. elegans genes are generally small with small introns, any large gene whose approximately 1 kb genomic sequence on the microarray represents mostly 5′ exons may not be detected in our analysis. In order to be more exhaustive in identifying all candidate genes, an Affymetrix oligonucleotide-based array might be preferred in future analysis (Hill et al., 2000).

Acknowledgments

The authors would like to thank Susan Mango for the plasmid pAP10, Morris Maduro and Joel Rothman for providing med-1, end-1, and end-3 cDNAs, Geraldine Seydoux for pes-10 cDNA, Keith Blackwell for skn-1 cDNA, Stuart Kim and the Stanford Microarray Facility for the microarray hybridizations, Norman Iscove for discussion of experimental details, Jim Waddle for the custom program os4D, and Manisha Patel for technical assistance. We are also grateful to the Japanese knockout consortium National BioResource Project (http://www.nbrp.jp/index.jsp), especially Shohei Mitani, which provided the tbx-11(tm278) mutant allele for analyses. All other strains used in this study were provided by the C. elegans Genome Consortium. This work is in part supported by grants from March of Dimes Birth Defect Foundation (1-FY01-265), American Heart Association (0455081Y) and NIH (HD37933) to R.L.
References


