Development 130, 889-900

doi:10.1242/dev.00302

Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome

L. Ryan Baugh[1], Andrew A. Hill[2], Donna K. Slonim[2], Eugene L. Brown[2] and Craig P. Hunter[1,*]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA
[2]Department of Genomics, Wyeth Research, Cambridge, MA 02140, USA
*Author for correspondence (e-mail: hunter@mcb.harvard.edu)

Accepted 18 November 2002

SUMMARY

Temporal profiles of transcript abundance during embryonic development were obtained by whole-genome expression analysis from precisely staged *C. elegans* embryos. The result is a highly resolved time course that commences with the zygote and extends into mid-gastrulation, spanning the transition from maternal to embryonic control of development and including the presumptive specification of most major cell fates. Transcripts for nearly half (8890) of the predicted open reading frames are detected and expression levels for the majority of them (>70%) change over time. The transcriptome is stable up to the four-cell stage where it begins rapidly changing until the rate of change plateaus before gastrulation. At gastrulation temporal patterns of maternal degradation and embryonic expression intersect indicating a mid-blastula transition from maternal to embryonic control of development. In addition, we find that embryonic genes tend to be expressed transiently on a time scale consistent with developmental decisions being made with each cell cycle. Furthermore, overall rates of synthesis and degradation are matched such that the transcriptome maintains a steady-state frequency distribution. Finally, a versatile analytical platform based on cluster analysis and developmental classification of genes is provided.

Supplemental data and methods available on-line

Key words: Genomics, Time series, Embryogenesis, mRNA amplification, Mid-blastula transition, Zygotic transcription, Microarray, *C. elegans*

INTRODUCTION

Molecular and genetic analysis has elucidated the mechanistic basis of embryogenesis. In addition to identifying and characterizing many key molecules and the processes they control, these analyses have also provided broad insight into embryogenesis, suggesting parameters to consider when thinking about global patterns of gene function and regulation. For example, genetic analysis indicates that most essential genes are pleiotropic (Perrimon et al., 1989; Thaker and Kankel, 1992) but that few function ubiquitously (Bucher and Greenwald, 1991; Ripoll, 1977; Thaker and Kankel, 1992). In addition, kinetic rehybridization (Cot and Rot) analysis has demonstrated that the composition of the embryonic transcriptome changes dramatically as maternal transcripts degrade and embryonic transcripts are synthesized, but that the number of unique transcripts (transcriptome complexity) remains roughly constant during embryogenesis (Davidson, 1986).

However, classic genetic and molecular techniques have limitations. In genetic screens, mutants with partially penetrant or variable phenotypes tend to be overlooked, while functionally redundant genes are missed entirely (Nusslein-Volhard, 1994). Thus, three to five times more genes are expressed during early embryogenesis than the estimated number of embryonic lethal genes (Davidson, 1986). In addition, because Cot and Rot analysis lack gene-specific and temporal information, the time-dependent expression of individual genes has not been characterized in any systematic way. More comprehensive analyses of gene function and expression are needed in order to model embryonic development.

The power of microarrays to quantitatively measure gene expression for the entire genome in parallel is widely appreciated and rapidly being applied to developmental systems. Temporal expression patterns can be resolved by analyzing staged populations of animals (Driessch et al., 2002; Hill et al., 2000; Jiang et al., 2001). Such analysis has been performed in *Drosophila* with dense sampling of time points over the entire life cycle (Arbeitman et al., 2002). In addition, expression analysis can be performed following experimental perturbation to identify tissue-, organ- or lineage-specific genes as well as direct and indirect targets of specific transcription factors (Arbeitman et al., 2002; Furlong et al., 2001; Gaudet and Mango, 2002). Furthermore, with sufficient temporal resolution it should be possible to see developmental processes unfold in the form of transcriptional cascades (Nasiadka and Krause, 1999).

The challenge in such microarray experiments is to translate large amounts of expression data into a deeper and more comprehensive understanding of development. High-throughput reverse genetic techniques will not only aid in this effort but will partially compensate for the limitations of forward genetic analysis by identifying co-expressed genes which may be functionally redundant (Molin et al., 2000). The *C. elegans* embryo, because of its rapid and invariant development (Sulston et al., 1983) and the ease of RNAi (Fire et al., 1998) and transgenic techniques (Fukushige et al., 1999; Mello et al., 1991), is an ideal system in which to pursue a developmental genomic approach.

After fertilization, the *C. elegans* embryo undergoes a series of stereotyped asymmetric cleavages that spatially segregate maternal factors (e.g. transcription factors, transmembrane receptors) with lineage specification activity. This maternal control of lineage identity is thought to result in embryonic expression of lineage-specific genes (Bowerman, 1998), the vast majority of which are unknown. In addition to lineage-based mechanisms, development appears to be controlled through gastrulation by regionalizing and organ-specific activities resulting in a larva with an invariant cell lineage but tissues and organs of polyclonal origin (Labousse and Mango, 1999; Sulston et al., 1983).

Embryonic transcription is first detected in somatic blastomeres at the four-cell stage (Edgar et al., 1994; Hope, 1991; Seydoux and Fire, 1994; Seydoux et al., 1996). However, the first observed developmental phenotype caused by inhibition of embryonic RNA polymerase II activity by RNAi is the absence of the initiation of gastrulation at the 26-cell stage, followed by developmental arrest at about the 100-cell stage (Nance and Priess, 2002; Powell-Coffman et al., 1996). Thus, maternal functions, provided in large part by maternal transcripts, must control much of early embryogenesis. Two classes of maternal transcripts have been described based on their localization patterns in early embryos (Seydoux and Fire, 1994). Class I maternal mRNAs are maintained in all blastomeres, while Class II mRNAs are specifically degraded in somatic blastomeres as early as the two-cell stage and are retained in the germ line precursors. Class I messages appear to encode genes with ubiquitous 'housekeeping' functions, while Class II messages are strongly associated with maternal functions restricted to the early embryo, including specification of embryonic transcription patterns.

Little is known about the complexity and dynamics of gene expression during *C. elegans* embryogenesis. How complexity and composition of the transcriptome change after fertilization and during the transition from maternal to embryonic (lineage-based) control remains uncharacterized. In the absence of sensitive techniques to measure global dynamics of gene expression, no mid-blastula transition has been reported. As with all other embryonic systems, a relatively small number and biased selection of embryonic gene expression patterns have been characterized. Thus, little is known regarding the temporal and spatial complexities of the expression pattern of a typical gene, how many patterns exist and the degree to which expression patterns serve as indicators of function.

In a step towards establishing the *C. elegans* embryo as a developmental genomic system, we describe the results and analysis of a set of wild-type time courses of transcript abundance profiles covering the first 3.5 hours (~1/4) of embryogenesis. Embryos were staged at the morphologically distinct four-cell stage for most of the data reported here. To observe changes in mRNA abundance before the four-cell stage, embryos were also staged at pseudocleavage (one-cell stage). The time course extends into mid-gastrulation ending at the 190-cell stage with typically two time points per cell cycle (12 in total). By the 102-cell stage, more than 70% of the cells in the embryo will contribute exclusively to a single tissue or organ (Labousse and Mango, 1999; Sulston et al., 1983), and by the 190-cell stage, more than 85% of the cells will have all their descendents share the same primary fate (Fig. 1C). This time course should therefore cover most specification events involved in embryonic patterning.

MATERIALS AND METHODS

Methods are briefly described here (see http://dev.biologists.org/supplemental/ for additional details). The complete dataset and analyses are available at www.mcb.harvard.edu/hunter.

Sample preparation

Embryos were collected from cut mothers by mouth pipette and washed thoroughly before being staged by morphology. See www.mcb.harvard.edu/hunter for a detailed protocol of the RNA isolation, amplification and labeling procedures. Briefly, RNA was isolated with TRIzol reagent (Invitrogen) and amplified by two rounds of in vitro transcription as described (Baugh et al., 2001). The estimated 10 million transcripts per embryo is based on bulk measurement of 200 pg total RNA per embryo (data not shown) and the assumptions of 1.5 kb average transcript length and 3.3% polyadenylated mRNA.

Hybridization and data reduction

Microarrays were custom manufactured by Affymetrix (Hill et al., 2000). Amplified biotinylated RNA (1 μg of) was used in each hybridization. Data was normalized and converted to average difference values using the dChip software (β-test version 2001) (Li and Wong, 2001). Average difference values were converted to transcript abundance estimates, in units of parts per million (ppm), by reference to a standard curve of eleven spiked in vitro transcripts as described elsewhere (Hill et al., 2001). Because probe sets can vary by two- or threefold in sensitivity (Hill et al., 2000) and because there may be compositional differences between amplified RNA and in vitro spike-ins (e.g. transcript lengths), transcript abundances should be treated as estimates and intergenic comparisons should be made cautiously.

Absolute decisions (present/absent/marginal calls) were computed by GeneChip 3.1. Sensitivity of each array was defined as the abundance at which each gene on the array had a 70% chance of being called present according to a logistic regression (Hill et al., 2001). The frequency of false-positive present calls for bacterial probesets was 0.015. The corresponding cumulative probability of getting two or more false-positive present calls among three or four replicates was $\sim 10^{-3}$ by binomial statistics.

Data analysis

For plotting gene expression profiles, clustering and phasing, the data were transformed by computing the moving average of means over two time points. Because the purpose of the moving average was to reduce systematic gene-specific differences between series 1 and series 2, PC6 and PC32 (both of which are part of series 2) were not averaged. Moving average transformed data was not used for statistics or developmental classification.

A modified Welch F statistic was used for ANOVA (Zar, 1999). For each gene, regressed error estimates were substituted for observed

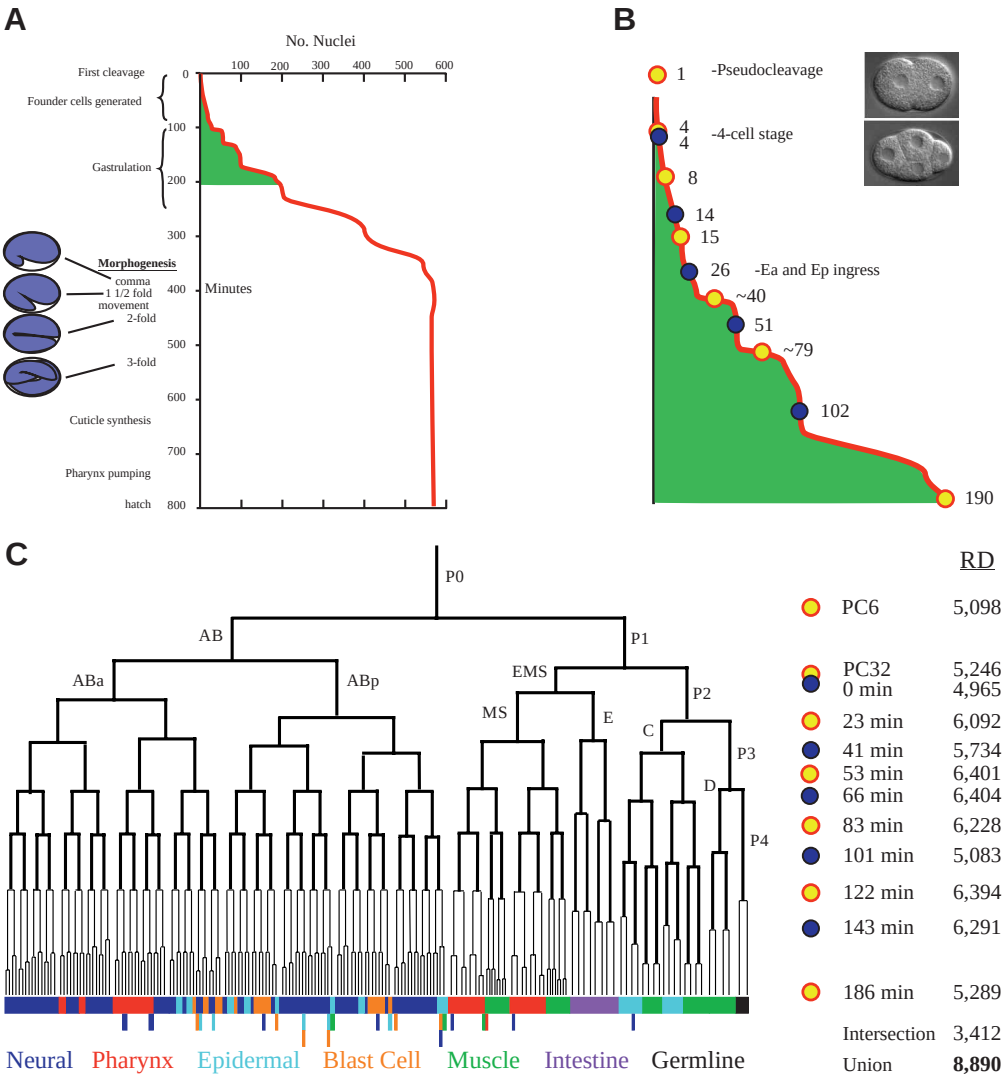


Fig. 1. An embryonic time course of transcript profiles based on precise staging of small cohorts of embryos. (A) Nucleus count versus minutes after the first cleavage at 22°C for all of embryogenesis (adapted, with permission, from Sulston et al., 1983). The area in green indicates the time domain covered by the time course. (B) The time points and the estimated average number of cells per time point in the time course. Embryos were staged at pseudocleavage and the four-cell stage as indicated. Samples for time points in blue and yellow were created, processed and assayed as independent time courses – series 1 and series 2. (C) The complete lineage through the 190-cell stage with the assayed time points indicated. The names given to each time point reflect how the embryos were staged and how long they were aged before being frozen (PC6, pseudocleavage plus 6 minutes; 0 min, four-cell stage plus 0 minutes).

error estimates. The substitution is justified by the lack of consistency among the most and least variable genes at each time point. Regressed error estimates were abundance-dependent pooled error estimates that represented a median error estimate from a window of genes of similar abundance to the gene of interest (see Fig. A at http://dev.biologists.org/supplemental/). A randomization test was used to compute the probability P_g of the observed F statistic for gene g under the null hypothesis that developmental time had no effect on expression. P-values were not corrected for multiple testing.

Clusters were generated by the QT clustering algorithm (Heyer et al., 1999), except that the distance metric used was $1-R_{avg}$, where R_{avg} was the average Pearson correlation coefficient between moving average profiles over 20 realizations of the data plus simulated noise.

Analysis of hypergeometric probability distributions was as described elsewhere (Tavazoie et al., 1999; Zar, 1999), except that depletions were also determined by considering P values near 1. Categories are considered significantly enriched when $P<0.001$ and at least two members of the category are in the group (cluster or class). Depletions are considered significant when $P>0.999$. Three-letter abbreviations correspond to RNAi phenotypes downloaded from WormBase on 5 April 2002 (www.wormbase.org). Chromosomal annotations are from the AceDB version concurrent with design of the arrays. All other annotations are from the Worm Proteome Database and are under one of the designations: ‘functional class’, ‘cellular role’, ‘genetic properties’, or ‘molecular environment’ (www.incyte.com/proteome) (Costanzo et al., 2001). A total of 355 distinct annotations were tested over 106 clusters and 45 developmental classes.

See http://dev.biologists.org/supplemental/ for details on phasing and classification of expression patterns.

RESULTS AND DISCUSSION

An embryonic system for developmental genomics

To generate high-resolution time course data, we amplified RNA from precisely staged and aged cohorts of 10-15 embryos (~2-3 ng total RNA) and hybridized it to whole genome, high-density oligonucleotide arrays (Hill et al., 2000). The arrays assay transcript levels for ~98% of the predicted *C. elegans* ORFs, and have been used previously to demonstrate that the combined RNA amplification and hybridization procedure is both sensitive and representative (Baugh et al., 2001). Initially (series 1) embryos were staged at the morphologically distinct four-cell stage and five time points were collected, each approx. one cell-cycle apart (Fig. 1). To enhance temporal resolution and to verify reproducibility, we assembled an additional time course (series 2) with staggered time points

relative to those in the first. To obtain measurements before the four-cell stage we also staged embryos at pseudocleavage, a transient stage immediately preceding pronuclear fusion. In total, twelve time points (~15-20 min spacing) were collected (Fig. 1), each in triplicate or quadruplicate.

Eleven in vitro synthesized and labeled transcripts were spiked at known concentrations into each hybridization reaction in order to estimate sensitivity and normalize signals between arrays. In addition, these in vitro transcripts were used to assemble a standard curve for each hybridization that allows signal intensity to be converted to transcript abundance reported in parts per million (ppm) (Hill et al., 2001). Sensitivity varied between hybridization reactions such that transcripts present at 5-26 ppm (average=12 ppm) were reliably detected, exact sensitivity depending on the hybridization. Given an estimate of 10 million transcripts per embryo we are able to detect as few as 30 transcripts per embryo, or ~0.2 transcripts per cell at the last time point. Because any two probe sets can vary by as much as two- to threefold in sensitivity (data not shown), transcript abundances should be treated more like estimates than exact measurements when making comparisons between genes. A gene is considered reproducibly detected (RD) when it is called present (see Materials and Methods) in at least two replicates of a given time point.

Given the sensitivity of the assay coupled with the number of replicates and in consideration of estimates of the complexity of embryonic gene expression and mean mRNA concentration (Davidson, 1986), we believe we detected nearly all polyadenylated transcripts present at each time point. The total number of RD transcripts (RD at any time point) was 8890 (Fig. 1C), comparable with measurements in other embryos (Davidson, 1986). However, the number of transcripts expressed simultaneously during embryogenesis appeared to be about 6000. Furthermore, only 3412 genes were RD at all 12 time points, suggesting that a majority of the expressed genes change in abundance during the time course.

As expected, the dynamics of gene expression cause there to be a strong dependence of sensitivity on temporal resolution. One-third of the genes detected in this time course were never called present in a RNA preparation representing all 12 hours of embryogenesis (Hill et al., 2000). In addition, 1084 of the 8890 RD genes are RD at only a single time point. As expected, these transcripts were all very low abundance, even at the time point where they are RD (average=7 ppm). The strong dependence of sensitivity on temporal resolution highlights the value of experimental designs focused on maximizing spatiotemporal resolution.

A primary concern of the experimental design was the possibility that real differences in gene expression would be obscured by excessive variance among replicates, resulting from either biological differences between cohorts of only 10-15 embryos or from staging or other technical issues. Two observations indicate that the observed variance is within acceptable limits. First and most important, adjacent time points are clearly distinct from each other; the average correlation coefficient among replicates is 0.973 compared with an average of 0.935 between adjacent time points ($P=10^{-4}$ by *t*-test). Second, the median coefficient of variation (CV) among replicates per time point for the RD genes is 24±3.5%. Although this CV is roughly twice that of controls where aliquots of the same RNA sample are independently purified, amplified, labeled and hybridized (L. R. B., A. A. H., D. K. S., E. L. B., C. P. H. and K. Hill-Harfe, data not shown) it allows for the reliable detection of less than twofold differences in expression. We do not know whether the additional variance is caused by variation in staging of pools or stochasticity of developmental rates or processes.

Unexpectedly, we also found evidence of a systematic gene-specific effect between the series 1 and series 2 time courses generated on separate occasions (yellow and blue circles in Fig. 1). Because the timepoints for each time course are interspersed, the expression levels of the most severely affected genes (less than 10%) appeared to oscillate with time. It seems

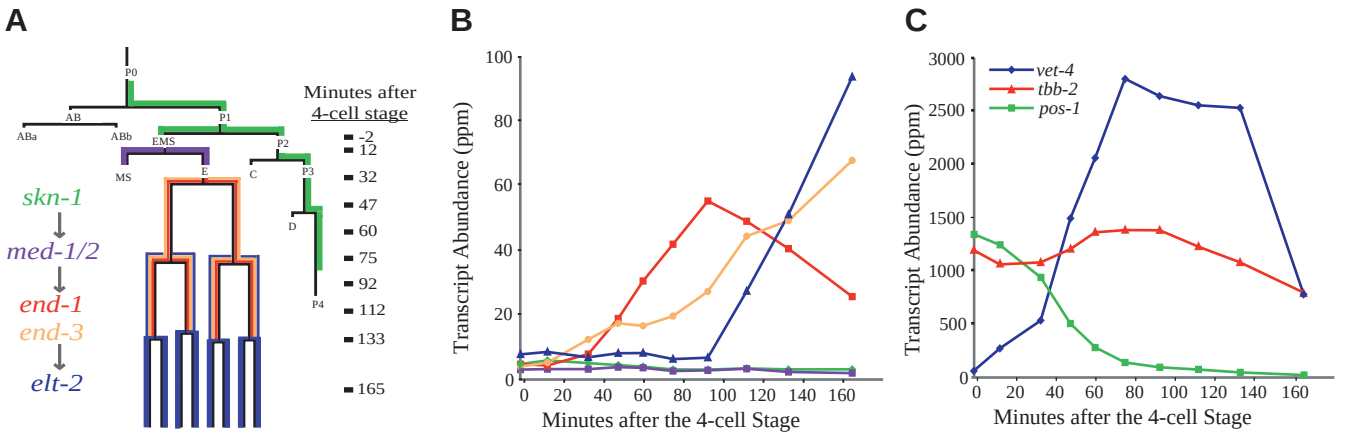


Fig. 2. Genes from a known transcriptional cascade and from three previously characterized expression classes are all detected. (A) Published localization patterns for five transcription factors involved in specification of intestinal fate are depicted on a partial lineage diagram. *skn-1*, *end-1* and *end-3* expression patterns were determined using in situ hybridization; *med-1/2* was determined using a combination of transgenic reporter and RT-PCR; and *elt-2* was determined using antibody. The known regulatory interactions among these genes and proteins are shown along with the moving average time points of gene expression profiles. (B) Gene expression profiles for each of the five genes in A. *med-1* and *med-2* are treated as a single gene as there is only a single probe set on the chip to assay either and it does not distinguish between the two highly similar sequences. Colors correspond to those in A. (C) Gene expression profiles are shown for representatives of each of three previously characterized expression classes: *vet-4*, very early [embryonic] transcripts (vet); *pos-1*, Class I maternal; and *tbb-2*, Class II maternal.

most likely that gene-specific bias was introduced in the amplification and labeling procedure on one occasion relative to the other. To eliminate artifactual differences caused by this bias, all statistical analysis is applied to the two time courses independently. However, to display all data for each gene in a single expression profile we plot the moving average of the data over two adjacent time points, one from each series. This approach assumes that measurements made on one occasion are no more accurate than those made on the other and has the added benefit of dampening biological and assay noise without changing the overall profile. Although more sophisticated time warping algorithms are available (Aach and Church, 2001), given the staggering of the two time courses and the roughly constant spacing of time points, we believe this is the most straightforward means of alignment.

Quantification and temporal resolution of known expression patterns

One goal of this work is to provide a quantitative baseline for future experiments intended to identify components of lineage and cell fate specification pathways. As a benchmark for this goal we examined five components of the well-characterized transcriptional cascade that specifies the E blastomere lineage. In Fig. 2A, we present expression patterns for five genes in this pathway that are derived from published data obtained by independent approaches, including antibody staining, GFP reporters, RT-PCR and in situ hybridization (Bowerman et al., 1993; Fukushige et al., 1998; Maduro et al., 2001; Seydoux and Fire, 1994; Zhu et al., 1997). We detected all five genes at the expected times (Fig. 2B). *skn-1* and *med-1/2* transcript abundances were too low to quantify, but were called present at the expected time points. By contrast, time of induction, rate of increase and maximum expression levels for *end-1*, *end-3* and *elt-2* transcripts were all readily determined. Considering the number of cells each gene is expressed in and our estimate of 10 million transcripts per embryo, transcripts for these three genes are present in excess of 100 copies per expressing cell at or before their time of genetic function. That these genes are so readily detected encourages us that we will be able to identify and resolve the expression patterns of additional genes that specify other lineage-specific cell fates.

To further validate the dataset, we examined the expression profiles of a larger set of genes with known expression patterns. Fig. 2C shows a representative gene of each of three previously characterized expression classes (Schauer and Wood, 1990; Seydoux and Fire, 1994). As expected, *vet-4* is induced very early and increases rapidly. Also as expected, *tbb-2* and *pos-1* are both supplied maternally, and while *tbb-2* remains fairly flat, *pos-1* shows a clear pattern of degradation. Overall, we detected, at the expected times, ten embryonic genes encoding transcription factors with spatially restricted expression patterns (Fig. 3A) (Krause et al., 1990; Ahringer, 1996; Labousse and Mango, 1999; Maduro, 2001; Molin, 2000). However, *vab-7*, which is expressed in only ~4% of embryonic cells, was, like *med-1/2*, detected at too low an abundance to quantify. The remaining embryonic genes in Fig. 3A were all detected at moderate abundance. In addition, we found that known Class I maternal genes tend to be high abundance and remain flat (Fig. 3B) and that Class II maternal genes range from low to high abundance and with the

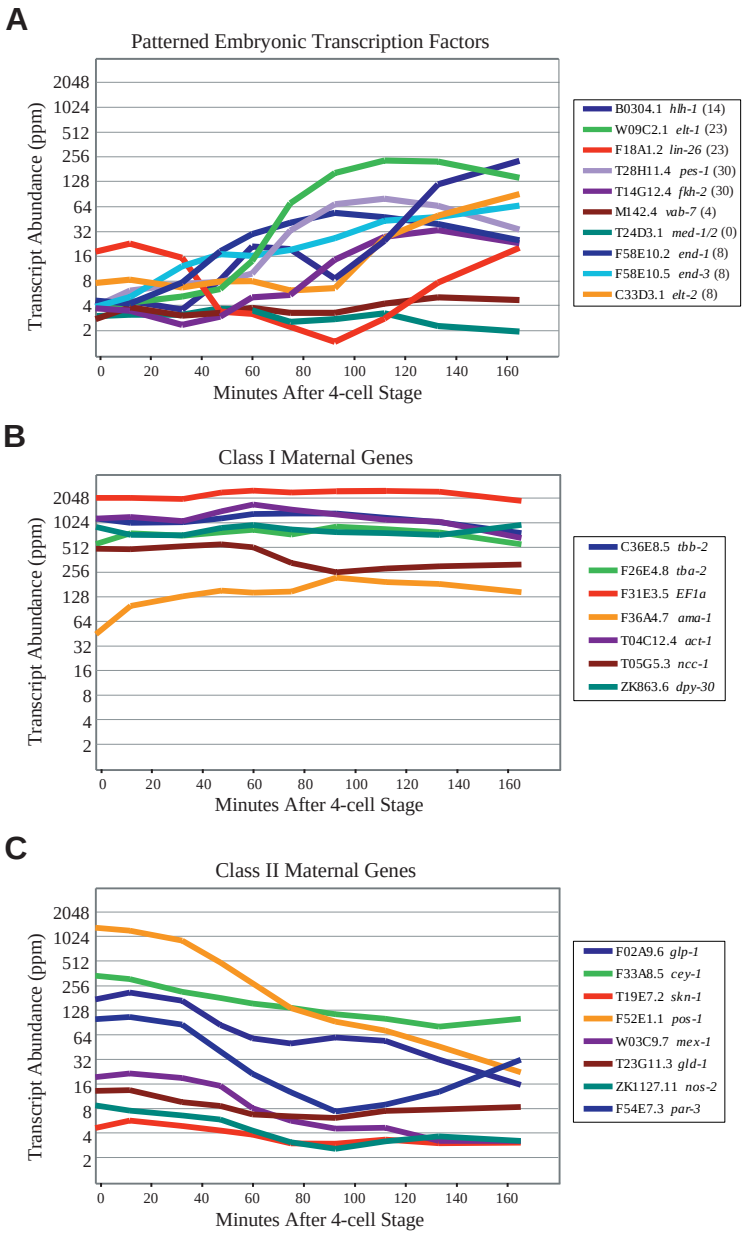


Fig. 3. Twenty-five genes with known expression patterns are detected. (A) Gene expression profiles for 10 embryonic transcription factors characterized by specific developmental phenotypes resulting from disruption of their function. Transcript abundance is plotted on a $\log_2$ scale. The key includes in parentheses the approximate number of cells (out of 102) in which each gene is expressed at 140 minutes. The maternal expression of *lin-26* and early transient induction of *hlh-1* are both consistent with reported expression patterns (Krause et al., 1990; Quintin et al., 2001). (B) Gene expression profiles for seven maternally expressed genes previously characterized as Class I (stable everywhere) by virtue of their in situ hybridization patterns (Seydoux and Fire, 1994). (C) Gene expression profiles for eight maternally expressed genes previously characterized as Class II (degraded in somatic blastomeres, stable in germline precursors) by virtue of their in situ hybridization patterns (Seydoux and Fire, 1994).

exception of *skn-1* all are abundant enough to show a significant decrease over time (Fig. 3C). In summary, all 25 known genes are appropriately detected and with the exception of three low abundance genes the expected expression pattern is readily resolved. This high rate of success provides additional evidence for the comprehensive detection of nearly all expressed genes.

Most genes are modulated in the transition from maternal to embryonic control

An important aspect of development is the transition from maternal to embryonic control. In order to identify genes whose expression is modulated during the transition, we examined changes in gene expression over the entire time course (by within-series ANOVA) and between pairs of time points (by paired-timepoint ANOVA) (see Fig. A at http://dev.biologists.org/supplemental/). We used both types of ANOVA so that we could determine exactly when significant increases and decreases in transcript abundance occurred for temporally modulated genes. The *P*-values for all tests performed can be found at www.mcb.harvard.edu/hunter.

The vast majority of genes expressed during early embryogenesis are temporally modulated. Table 1 shows the number of RD genes that were modulated across the two time series, at three levels of statistical confidence. The minimum of the two series ANOVA *P*-values is considered most relevant, because changes in abundance that occur before the four-cell stage or after the 100-cell stage are not captured in the series 1 analysis. With the most permissive cutoff ($P<0.05$), the smallest fold-change considered significant is ~1.7. At this cutoff, we see that 6963 of 8890 RD genes are significantly modulated (78%), and with a Bonferroni correction for multiple testing it drops to 68% of the RD genes. However, given variable translation rates and protein stabilities, as well as compartmentalization effects, we cannot conclude that a statistically significant change in transcript abundance necessarily correlates with a change in available protein levels. Nevertheless, it is clear that gene regulation is remarkably complex during the transition from maternal to embryonic control.

The fraction of modulated genes seen here is similar to what has been reported for *Drosophila* embryogenesis (Arbeitman et al., 2002) and the analogous unicellular to multicellular transition in *D. discoideum* development (Driessch et al., 2002). In contrast to what is suggested by genetic analysis of *Drosophila* embryogenesis (Lawrence, 1992), genes that define the specification state of cells (e.g. signaling pathway components, transcription factors and co-factors) make up a minority of the modulated genes, while the majority consists of genes encoding sundry biochemical activities not usually thought of as developmentally interesting. This discrepancy probably reflects bias in phenotypic selection for mutants with specific alterations (e.g. cuticle patterns), rather than non-specific lethality and suggests that the importance of transcriptional control of metabolic processes in the early embryo is under appreciated. It will be interesting to investigate the involvement of these genes in developmental processes.

Table 1. Modulation of RD genes

Cut-off	Series 1 ANOVA (five timepoints)	Series 2 ANOVA (seven timepoints)	Union
0.05	6040	6384	6963
0.01	4391	4544	5152
0.001	2455	2718	3157

The majority of expressed genes are temporally modulated. ANOVA was carried out for both experimental series (blue and yellow circles in Fig. 1). The number of genes with *P*-value less than each of three cut-offs is shown with the number of genes in the union of the gene lists from each within-series ANOVA. The null hypothesis is that expression was unchanged across the time course.

The four-cell stage marks a dramatic transition in transcriptome dynamics

The above analysis indicates that mRNA metabolism in the embryo is very dynamic. To investigate the initiation and kinetics of embryonic transcription and depletion of maternal transcripts, we examined the dynamics of transcript abundance on a shorter time scale. For this analysis we examined the difference between adjacent timepoints by 'paired-timepoint' ANOVA. Consistent with expectations from previous work (Edgar et al., 1994; Seydoux and Fire, 1994; Seydoux et al., 1996), relatively few transcripts changed between the one-cell and early four-cell stage (PC6×PC32; Fig. 4). To evaluate the small subset of genes that do show relatively modest increases or decreases in abundance, we asked how many of these genes maintain a trajectory after the four-cell stage, consistent with the change seen up to the four-cell stage. By this criterion ~70% of the 179 decreasing genes ($P<0.01$) appear to continue decreasing after the four-cell stage. By contrast, only eight out of 39 increasing genes (21%) appear to continue increasing (including *ama-1*, *vet-4*, *skr-8* and *skr-9*). Consistent with the expected number of false positives (~89), there are 80 to 90 genes whose overall expression patterns are not consistent with their observed increases or decreases up to the four-cell stage.

The fact that before the four-cell stage more genes decrease than increase suggests that degradation of maternal mRNAs may be either continuing or beginning earlier than embryonic transcription. However, as our measurements rely on the presence of poly A tails, this discrepancy could result from the fact that polyadenylation of new transcripts represents the end of the synthetic process while deadenylation of existing transcripts represents the beginning of the decay process (Wang et al., 2002). In addition, the relatively small number of maternal transcripts that do degrade before the four-cell stage could indirectly reflect the completion of oogenesis, rather than regulation during embryogenesis.

In contrast to transcriptional inhibition in the early embryo, there is no proposed mechanism for the delayed degradation of the vast majority of maternal transcripts (Seydoux et al., 1996). It is possible that early embryonic gene products regulate the timing of degradation, establishing coordination between transcription and degradation. Alternatively, coordinated degradation of maternal transcripts could be an autonomous process that is mediated by a degradation cascade affecting both the transcript and its protein product. Time course data following RNA polymerase II inactivation should distinguish between these possibilities.

The stability of the transcriptome before the 4-cell stage suggests that all embryonic processes occurring up until then are under maternal control. After the four-cell stage, the number of genes changing in transcript abundance increases dramatically through the next two cell cycles until just before

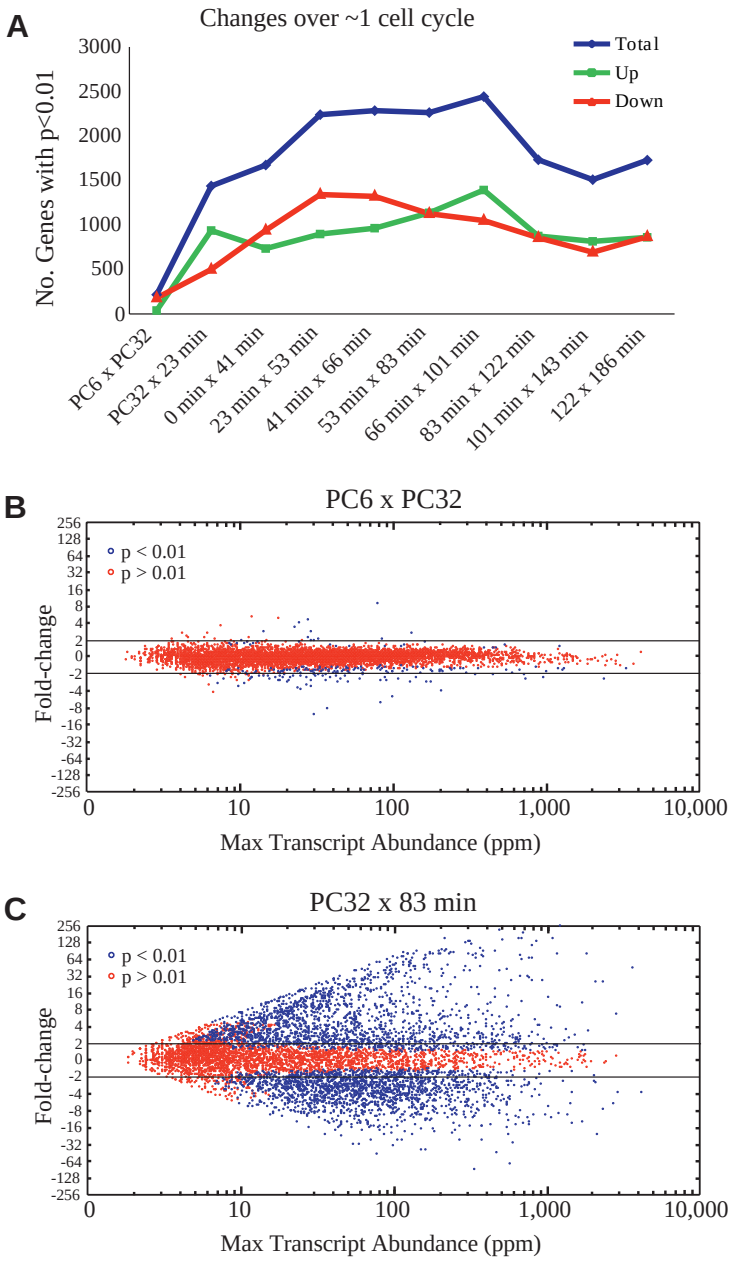


Fig. 4. The transcriptome is stable up to the four-cell stage and changes dramatically thereafter. (A) The *x*-axis shows ten pairs of time points analyzed by paired-timepoint ANOVA. The *y*-axis shows the number of RD genes with $P<0.01$. The number of genes making the cut-off is also split according to whether the change in abundance is positive (Up) or negative (Down). (B) A scatter plot of the 8890 RD genes showing changes in abundance that occur between the PC6 and PC32 time points (one-cell and early four-cell stages, respectively). The max of the two mean transcript abundances is plotted on the *x*-axis on a $\log_{10}$ scale. Fold-change (PC32/PC6) is plotted on the *y*-axis on a $\log_2$ scale. The two lines crossing the *y*-axis at ±2 mark twofold changes. Each point is color coded according to whether or not the observed difference is statistically significant ($P<0.01$) according to a paired-timepoint ANOVA. The number of genes that are considered to be significantly different is 217, 38 of which show an increase and 179 show a decrease. (C) A scatter plot of the 8890 RD genes reflecting changes in transcript abundance that occur between the PC32 and 83 minute time points (early four-cell and ~40-cell stages, respectively). The plot is otherwise identical to B. Of the 3773 genes that are considered significantly different, 1911 show an increase and 1862 show a decrease.

the beginning of gastrulation, where it plateaus and appears to remain roughly constant thereafter, reflecting the onset of embryonic control (Fig. 4A). Furthermore, it appears that after the 26-cell stage the number of genes increasing and decreasing are closely matched (see below).

To examine the transition from maternal to embryonic control of development in detail, we compared an early time point (the four-cell stage) to the 83 minute timepoint (~40-cell stage), the first time point after the initiation of gastrulation (Fig. 4C). In this paired timepoint comparison, over 40% of the RD genes (3773) are significantly modulated ($P<0.01$), again highlighting the extent and magnitude of the transition from maternal to embryonic control of development. Increases in excess of 100-fold are common and many genes go from 'on' to 'off' or vice versa. The diagonal edge of the scatter reflects transcript abundance measurements of genes that were at or below the detection limit in one of the two time points. Many more such gene expression transitions occur among transcripts that increase rather than decrease in abundance, consistent with the embryonic genome assuming control of spatial and temporally restricted developmental processes. That many maternal transcripts do not go to zero may reflect the germline stability of Class II mRNAs or indicate that many maternal transcripts are either stable throughout embryogenesis or synthesized anew in the embryo.

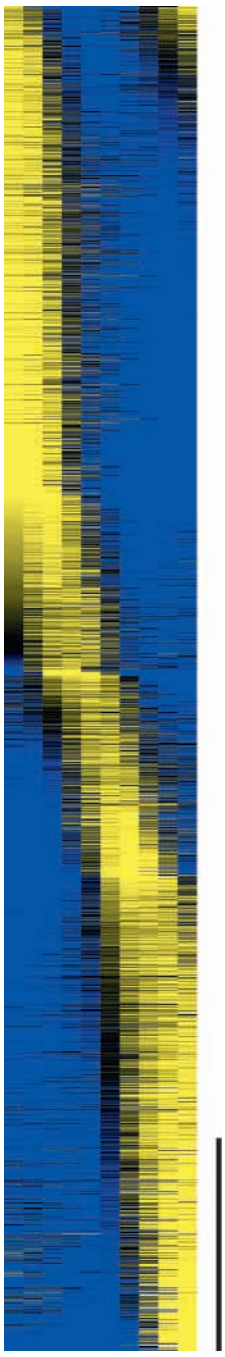


Fig. 5. A phasegram reveals symmetry in the dynamics of the transcriptome, including a wave of roughly constant length. 3157 RD genes with $P<0.001$ in either of the two within-series ANOVAs were sorted according to their time of maximum expression. Columns correspond to moving average timepoints and rows to individual genes. There are roughly two timepoints per cell cycle. Each gene was mean normalized and $\log_2$ transformed. Yellow corresponds to positive values after log transformation (above the mean) and blue corresponds to negative values. Scale bar: 500 genes.

Widespread transient expression suggests developmental decisions are made rapidly

In order to present the expression profiles of the most dynamic genes, in one graph we sorted them by peak expression timepoint (Spellman et al., 1998). The 'phasegram' in Fig. 5 reveals a striking symmetry in the patterns of expression including a wave of genes induced embryonically but transiently. The profile of this wave suggests that the time scale of regulation for the vast majority of dynamic genes is only one cell cycle (increasing over one cell cycle and decreasing over one cell cycle), consistent with cluster analysis and developmental classification of genes (see below, Figs 7, 8). This observation suggests that developmental decisions are made rapidly throughout early embryogenesis, consistent with the observation that there is a narrow temporal window for cell fate transformation by ectopic expression of transcription factors (Gilleard and McGhee, 2001; Quintin et al., 2001; Zhu et al., 1998) and in support of the idea that embryonic regulatory networks achieve a different steady state with each cell cycle (Maduro and Rothman, 2002) accomplishing patterning through a series of binary decisions (Kaletta et al., 1997; Lin et al., 1998). It remains to be determined if this time scale of regulation is constant throughout embryogenesis or if it changes as cell cycles slow down and differentiation commences.

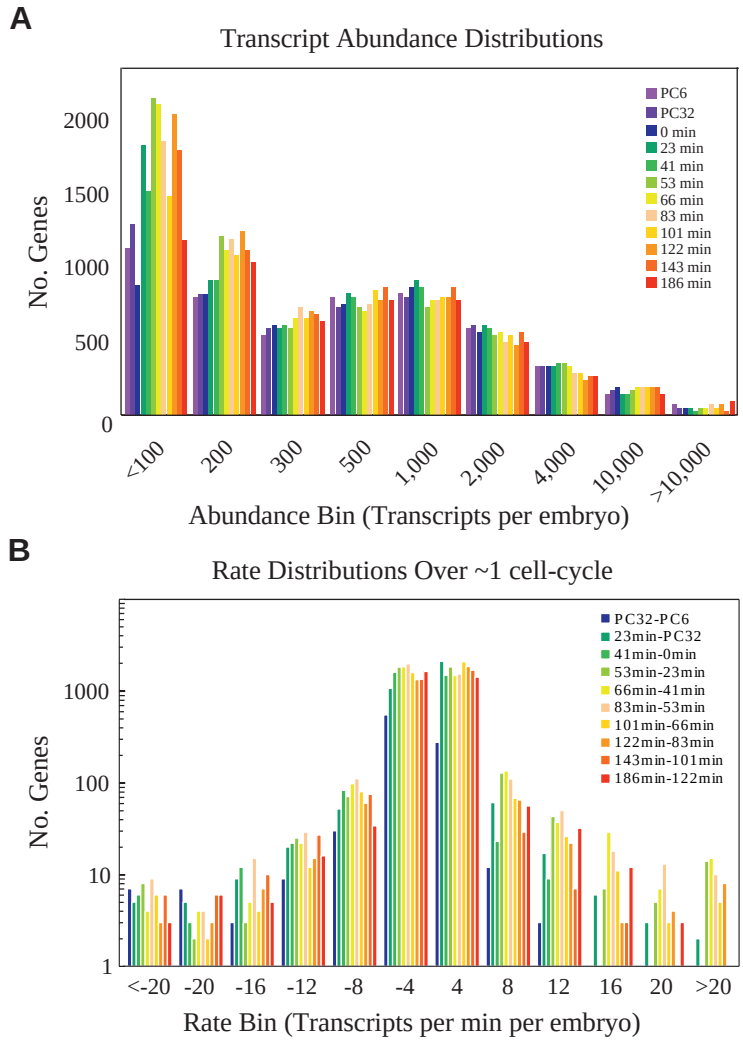


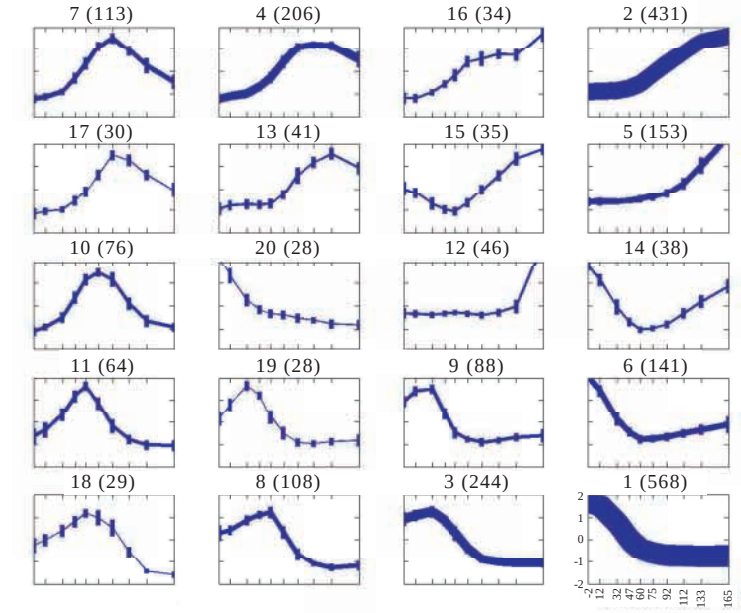
Fig. 6. The transcriptome maintains a steady-state distribution of transcript abundances during early embryogenesis. (A) A histogram plotting the distribution of transcript abundances among the RD genes for each of twelve time points assayed. Binned units along the *x*-axis are transcripts per embryo and the *y*-axis relates how many of the genes RD in that time point are in each bin. (B) A histogram plotting the distribution of rates of change in transcript abundance for each of ten time intervals. Binned units along the *x*-axis are transcripts per minute per embryo and the *y*-axis relates how many genes fall into each bin (note log scale). Time intervals are equivalent to those in Fig. 4A. Rates were calculated by dividing the difference in abundance between timepoints by the corresponding time interval, and converting to transcripts min^{-1} $embryo^{-1}$, assuming 10^7 transcripts per embryo. Only those RD genes with $P<0.05$ in the paired-time point ANOVA corresponding to each time interval are included.

The transcriptome maintains a steady-state frequency distribution

The synthesis, use and turnover of maternal and embryonic transcripts are very different. Maternal transcripts are synthesized well in advance of their use and, at least for Class II transcripts, are rapidly depleted from the embryo. Furthermore, most maternal messages are ubiquitously distributed in the embryo. By contrast, embryonic transcripts are synthesized nearer their time of use and often in only a subset of cells. Therefore we wondered whether the frequency distributions of either maternal or embryonic transcripts would be skewed towards either abundant or rare transcripts. Despite the disparate nature of transcription and degradation during oogenesis and embryogenesis, the distribution of transcript abundances in the early embryo is roughly constant (Fig. 6A).

As the embryo does not grow and the total mRNA content is maintained at an estimated ten million transcripts per embryo, global rates of transcription and degradation must be matched over this time course. We determined the rates of increase or decrease in transcript abundance for statistically significant changes in abundance over short time intervals (~one cell cycle). The number of transcripts increasing or decreasing in abundance at each estimated rate is nearly the same in each time window, with the exception of the earliest window, as expected (see Fig. 4 and accompanying discussion). This asymmetry appears to persist in that there are relatively few high velocity increases over the early time windows until about 66 minutes (26-cell stage). As expected given a constant frequency distribution (Fig. 6A), the distributions of rates are otherwise essentially symmetric (Fig. 6B). Were the rates not matched, then the frequency distribution of the transcripts would change with developmental time, indicating that one of the two processes may be rate limiting. The fact that the rates are matched suggests that the two processes may be functionally coupled, begging the question of whether such a steady state is a universal property of developmental systems or if it is a peculiarity of assaying whole embryos or embryos that do not grow. Similar analyses in other systems with embryos that either do (e.g. vertebrates) or do not (e.g. flies) grow, or that follow defined cell lineages (e.g. hematopoiesis) will help to distinguish between these possibilities.

A



B

Cluster	Coordinate	No. Genes	Enrichments
1	5, 4	568	Cytokinesis
			Meiosis
			Nuclear_cytoplasmic_transport
			Protein_degradation
			Proteasome_subunit
			Embryonic_partitioning_is_defective
			Nuclear_pore
			Chromosome_IV
3	5, 3	244	Mei
5	2, 4	153	Plasma_membrane
			Chromosome_X
6	4, 4	141	
7	1, 1	113	Protein_conjugation_factor
8	5, 2	108	Chromosome_III
9	4, 3	88	Protein_phosphatase
10	3, 1	76	Methylation
			Led
			Chromosome_II
12	3, 3	46	Unknown
			Actin_cytoskeleton_associated
			Chromosome_X
13	2, 2	41	Abnormal_male_specific_structures
			Cell_fate_lineage_defects
			Bmd
			Chromosome_X
14	3, 4	38	9_introns_
15	2, 3	35	RNA_processing_modification
17	2, 1	30	Protein_modification
			Protein_conjugation_factor
19	4, 2	28	Cell_structure
			Abnormal_vulva

Fig. 7. Cluster analysis of expression profiles reveals the most predominant expression patterns as well as significant associations of gene annotations. (A) Means of the 20 largest of 106 total clusters are presented (including 79% of 3157 genes with $P<0.001$ in either within-series ANOVA). Clusters are numbered according to size (number of member genes in brackets) and arranged so that similar patterns are near each other. Axis labels are included for cluster 1 and are the same throughout. Line width reflects relative cluster size on a log scale. Bars reflect 1 s.d. among the members of the cluster. (B) Significant enrichments and depletions of gene annotations in expression clusters determined by a hypergeometric probability analysis ($P<0.001$). Coordinates for the clusters corresponding to A are given (row, column). Functional categories are from Worm Proteome Database (4 March 2002) (Costanzo, 2001). Abbreviations correspond to RNAi phenotypes from WormBase (Mei, defective meiosis; Led, late embryo defect; Bmd, body morphology defective).

Cluster analysis reveals common patterns of co-regulation

Embryos that develop rapidly from fertilization are expected to be near maximally dependent on maternal control of early events. Therefore, transcripts first expressed in the early embryo are expected to be enriched for spatially and temporally restricted functions (Wieschaus, 1996). Conversely, maternal transcripts that are rapidly cleared from the embryo may encode functions that would interfere with later embryonic processes. These two patterns are readily apparent among clustered expression patterns (Fig. 7A): clusters that contain maternal transcripts that rapidly decay (e.g. clusters 1, 3 and 6) and clusters of genes induced in the embryo (e.g. clusters 2, 4 and 5). These are distinct clusters because of differences in timing rather than gross differences in pattern. Unexpectedly, many genes were found in complex clusters. Transcripts that are detected only transiently are common (clusters 7, 8, 10, 11, 13, 17, 18 and 19). This is an intriguing expression pattern that suggests many embryonic genes perform temporally restricted functions, although the protein products may be substantially more stable than their messages. Multi-component expression patterns are also present; in particular, maternal expression followed by degradation and then embryonic induction (clusters 14 and 15). The distinct components of these expression patterns may reflect distinct functions in maternal and embryonic processes or may reflect the relative stability of the proteins translated from maternal RNA. The full set of 106 clusters includes many smaller clusters representing a variety of very complex expression patterns (see Fig. B at http://dev.biologists.org/supplemental/).

Many of the clusters are enriched and depleted for specific functional classes, indicating that temporal expression patterns can correlate with function (Fig. 7B, see Table A at http://dev.biologists.org/supplemental/). For example, cluster 1 is enriched with genes that function in the earliest developmental processes following fertilization. In addition, genes expressed in the germline tend to be excluded from the X chromosome (Reinke et al., 2000), and we see that X-linked genes are depleted from the maternal clusters 1, 3, 6 and 8. By contrast, clusters with relatively late increases in expression are enriched for X-linked genes (clusters 5, 12 and 13) and some of these same clusters are enriched for genes involved in embryonic patterning and morphogenesis (clusters 12 and 13). This analysis is extended for all 106 clusters in Table A. The power of this analysis is limited by the small fraction of annotated *C. elegans* genes, but will improve as more genes are characterized. However, this limitation does not apply to the identification of common regulatory motifs among co-expressed genes. Preliminary analysis indicates that clustered genes are enriched for putative regulatory motifs in 5′ non-coding regions (A. A. H. and D. K. S., unpublished).

Developmental classification of genes identifies a mid-blastula transition

As a complement to cluster analysis, we have used developmental genetic concepts, such as maternal, embryonic

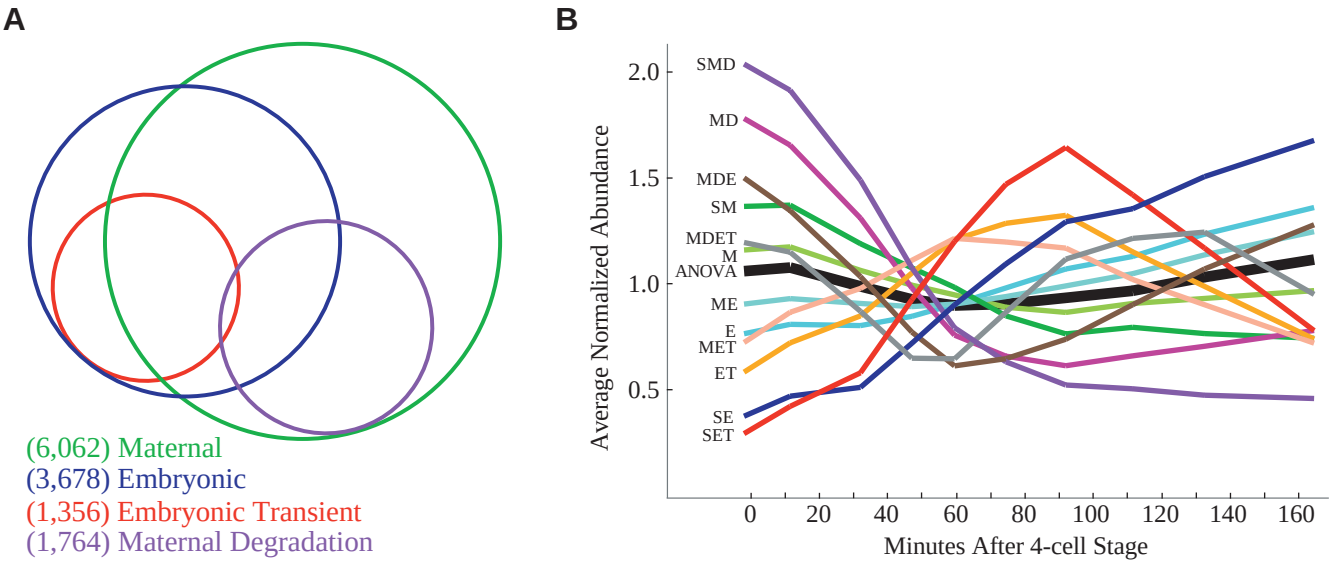


Fig. 8. Defined expression classes based on developmental concepts reveal an inflection point in the transition from maternal to embryonic control. (A) A Venn diagram relating by area the relative sizes and intersections of the four basis classes. The number of genes in each class is in parentheses. (B) Gene expression profiles for the average of each of the defined expression classes. Each gene was mean normalized before computing the class average. The heavy black line labeled ANOVA plots the average of all genes with $P<0.01$ in either of the two within-series ANOVAs, and is included as a point of reference. M, maternal; E, embryonic,; ET, embryonic transient; MD, maternal degradation; S, strictly (e.g. SE implies E but not M; SMD implies MD but not E).

and strictly embryonic to classify genes by expression pattern. For this, we took advantage of the present calls and paired-timepoint ANOVAs to consider when a gene is detected (e.g. PC6 implies maternal expression) and when it shows significant increases and decreases in abundance. The intersections and subdivisions of these basis classes describe overlapping classes with independent descriptors that allow more refined correlations between gene expression patterns and gene annotation to be discerned.

The composition of the transcriptome in terms of the four basis classes is presented in Fig. 8A and Table B (see http://dev.biologists.org/supplemental/). Almost 70% of the RD genes are Maternal (M, 6062), consistent with most embryonic lethal mutations showing maternal effects (Perrimon et al., 1989). Thirty percent of M genes are degraded [maternal degradation (MD); 1764], as are Class II maternal genes, and are likely to be enriched for genes that function to pattern the early embryo. Forty percent of all detected transcripts increase in abundance at some point during embryogenesis [embryonic (E), 3678], indicating zygotic expression. Overlap between M and E is extensive [maternal-embryonic (ME); 2705], indicating the requirement of many genes continuously during the transition from maternal to embryonic control. As a result, strictly embryonic (SE) genes, detected only after the four-cell stage, make up only 11% (973) of the RD genes, consistent with the frequency of 'late' genes in other embryos (Davidson, 1986). In addition, almost 40% of E genes are transient [embryonic transient (ET), 1356], again suggesting that transient gene function is common. As seen in Fig. 5, most of the E genes that are not transient are induced late. Intersections of these classes give smaller classes with multiple descriptors. For example, maternal degradation-embryonic (MDE) has 643 members, suggesting that many genes may have distinct maternal and embryonic functions.

The average expression profile of the 12 expression classes reveals the fundamental expression pattern of each class (Fig. 8B). M genes show a slight decrease over time, even though a decrease is not required in its definition. Interestingly, the class averages intersect 50-60 minutes after the four-cell stage, which coincides with the initiation of gastrulation (~66 minutes). Although embryonic transcription commences earlier, this inflection point in the dynamics of the transcriptome is reminiscent of a mid-blastula transition as it marks a transition between maternal and embryonic control of development. This observation suggests that our understanding of other fundamental embryonic stages that may otherwise be difficult to detect could be improved by analysis of transcriptome dynamics (e.g. phylotypic stage) (Gerhart and Kirschner, 1997).

Classification of genes by time of increase or decrease in abundance is expected to be relevant to their regulation and function. The three dynamic expression classes (MD, E and ET) were therefore subdivided by timing of defining features of the expression profile of each (Fig. C at http://dev.biologists.org/supplemental/). MD subclasses are based on the time of the first significant decrease in abundance, E subclasses are based on time of the first significant increase and ET subclasses are based on time of max abundance. Sizes of each of the 33 subclasses are in Table B (at http://dev.biologists.org/supplemental/).

Enrichments and depletions of gene annotations among the members of all 45 classes and subclasses (see Table C at http://dev.biologists.org/supplemental/) support conclusions from cluster analysis and reveal novel insights. The SE class is enriched for X-linked genes, consistent with the deficiency of X-linked germline genes (Reinke et al., 2000) and cluster analysis (Fig. 7; see above). Furthermore, as expected from the observation that dosage compensation is inactive in the early

embryo (Meyer, 2000), these X-linked SE genes are under represented prior to the 40-cell stage (Table C, http://dev.biologists.org/supplemental/). The subclasses ought to be useful in ongoing informatic analysis: the search for 3′ UTR sequences responsible for different degradation kinetics, hypothesis testing regarding early versus late genes, predicting order of gene function, etc.

We thank Kate Hill-Harfe for helping with a control experiment assessing the reproducibility of the combined RNA isolation and amplification procedure. This work was supported in part by a Beckman Young Investigator Award to C. P. H.

REFERENCES

Aach, J. and Church, G. M. (2001). Aligning gene expression time series with time warping algorithms. *BioInformatics* **17**, 495-508.

Ahringer, J. (1996). Posterior patterning by the *Caenorhabditis elegans* even-skipped homolog *vab-7*. *Genes Dev.* **10**, 1120-1130.

Arbeitman, M. N., Furlong, E. E. M., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W. and White, K. P. (2002). Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270-2275.

Baugh, L. R., Hill, A. A., Brown, E. L. and Hunter, C. P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* **29**, E29.

Bowerman, B. (1998). Maternal control of pattern formation in early *Caenorhabditis elegans* embryos. In *Current Topics in Developmental Biology*, Vol. 39, pp. 73-117. San Diego, CA: Academic Press.

Bowerman, B., Draper, B. W., Mello, C. C. and Priess, J. R. (1993). The maternal gene *skn-1* encodes a protein that is distributed unequally in early *C. elegans* embryos. *Cell* **74**, 443-452.

Bucher, E. A. and Greenwald, I. (1991). A genetic mosaic screen of essential zygotic genes in *Caenorhabditis elegans*. *Genetics* **128**, 281-292.

Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P. et al. (2001). YPD, PombePD and Worm PD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.* **29**, 75-79.

Davidson, E. H. (1986). *Gene Activity in Early Development*. Orlando: Academic Press.

Driessch, N. V., Shaw, C., Katoh, M., Morio, T., Sucgang, R., Ibarra, M., Kuwayama, H., Saito, T., Urushihara, H., Maeda, M. et al. (2002). A transcriptional profile of multicellular development in *Dictyostelium discoideum*. *Development* **129**, 1543-1552.

Edgar, L. G., Wolf, N. and Wood, W. B. (1994). Early transcription in *Caenorhabditis elegans* embryos. *Development* **120**, 443-451.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. A. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.

Fukushige, T., Hawkins, M. G. and McGhee, J. D. (1998). The GATA-Factor *elt-2* Is essential for formation of the *Caenorhabditis elegans* intestine. *Dev. Biol.* **198**, 286-302.

Fukushige, T. H., Hendzel, M. J., Bazett-Jones, D. P. and McGhee, J. D. (1999). Direct visualization of the the *elt-2* gut-specific GATA factor binding to a target promoter inside the living *Caenorhabditis elegans* embryo. *Proc. Natl. Acad. Sci. USA* **96**, 11883-11888.

Furlong, E. E. M., Andersen, E. C., Null, B., White, K. P. and Scott, M. P. (2001). Patterns of gene expression during drosophila mesoderm development. *Science* **293**, 1629-1633.

Gaudet, J. and Mango, S. E. (2002). Regulation of organogenesis by the *C. elegans* FoxA Protein PHA-4. *Science* **295**, 821-825.

Gerhart, J. and Kirschner, M. (1997). *Cells, Embryos, and Evolution: Toward a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability.* Boston: Blackwell Science.

Gilleard, J. S. and McGhee, J. D. (2001). Activation of hypodermal differentiation in the *Caenorhabditis elegans* embryo by GATA transcription factors ELT-1 and ELT-3. *Mol. Cell. Biol.* **21**, 2533-2544.

Heyer, L. J., Kruglyak, S. and Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* **11**, 1106-1115.

Hill, A. A., Hunter, C. P., Tsung, B. T., Tucker-Kellogg, G. and Brown, E. L. (2000). Genomic analysis of gene expression in *C. elegans*. *Science* **290**, 809-812.

Hill, A. A., Brown, E. L., Whitley, M. Z., Tucker-Kellog, G., Hunter, C. P. and Slonim, D. K. (2001). Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. *Genome Biol.* **2**, 0055.1-0055.13.

Hope, I. A. (1991). 'Promoter trapping' in *Caenorhabditis elegans*. *Development* **113**, 388-408.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S. K. (2001). Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **98**, 218-223.

Kaletta, T., Schnabel, H. and Schnabel, R. (1997). Binary specification of the embryonic lineage in *Caenorhabditis elegans*. *Nature* **390**, 294-298.

Krause, M. W., Fire, A., Harrison, S. W., Priess, J. R. and Weintraub, H. (1990). CeMyoD accumulation defines the body wall muscle cell fate during *C. elegans* embryogenesis. *Cell* **63**, 907-919.

Labousse, M. and Mango, S. E. (1999). Patterning the *C. elegans* embryo: moving beyond the cell lineage. *Trends Genet.* **15**, 307-313.

Lawrence, P. A. (1992). *The Making of a Fly: The Genetics of Animal Design*. Oxford: Blackwell Scientific.

Li, C. and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **98**, 31-36.

Lin, R., Hill, R. J. and Priess, J. R. (1998). POP-1 and anterior-posterior fate decisions in *C. elegans* embryos. *Cell* **92**, 229-239.

Maduro, M. F. and Rothman, J. H. (2002). Making worm guts: the gene regulatory network of the *Caenorhabditis elegans* endoderm. *Dev. Biol.* **246**, 68-85.

Maduro, M. F., Meneghini, M. D., Bowerman, B., Broitman-Maduro, G. and Rothman, J. H. (2001). Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3β homolog is mediated by MED-1 and -2 in *C. elegans*. *Mol. Cell* **7**, 475-485.

Mello, C. C., Kramer, J. M., Stinchcomb, J. D. and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **12**, 3959-3970.

Meyer, B. J. (2000). Sex in the worm: counting and compensating X-chromosome dose. *Trends Genet.* **16**, 247-253.

Molin, L. F., Mounsey, A., Aslam, S., Bauer, P. K., Young, J., James, M., Sharma-Oates, A. and Hope, I. A. (2000). Evolutionary conservation of redundancy between a diverged pair of forkhead transcription factor homologues. *Development* **127**, 4825-4835.

Nance, J. and Priess, J. R. (2002). Cell polarity and gastrulation in *C. elegans*. *Development* **129**, 387-397.

Nasiadka, A. and Krause, H. M. (1999). Kinetic analysis of segmentation gene interactions in *Drosophila* embryos. *Development* **126**, 1515-1526.

Nusslein-Volhard, C. (1994). Of flies and fishes. *Science* **266**, 572-574.

Perrimon, N., Engstrom, L. and Mahowald, A. P. (1989). Zygotic lethals with specific maternal effect phenotypes in *Drosophila melanogaster*. I. Loci in the X chromosome. *Genetics* **121**, 333-352.

Powell-Coffman, J. A., Knight, J. and Wood, W. B. (1996). Onset of *C. elegans* gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. *Dev. Biol.* **178**, 472-483.

Quintin, S., Michaux, G., McMahon, L., Gansmuller, A. and Labouesse, M. (2001). The *Caenorhabditis elegans* gene *lin-26* can trigger epithelial differentiation without conferring tissue specificity. *Dev. Biol.* **235**, 410-421.

Reinke, V., Smith, H. E., Nance, J., Wang, J., van Doren, C., Begley, R., Jones, S. J. M., Davis, E. B., Scherer, S., Ward, S. and Kim, S. K. (2000). A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605-616.

Ripoll, P. (1977). Behavior of somatic cells homozygous for zygotic lethals in *Drosophila melanogaster*. *Genetics* **86**, 357-376.

Schauer, I. E. and Wood, W. B. (1990). Early *C. elegans* embryos are transcriptionally active. *Development* **110**, 1303-1317.

Seydoux, G. and Fire, A. (1994). Soma-germline asymmetry in the distributions of embryonic RNAs in *Caenorhabditis elegans*. *Development* **120**, 2823-2834.

Seydoux, G., Mello, C. C., Pettit, J., Wood, W. B., Priess, J. R. and Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of *C. elegans*. *Nature* **382**, 713-716.

Spellman, P. T., Sherlock, G., Zhang, M. O., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes in the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **12**, 3273-3297.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119.

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. and Church, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.* **22**, 281-285.

Thaker, H. M. and Kankel, D. R. (1992). Mosaic analysis gives an estimate of the extent of genomic involvement in the development of the visual system in *Drosophila melanogaster*. *Genetics* **131**, 883-894.

Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D. and Brown, P. O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* **99**, 5860-5865.

Wieschaus, E. (1996). Embryonic transcription and the control of developmental pathways. *Genetics* **142**, 5-10.

Zar, J. H. (1999). *Biostatistical Analysis*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Zhu, J., Hill, R. J., Heid, P. J., Fukuyama, M., Sugimoto, A., Priess, J. R. and Rothman, J. H. (1997). end-1 encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans* embryos. *Genes Dev.* **11**, 2883-2896.

Zhu, J., Fukushige, T., McGhee, J. D. and Rothman, J. H. (1998). Reprogramming of early embryonic blastomeres into endodermal progenitors by a *Caenorhabditis elegans* GATA factor. *Genes Dev.* **12**, 3809-3814.

Supplemental data

Embryo staging and collection

Bristol N2 worms were grown on *E. coli* strain OP50 at 25°C. For the dissection, staging and aging that follow, everything was performed in a climate-controlled room at 22°C. Young gravid hermaphrodites were cut in 100 µl water in a depression well slide. Bleach solution (10 µl) [4:1 NaOCl (6% available chlorine), 0.5 M KOH] was added and worms were triturated by pipet for ~10 seconds before adding 10 µl 20% BSA and triturating for an additional ~10 seconds. For staging embryos at the four-cell stage, one- and two-cell embryos were collected and washed by serial transfer via mouth pipet through a series of five 100 µl drops of water in 1 cm^2 hydrophobic barrier wells on the surface of a microscope slide previously treated with Sigmacote (Sigma). Four-cell and older embryos were set aside, and two-cell embryos were pooled. Embryos were pulled from the pool and put into a new pool as they reached the four-cell stage. The pool of four-cell embryos was expanded in this way for approximately 5 minutes or until any of the oldest members of the pool reached the six-cell stage. The pool was then double-checked and six-cell or abnormal-looking embryos were eliminated. The pool was then transferred to another 100 µl drop of water on a separate slide and a stopwatch was started so that time zero is relatively late in the four-cell stage. After the appropriate amount of time aging, embryos were examined and any that did not appear to have developed normally were eliminated and all others were transferred to the lid of a 0.6 ml eppendorf tube (in 1-3 µl) and frozen in liquid nitrogen. Nuclei counts were made in controls that were staged and aged as above but frozen on poly-L-lysine treated slides then fixed and stained with DAPI, in order to calibrate aging time with the published lineage and to measure temporal dispersion (data not shown).

For staging embryos at pseudocleavage (PC), mothers were cut as above, except only those embryos with a single partial cleavage furrow were collected and for a period of exactly 3 minutes. A stopwatch was started at the end of the 3 minute collection (PC plus 0 minutes), and the embryos were transferred to the first wash of five serial washes. By the fifth wash (PC plus ~4 minutes) true cleavages had resulted in two-cell embryos, while PC

furrows had relaxed resulting in one-cell embryos. One-cell embryos were collected and either frozen at PC plus 6 minutes or aged until PC plus 32 minutes, by which time they had all made it to the early four-cell stage. Because PC is more transient than the four-cell stage, such staging results in smaller cohorts of embryos with less temporal dispersion.

RNA isolation and amplification

For a detailed protocol of the RNA isolation, amplification and labeling procedures please see http://www.mcb.harvard.edu/hunter/. Briefly, RNA was isolated by adding 100 μl TRIzol reagent (Invitrogen), vortexing briefly, pipetting up and down eight to ten times, adding 7 μl water and 1 μl linear polyacrylamide (5 μg/μl; GenElute LPA, Sigma), vortexing for 10 seconds, adding 20 μl $CHCl_3$, vortexing for 30 seconds, spinning at 13,000 for 5 minutes, transferring the aqueous phase, adding 60 μl isopropanol and incubating overnight at –20°C. RNA was pelleted by spinning at 16,000 ***g*** for 25 minutes, washed once with 75% ethanol and resuspended in 4 μl DEPC-treated dH_2O (including 20 ng of the (dT)-T7 primer). In some cases embryo collections were pooled at the TRIzol step in order to obtain either 10 embryos per RNA prep (PC6, PC32, 0 minutes) or 15 (all other samples) in a final volume of 100 μl TRIzol.

mRNA was amplified and labeled as described elsewhere (Baugh, 2001), with the notable exception that 20 ng (dT)-T7 primer was used in a 2 μl reverse transcription reaction in the first round of amplification (as opposed to 10 ng and 1 μl). Amplified RNA was quantified by UV absorbance at 260 nm and analyzed by electrophoresis. Yields for the samples used ranged between 1.7 and 18.7 μg.

Array hybridizations

Hybridizations were performed essentially as described in the Affymetrix Expression Analysis Technical Manual, except 1 μg amplified RNA was used per hybridization. Arrays were stained using the Affymetrix recommended antibody amplification method, and scanned with the Affymetrix GeneChip scanner. Four replicate scans were averaged for each array; this averaging improved the signal-to-noise ratio for the arrays, compared with an average of two scans.

Data reduction

Array images were reduced to probe intensity values and stored in .cel file format using Affymetrix GeneChip 3.1. Data in .cel files were normalized and converted to average difference values using the dChip software (β-test version 2001) (Li and Wong, 2001). Average difference values were converted to transcript abundance estimates, in units of ppm, by reference to a standard curve of 11 spiked in vitro transcripts as described elsewhere (Hill, 2001).

Absolute decisions (present/absent/marginal calls) were computed by GeneChip 3.1. The absolute decision is based on the magnitude of the difference between hybridization intensity and array background and on the fraction of probe pairs with fluorescence above background and noise (see the Affymetrix GeneChip analysis suite user guide for details).

Moving average

For the purposes of plotting gene expression profiles, clustering and phasing, the data were transformed by computing the moving average of means over two time points. Ten averages of adjacent timepoints that were part of distinct series were computed, starting with PC32 (about 4 minutes younger than 0 minutes) and 0 minutes. As a result, the first moving average time point is –2 minutes relative to the four-cell stage and the last time point is 165 minutes (the average of 122 minutes and 186 minutes). The purpose of the moving average was to reduce systematic gene-specific differences between series 1 and series 2. Hence, PC6 and PC32, both part of series 2, were not averaged. Moving average transformed data was not used for statistics or developmental classification.

ANOVA

A modified Welch F statistic (Zar, 1999) was used for all hypothesis testing. Individual replicate data was $\log_e$ transformed as the first step of all statistical analyses. The calculation of the modified Welch F statistic for each gene was as described [see Eqns 10.22-10.27 by Zar (Zar, 1999)], except that 'regressed' error estimates r^2 were substituted for the s^2 error terms in the equations. For each gene, these regressed error estimates were abundance-dependent pooled error estimates that represented a median error

estimate from a window of genes of similar abundance to the gene of interest.

Regressed error estimates were computed as follows. Replicate data (containing K timepoints and G genes) was log-transformed, and the (KG) means u_{kg} and variances s_{kg}^2 (k=1..K, g=1..G) were computed for all genes on a given array design (A, B or C). Regressed error estimates r_{kg}^2 were windowed medians of the observed variances s_{kg}^2, using a window size of W=0.01KG. To reduce computation time, we applied a 'jumping', not a 'running' median. That is, all r_{ij}^2 within the first window were assigned the median of that window, the window was shifted by W and the process repeated. Based on empirical testing of windowed medians to improve the median fit to s_{kg}^2, we applied two constraints to the windowed median estimates to make the fit robust and consistent with a simple two component (additive background + multiplicative sampling error) noise model: (1) r_{ij}^2 was constrained to be a decreasing function of the mean frequency in log space, i.e $r_{ij}^2 \leq r_{kl}^2$, for $u_{ij} > u_{kl}$; (2) as the windowed median simply assigned the median of a window of s_{ij}^2 values to each u_{ij}, a strict functional relationship was not guaranteed by this fit alone. Therefore, in rare cases when the median windows assigned multiple r_{ij}^2 values to a single u_{ij} value, we re-assigned to that mean u_{ij} the largest regressed error r_{ij}^2 that was associated with that mean in the dataset.

A randomization test was used to compute the probability P_g of the observed F statistic for gene *g* under the null hypothesis that developmental time had no effect on expression. As the number of experimental replicates was different at some timepoints on some array designs, each array design (A, B or C) was randomized independently.

The randomization test was carried out as follows. For each array design, the log-transformed (G×N) data matrix was assembled, where G was the number of genes on the array, and N the total number of observations (for example, for the A array design, G=6617 and N=50). For each of the G genes the F statistic was computed, within series 1 (K=5 timepoints), within series 2 (K=7 timepoints), and for each paired-timepoint contrast of interest (K=2 for each contrast). The N timepoint labels were then randomly shuffled, and all F statistics recomputed. The random permutation was repeated N_P=200 times to generate one G×N_P matrix containing the null distribution of F for each of the two within-series ANOVAs, and equivalent

$G \times N_P$ matrices for each paired-timepoint contrast. Each of the G gene-specific F statistics from the observed data were referred to their corresponding null distribution, and the p-value for each gene *g* was computed as:

P_g=(count of $F_{null} \geq F_{obs}$)/GN_P

In the null distribution we included all genes, as opposed to referring each gene to the null distribution arising from random shuffling of the observations of that gene only. Thus, each null distribution contained $G \times N_P \sim 6 \times 10^5$ observations of F. To validate this approach, we examined the null distribution of the F statistic for 22 probesets corresponding to 11 cRNAs spiked into the A array hybridizations at levels from ~3-1000 transcripts per million. The null distribution of F was not correlated with expression level for these spiked messages, i.e. F was pivotal in the sense of Westfall and Young (Westfall and Young, 1993).

Phasing

Moving average transformed data of the 3157 RD genes with $P<0.001$ in either within-series ANOVA was used for phasing. Each gene was normalized by its mean over all ten moving average time points. Normalized abundances were $\log_2$ transformed. The ten moving average time points were subdivided into four time windows: –2 minutes; 12 minutes, 32 minutes and 47 minutes; 60 minutes, 75 minutes and 92 minutes; and 112 minutes, 133 minutes and 165 minutes. A mean value was calculated for each of the four windows and the values were ranked 1-4 for each gene. The genes were then sorted in an iterative, nested fashion. First, they were sorted according to earliest window rank. Genes ranking highest in the earliest window were set aside and the remaining genes were sorted according to second window rank. Again, genes ranking highest in the second window were set aside and the remaining genes were sorted according to third window rank. The process was repeated until four groups of genes had been defined. Each group was sorted again, independent of the other three groups, according to the mean value for the time window preceding the highest rank of the group. For the earliest time window the latest time window was used as the preceding window. This second sort, performed four independent times, makes transitions between and down each of the four groups of genes smooth. Breakpoints marking the boundaries of the four groups are

nevertheless apparent and should not be misinterpreted. The phasegram was plotted using TreeView (Eisen, 1998).

Cluster analysis

We desired a clustering algorithm that is insensitive to experimental noise, does not force all input genes into a cluster, and does not require an a priori determination of the number of output clusters. Clusters were generated by the QT clustering algorithm (Heyer, 1999). The algorithm assembles a series of clusters ordered by size, largest first, with no limit on the number of clusters other than the coherence (cluster diameter) defined a priori (0.7 in our clusters). To ensure robust clusters the distance metric used for clustering was $1\text{-}R_{avg}$, where R_{avg} was the average Pearson correlation coefficient between moving average profiles over 20 realizations of the data plus simulated noise. Noise was generated by a two-component model consisting of an additive Gaussian background with standard deviation 2 ppm, and a multiplicative Gaussian sampling error with s.d.=0.1. Simulated data were floored at 1 ppm.

Expression pattern classification

Paired timepoint ANOVA tests serve as the primary basis for classification, though within-series ANOVA tests as well as present calls in the first time point (PC6) are also considered. A cutoff of $P<0.01$ was used with all statistical tests unless otherwise noted. Paired timepoint tests used include ten spanning roughly one cell cycle (PC6×PC32, PC32×23 minutes, 0×41 minutes, 23×53 minutes, 41×66 minutes, 53×83 minutes, 66×101 minutes, 83×122 minutes, 101×143 minutes, 122×186 minutes) as well as eight more spanning roughly two cell cycles (PC6×23 minutes, PC32×53 minutes, 0×66 minutes, 23×83 minutes, 41×101 minutes, 53×122 minutes, 66×143 minutes, 83×186 minutes). All of these tests are within only one of the two time series, the former consisting of adjacent timepoints within a series and the latter consisting of alternate timepoints within a series. Significant increases and decreases in abundance observed within defined time domains were used to classify genes (see below). Time domains were selected following visual inspection of the clustered data and were defined as follows: ‘maternal degradation’ domain equals PC6 to 83 minutes; ‘embryonic’

domain equals PC6 to 186 minutes, ‘induction following degradation’ domain equals 53-186 minutes.

The definition of each class is as follows: ‘maternal’ genes are called present in at least one of the three PC6 replicates; ‘embryonic’ genes increase significantly during either time course. Specifically, among the genes flagged as dynamic in either of the two within-series ANOVAs, embryonic genes are the subset that also significantly increase in at least two of the eighteen total paired timepoint tests or significantly increase in either the 122×186 minutes or 83×186 minutes comparison. ‘Maternal degradation’ (MD) genes are the subset of maternal genes that decrease without first increasing in abundance. Specifically, among the genes flagged as dynamic in either of the two within-series ANOVAs, MD genes decrease significantly in at least two of the ten total paired timepoint tests, but do not significantly increase (P<0.05) in any paired timepoint test before the earliest significant decrease (P<0.01). ‘Embryonic transient’ genes are the subset of embryonic genes in which the latest significant increase is earlier than their latest significant decrease. ‘Maternal-embryonic’ genes are in the intersection of the maternal and embryonic classes. ‘Maternal degradation-embryonic’ genes are the subset of maternal degradation genes that significantly increase in at least two of the eight total paired timepoint tests in the ‘induction following degradation’ time domain. ‘Maternal-embryonic transient’ genes are in the intersection of the maternal and embryonic transient classes. ‘Maternal degradation-embryonic transient’ genes are in the intersection of the Maternal degradation-embryonic and embryonic transient classes. ‘Strictly maternal’ genes are the subset of maternal genes that are not also classified as embryonic. ‘Strictly embryonic’ genes are the subset of embryonic genes that are not also classified as maternal. ‘Strictly maternal degradation’ genes are the subset of maternal degradation genes that are not also classified as embryonic. ‘Strictly embryonic transient’ genes are the subset of embryonic transient genes that are not also classified as maternal.

Select classes were subclassed by the defining timepoint in the expression profile of each gene; there is no overlap between the subclasses of a particular class. Maternal degradation subclasses are based on the earliest significant decrease (abbreviated ‘pd’ for primary decrease). Embryonic and strictly embryonic subclasses are based on the earliest significant increase

(abbreviated ‘pi’ for primary increase). Embryonic transient subclasses are based on the time of max expression (abbreviated ‘max’).

REFERENCES

Baugh, L. R., Hill, A. A., Brown E. L. and Hunter C. P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* **29**, E29.

Eisen, M., Spellman, P., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.

Hill, A. A., Brown, E. L., Whitley, M. Z., Tucker-Kellog, G., Hunter, C. P. and Slonim, D. K. (2001). Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. *Genome Biol.* **2**, 0055.1-0055.13.

Li, C. and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **98**, 31-36.

Westfall, P. H., and Young, S. S. (1993). *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment*. New York: Wiley.

Zar, J. H. (1999). *Biostatistical Analysis*. Upper Saddle River, NJ: Prentice-Hall.

Table A. Gene annotations of expression clusters

Expression group	Gene class	Class in group	Total in group	Class in background	Fold enrichment	*P*
Enrichments						
Cluster 1	Cytokinesis	2	568	2	5.56	0.0E+00
Cluster 1	Meiosis	10	568	20	2.78	2.0E-04
Cluster 1	Nuclear cytoplasmic transport	4	568	5	4.45	1.9E-04
Cluster 1	Protein degradation	15	568	37	2.25	3.0E-04
Cluster 1	Proteasome subunit	13	568	21	3.44	1.1E-06
Cluster 1	Embryonic partitioning is defective	3	568	3	5.56	0.0E+00
Cluster 1	Nuclear pore	5	568	6	4.63	3.3E-05
Cluster 1	Chromosome IV	115	568	497	1.29	6.0E-04
Cluster 3	Mei	2	87	3	8.54	4.6E-04
Cluster 5	Plasma membrane	5	153	25	4.13	9.7E-04
Cluster 5	Chromosome X	35	153	335	2.16	1.8E-06
Cluster 7	Protein conjugation factor	3	113	8	10.48	9.8E-05
Cluster 8	Chromosome III	31	108	542	1.67	7.9E-04
Cluster 9	Protein phosphatase	3	88	9	11.96	6.4E-05
Cluster 10	Methylation	2	76	8	10.38	6.9E-04
Cluster 10	Led	2	21	8	13.27	3.0E-04
Cluster 10	Chromosome II	26	76	623	1.73	8.3E-04
Cluster 12	Unknown	11	46	186	4.06	8.2E-06
Cluster 12	Unknown	9	46	185	3.34	2.4E-04
Cluster 12	Actin cytoskeleton associated	2	46	10	13.73	3.2E-04
Cluster 12	Chromosome X	14	46	335	2.87	4.2E-05
Cluster 13	Abnormal male specific structure s	2	41	13	11.85	5.3E-04
Cluster 13	Cell fate lineage defects	3	41	26	8.88	3.0E-04
Cluster 13	Bmd	3	15	12	18.58	9.9E-06
Cluster 13	Chromosome X	11	41	335	2.53	7.6E-04
Cluster 14	9 introns	2	38	13	12.78	4.2E-04
Cluster 15	RNA processing modification	2	35	17	10.61	7.6E-04
Cluster 17	Protein modification	3	30	42	7.52	5.8E-04
Cluster 17	Protein conjugation factor	2	30	8	26.31	4.2E-05
Cluster 19	Cell structure	2	28	19	11.87	5.5E-04
Cluster 19	Abnormal vulva	2	28	22	10.25	8.6E-04
Cluster 21	Led	2	9	8	30.97	2.0E-05
Cluster 22	Etv	2	16	8	17.42	1.3E-04
Cluster 23	Plasma membrane	2	25	25	10.10	9.0E-04
Cluster 24	2 introns	2	19	24	13.85	3.5E-04
Cluster 24	7 introns	2	19	23	14.45	3.0E-04
Cluster 28	WT	6	6	798	1.40	1.0E-12
Cluster 37	WT	2	2	798	1.40	5.4E-13
Cluster 41	Lethal larval	2	9	34	20.63	9.2E-05
Cluster 49	WT	3	3	798	1.40	1.4E-12
Cluster 49	Chromosome I	5	6	646	4.07	7.2E-05
Cluster 50	Emb	2	2	256	4.36	3.5E-14
Cluster 52	1 intron	2	6	14	75.17	1.4E-06
Cluster 52	Chromosome V	4	6	510	4.13	5.6E-04
Cluster 53	WT	3	3	798	1.40	1.4E-12
Cluster 55	WT	2	2	798	1.40	5.4E-13
Cluster 60	Protein modification	2	5	42	30.07	2.2E-05
Cluster 62	Signal transduction	2	5	50	25.26	3.7E-05
Cluster 69	WT	3	3	798	1.40	1.4E-12
Cluster 72	WT	2	2	798	1.40	5.4E-13
Cluster 74	Emb	2	2	256	4.36	3.5E-14
Cluster 76	WT	3	3	798	1.40	1.4E-12
Cluster 76	Chromosome I	3	3	646	4.89	0.0E+00
Cluster 81	Chromosome III	3	3	542	5.82	0.0E+00
Cluster 86	WT	2	2	798	1.40	5.4E-13
Cluster 91	Chromosome I	2	2	646	4.89	0.0E+00
Cluster 102	WT	2	2	798	1.40	5.4E-13
Cluster 102	Chromosome I	2	2	646	4.89	0.0E+00
Depletions						
Cluster 1	Chromosome X	23	568	335	0.38	1.00000
Cluster 3	Chromosome X	5	244	335	0.19	1.00000
Cluster 6	Chromosome X	3	141	335	0.20	0.99990
Cluster 8	Chromosome X	1	108	335	0.09	0.99994

Enrichments and depletions of gene annotations in all 106 expression clusters. Enrichments and depletions are determined by a hypergeometric probability analysis and are shown where significant ($P<0.001$). Functional categories are from Worm Proteome Database (www.incyte.com/proteome) (Costanzo, 2001) and three-letter abbreviations correspond to RNAi phenotypes from WormBase. Mei, defective meiosis; Led, late embryo defect; Bmd, body morphology defective; Etv, embryonic terminal arrest variable; Emb, embryonic lethal. For RNAi phenotypes 'Total in group' does not correspond to the cluster size but rather the number of genes in the cluster for which an RNAi assay has been published; for all other annotations 'Total in group' is equivalent to cluster size.

Table B. The number of genes in each developmental class and subclass

Expression class	Number of genes
RD	8890
Maternal	6062
Embryonic	3678
Maternal-embryonic	2705
Strictly maternal	3357
Strictly embryonic	973
Maternal degradation	1764
Embryonic transient	1356
Strictly maternal degradation	953
Strictly embryonic transient	441
Maternal degradation-embryonic	643
Maternal-embryonic transient	915
Maternal degradation-embryonic transient	44
RD not classified	1855
MD pd(PC32)	155
MD pd(23 minutes)	551
MD pd(41 minutes)	367
MD pd(53 minutes)	733
MD pd(66 minutes)	159
E pi(PC32)	31
E pi(23 minutes)	925
E pi(41 minutes)	361
E pi(53 minutes)	888
E pi(66 minutes)	284
E pi(83 minutes)	425
E pi(101 minutes)	335
E pi(122 minutes)	370
E pi(143 minutes)	229
E pi(186 minutes)	277
SE pi(PC32)	8
SE pi(23 minutes)	170
SE pi(41 minutes)	61
SE pi(53 minutes)	342
SE pi(66 minutes)	55
SE pi(83 minutes)	123
SE pi(101 minutes)	97
SE pi(122 minutes)	93
SE pi(143 minutes)	67
SE pi(186 minutes)	53
ET max(23 minutes)	69
ET max(41 minutes)	59
ET max(53 minutes)	260
ET max(66 minutes)	181
ET max(83 minutes)	278
ET max(101 minutes)	216
ET max(122 minutes)	118
ET max(143 minutes)	125

The sizes of 45 developmental classes and subclasses defined by expression pattern. Only RD genes are classified.

Table C. Gene annotations of developmental classes and subclasses

Expression group	Gene class	Class in group	Total in group	Class in background	Fold enrichment	*P*
Enrichments						
Class E pi(PC32)	Methylation	2	31	10	45.39	9.1E-06
Class E pi(23 minutes)	Protein conjugation factor	7	925	12	4.44	2.6E-05
Class E pi(23 minutes)	Protein phosphatase	6	925	13	3.51	5.5E-04
Class E pi(23 minutes)	Germline maintenance is defective	4	925	6	5.07	2.1E-04
Class E pi(23 minutes)	Lethal embryonic	37	925	171	1.65	6.5E-04
Class E pi(23 minutes)	Pvl	11	347	27	2.74	2.0E-04
Class E pi(41 minutes)	Helicase	4	361	16	4.87	9.4E-04
Class E pi(41 minutes)	Abnormal vulva	7	361	36	3.79	3.8E-04
Class E pi(41 minutes)	Oogenesis is defective	6	361	32	3.65	9.8E-04
Class E pi(41 minutes)	Emb	42	133	476	1.55	5.9E-04
Class E pi(41 minutes)	Mul	2	133	4	8.77	7.0E-04
Class E pi(53 minutes)	Chromatin chromosome structure	21	888	47	3.54	8.6E-09
Class E pi(53 minutes)	Nucleotide metabolism	4	888	7	4.53	5.3E-04
Class E pi(53 minutes)	DNA binding protein	23	888	90	2.02	2.3E-04
Class E pi(53 minutes)	Dosage compensation defects	5	888	9	4.40	2.4E-04
Class E pi(53 minutes)	Lethal	4	888	6	5.28	1.7E-04
Class E pi(53 minutes)	DNA associated direct or indirect	40	888	168	1.89	1.7E-05
Class E pi(53 minutes)	Ribosome associated	8	888	20	3.17	3.6E-04
Class E pi(53 minutes)	Glycosylation unknown type	2	888	2	7.92	3.0E-12
Class E pi(53 minutes)	Methylation	6	888	10	4.75	4.2E-05
Class E pi(53 minutes)	Nuclear	51	888	244	1.66	7.6E-05
Class E pi(53 minutes)	Chromosome X	138	888	865	1.26	9.0E-04
Class E pi(66 minutes)	Protein synthesis	7	284	39	4.45	1.3E-04
Class E pi(66 minutes)	RNA splicing	4	284	15	6.61	2.2E-04
Class E pi(66 minutes)	RNA binding protein	10	284	56	4.42	1.1E-05
Class E pi(66 minutes)	Cell death is defective	3	284	12	6.19	1.0E-03
Class E pi(66 minutes)	Ribosome associated	7	284	20	8.67	5.3E-07
Class E pi(66 minutes)	Emb	37	102	476	1.78	4.4E-05
Class E pi(66 minutes)	Etv	4	102	13	7.04	1.4E-04
Class E pi(66 minutes)	Gro	12	102	100	2.74	2.9E-04
Class E pi(66 minutes)	Oth	2	102	5	9.15	7.6E-04
Class E pi(66 minutes)	Ste	14	102	123	2.60	2.0E-04
Class E pi(83 minutes)	Transcription factor	14	425	99	2.34	8.1E-04
Class E pi(83 minutes)	Muscle fibers are abnormal	4	425	12	5.52	4.4E-04
Class E pi(83 minutes)	Paralyzed	4	425	13	5.09	6.7E-04
Class E pi(83 minutes)	Lva	17	137	117	2.47	1.0E-04
Class E pi(101 minutes)	11 introns	5	335	18	5.83	1.3E-04
Class E pi(101 minutes)	Cell fate lineage defects	11	335	44	5.25	5.9E-07
Class E pi(101 minutes)	Spn	2	125	4	9.33	5.8E-04
Class E pi(186 minutes)	Lva	15	101	117	2.96	2.5E-05
Class E	Chromatin chromosome structure	35	3678	47	1.42	5.2E-04
Class E	Pol II transcription	64	3678	92	1.33	2.3E-04
Class E	DNA binding protein	63	3678	90	1.34	1.9E-04
Class E	Guanine nucleotide exchange factor	2	3678	2	1.91	0.0E+00
Class E	Transcription factor	77	3678	99	1.49	3.6E-08
Class E	Cell migration defects	23	3678	26	1.69	1.4E-05
Class E	Dark intestine	2	3678	2	1.91	0.0E+00
Class E	Defective dye filling of amphid phasmid sensory neurons	4	3678	4	1.91	0.0E+00
Class E	Extra cell deaths occur	4	3678	4	1.91	0.0E+00
Class E	Germline maintenance is defective	6	3678	6	1.91	0.0E+00
Class E	Hatching is defective	2	3678	2	1.91	0.0E+00
Class E	Lethal larval	43	3678	55	1.50	1.8E-05
Class E	Pharynx development is defective	2	3678	2	1.91	0.0E+00
Class E	Touch sensation is defective	5	3678	5	1.91	3.3E-13
Class E	XO animals are hermaphrodites	2	3678	2	1.91	0.0E+00
Class E	DNA associated direct or indirect	115	3678	168	1.31	5.8E-06
Class E	Ribosome associated	17	3678	20	1.63	4.1E-04
Class E	C terminal geranylgeranylation	2	3678	2	1.91	0.0E+00
Class E	Glycosyl phosphatidylinositol anchor	4	3678	4	1.91	0.0E+00
Class E	Glycosylation unknown type	2	3678	2	1.91	0.0E+00
Class E	N terminus unmodified	3	3678	3	1.91	0.0E+00
Class E	Basolateral plasma membrane	2	3678	2	1.91	0.0E+00
Class E	Mitochondrial matrix	3	3678	3	1.91	0.0E+00
Class E	Nuclear	160	3678	244	1.25	7.2E-06
Class E	Peroxisome	4	3678	4	1.91	0.0E+00
Class E	Emb	291	1268	476	1.12	3.4E-04
Class E	Lva	81	1268	117	1.27	2.7E-04
Class E	Muv	3	1268	3	1.84	0.0E+00
Class E	Prl	2	1268	2	1.84	0.0E+00
Class E	Pvl	22	1268	27	1.50	7.2E-04
Class E	Chromosome X	499	3678	865	1.10	2.9E-04
Class ET max(41 minutes)	Helicase	2	59	16	14.90	2.9E-04
Class ET max(41 minutes)	Other kinase	2	59	5	47.69	5.5E-06
Class ET max(41 minutes)	Altered fertility	2	59	20	11.92	5.8E-04

Table C. Gene annotations of developmental classes and subclasses

Expression group	Gene class	Class in group	Total in group	Class in background	Fold enrichment	*P*
Class ET max(41 minutes)	Oogenesis is defective	3	59	32	11.18	1.3E-04
Class ET max(53 minutes)	Regulatory subunit	2	260	5	10.82	4.7E-04
Class ET max(53 minutes)	Spd	2	87	4	13.41	2.0E-04
Class ET max(83 minutes)	Chromatin chromosome structure	10	278	47	5.38	1.5E-06
Class ET max(83 minutes)	DNA binding protein	14	278	90	3.94	1.9E-06
Class ET max(83 minutes)	Transcription factor	12	278	99	3.07	1.3E-04
Class ET max(83 minutes)	DNA associated direct or indirect	19	278	168	2.86	9.0E-06
Class ET max(83 minutes)	Methylation	4	278	10	10.12	2.0E-05
Class ET max(83 minutes)	Nuclear	27	278	244	2.80	2.9E-07
Class ET max(101 minutes)	Protein conjugation factor	3	216	12	8.14	3.5E-04
Class ET max(101 minutes)	Defective neuronal development	3	216	15	6.51	9.0E-04
Class ET max(101 minutes)	Touch sensation is defective	2	216	5	13.03	2.7E-04
Class ET max(101 minutes)	Pvl	4	64	27	5.40	6.7E-04
Class ET max(122 minutes)	Abnormal alae	2	118	6	19.87	8.9E-05
Class ET max(122 minutes)	Abnormal vulva	4	118	36	6.62	3.0E-04
Class ET max(122 minutes)	Chromosome X	26	118	865	1.79	8.9E-04
Class ET max(143 minutes)	Transcription factor	7	125	99	3.98	3.5E-04
Class ET max(143 minutes)	Multiple vulva like structures	3	125	12	14.07	4.2E-05
Class ET max(143 minutes)	DNA associated direct or indirect	9	125	168	3.02	7.7E-04
Class ET max(143 minutes)	Spn	2	46	4	25.36	2.8E-05
Class M	Energy generation	15	6062	15	1.16	3.3E-12
Class M	Nuclear cytoplasmic transport	9	6062	9	1.16	5.1E-12
Class M	Other metabolism	4	6062	4	1.16	1.1E-11
Class M	Pol III transcription	2	6062	2	1.16	8.1E-12
Class M	Protein folding	16	6062	16	1.16	4.0E-12
Class M	Protein synthesis	39	6062	39	1.16	9.1E-12
Class M	RNA turnover	7	6062	7	1.16	8.8E-12
Class M	Recombination	7	6062	7	1.16	8.8E-12
Class M	Active transporter secondary	8	6062	8	1.16	6.4E-12
Class M	Chaperones	15	6062	15	1.16	3.3E-12
Class M	Complex assembly protein	9	6062	9	1.16	5.1E-12
Class M	Conserved ATPase domain	4	6062	4	1.16	1.1E-11
Class M	Cyclin	2	6062	2	1.16	8.1E-12
Class M	GTPase activating protein	3	6062	3	1.16	7.0E-12
Class M	Helicase	16	6062	16	1.16	4.0E-12
Class M	Ligase	13	6062	13	1.16	5.8E-12
Class M	Major Facilitator Superfamily	2	6062	2	1.16	8.1E-12
Class M	Other kinase	5	6062	5	1.16	8.1E-12
Class M	Other phosphatase	2	6062	2	1.16	8.1E-12
Class M	Proteasome subunit	25	6062	25	1.16	1.0E-11
Class M	RNA polymerase subunit	4	6062	4	1.16	1.1E-11
Class M	Ribosomal subunit	2	6062	2	1.16	8.1E-12
Class M	Spliceosomal subunit	6	6062	6	1.16	0.0E+00
Class M	Topoisomerase	2	6062	2	1.16	8.1E-12
Class M	Translation factor	13	6062	13	1.16	5.8E-12
Class M	tRNA synthetase	3	6062	3	1.16	7.0E-12
Class M	13 introns	7	6062	7	1.16	8.8E-12
Class M	16 introns	5	6062	5	1.16	8.1E-12
Class M	17 introns	6	6062	6	1.16	0.0E+00
Class M	20 introns	2	6062	2	1.16	8.1E-12
Class M	22 introns	2	6062	2	1.16	8.1E-12
Class M	Abnormal alae	6	6062	6	1.16	0.0E+00
Class M	Abnormal rhythms	7	6062	7	1.16	8.8E-12
Class M	Abnormal rolling	3	6062	3	1.16	7.0E-12
Class M	Constipated	5	6062	5	1.16	8.1E-12
Class M	Constitutive dauer formation	10	6062	10	1.16	0.0E+00
Class M	Dark intestine	2	6062	2	1.16	8.1E-12
Class M	Defective dauer formation	5	6062	5	1.16	8.1E-12
Class M	Defects in neurotransmitter metabolism	2	6062	2	1.16	8.1E-12
Class M	Embryonic partitioning is defective	5	6062	5	1.16	8.1E-12
Class M	Extracellular matrix defects cuticle	3	6062	3	1.16	7.0E-12
Class M	Germline maintenance is defective	6	6062	6	1.16	0.0E+00
Class M	Hermaphrodite germline is feminized	3	6062	3	1.16	7.0E-12
Class M	High incidence of males	9	6062	9	1.16	5.1E-12
Class M	Increased frequency of chromosome nondisjunction	9	6062	9	1.16	5.1E-12
Class M	Increased thermotolerance	2	6062	2	1.16	8.1E-12
Class M	Lethargic	10	6062	10	1.16	0.0E+00
Class M	Long	2	6062	2	1.16	8.1E-12
Class M	Meiosis is defective	21	6062	21	1.16	2.7E-12
Class M	Short	9	6062	9	1.16	5.1E-12
Class M	Protein synthesis factor	11	6062	11	1.16	1.1E-11
Class M	Ribosome associated	20	6062	20	1.16	4.1E-12
Class M	Diphthamide formation	2	6062	2	1.16	8.1E-12
Class M	Glycosylation unknown type	2	6062	2	1.16	8.1E-12
Class M	Lysine methylation	2	6062	2	1.16	8.1E-12

Table C. Gene annotations of developmental classes and subclasses

Expression group	Gene class	Class in group	Total in group	Class in background	Fold enrichment	*P*
Class M	N terminal myristylation	2	6062	2	1.16	8.1E-12
Class M	N terminus unmodified	3	6062	3	1.16	7.0E-12
Class M	O linked glycosylation	2	6062	2	1.16	8.1E-12
Class M	Apical plasma membrane	2	6062	2	1.16	8.1E-12
Class M	Basolateral plasma membrane	2	6062	2	1.16	8.1E-12
Class M	Cell body soma	5	6062	5	1.16	8.1E-12
Class M	Centrosome spindle pole body	15	6062	15	1.16	3.3E-12
Class M	Extracellular matrix cuticle and basement membrane	2	6062	2	1.16	8.1E-12
Class M	Mitochondrial inner membrane	3	6062	3	1.16	7.0E-12
Class M	Mitochondrial matrix	3	6062	3	1.16	7.0E-12
Class M	Nuclear matrix	2	6062	2	1.16	8.1E-12
Class M	Nuclear nucleolus	2	6062	2	1.16	8.1E-12
Class M	Nuclear transport factor	3	6062	3	1.16	7.0E-12
Class M	Abs	3	2090	3	1.12	0.0E+00
Class M	Bmd	25	2090	25	1.12	0.0E+00
Class M	Clr	7	2090	7	1.12	0.0E+00
Class M	Cyk	2	2090	2	1.12	0.0E+00
Class M	Dpy	20	2090	20	1.12	0.0E+00
Class M	Emb	459	2090	476	1.08	3.3E-10
Class M	Etv	13	2090	13	1.12	0.0E+00
Class M	Gro	98	2090	100	1.09	1.7E-04
Class M	Lon	3	2090	3	1.12	0.0E+00
Class M	Mei	5	2090	5	1.12	0.0E+00
Class M	Mul	4	2090	4	1.12	0.0E+00
Class M	Nmo	6	2090	6	1.12	0.0E+00
Class M	Ocs	8	2090	8	1.12	0.0E+00
Class M	Pna	4	2090	4	1.12	0.0E+00
Class M	Pnm	2	2090	2	1.12	0.0E+00
Class M	Sle	4	2090	4	1.12	0.0E+00
Class M	Spd	4	2090	4	1.12	0.0E+00
Class M	Spn	4	2090	4	1.12	0.0E+00
Class M	Chromosome I	1139	6062	1275	1.04	8.5E-05
Class M	Chromosome III	1128	6062	1261	1.04	5.3E-05
Class MD pd(PC32)	Protein translocation	2	155	8	11.35	5.4E-04
Class MD pd(PC32)	Receptor protein translocation	2	155	4	22.69	4.1E-05
Class MD pd(23 minutes)	Lva	24	230	117	2.08	1.0E-04
Class MD pd(23 minutes)	Chromosome III	126	551	1261	1.28	8.9E-04
Class MD pd(41 minutes)	Proteasome subunit	6	367	25	4.60	2.1E-04
Class MD pd(53 minutes)	Proteasome subunit	8	733	25	3.07	6.0E-04
Class MD pd(53 minutes)	Topoisomerase	2	733	2	9.60	0.0E+00
Class MD pd(53 minutes)	Chromosome I	169	733	1275	1.27	1.5E-04
Class MD pd(66 minutes)	Cell stress	4	159	20	8.85	6.5E-05
Class MD pd(66 minutes)	Heat shock protein	2	159	9	9.83	8.6E-04
Class MD pd(66 minutes)	Cytoplasmic	10	159	142	3.12	3.5E-04
Class MD	Isomerase	11	1764	19	2.31	4.9E-04
Class MD	Proteasome subunit	16	1764	25	2.55	7.4E-06
Class MD	Topoisomerase	2	1764	2	3.99	3.7E-12
Class MD	Lva	52	700	117	1.48	2.4E-04
Class MD	Chromosome I	387	1764	1275	1.21	9.4E-07
Class MDE	Carbohydrate metabolism	7	643	16	4.79	3.1E-05
Class MDE	Isomerase	6	643	19	3.46	9.8E-04
Class MDE	Lva	27	254	117	2.12	2.8E-05
Class MDE	Chromosome III	147	643	1261	1.28	3.5E-04
Class ME	Carbohydrate metabolism	12	2705	16	1.95	5.9E-04
Class ME	Chromatin chromosome structure	29	2705	47	1.60	3.6E-04
Class ME	Mitosis	24	2705	37	1.69	3.1E-04
Class ME	Protein synthesis	26	2705	39	1.73	9.3E-05
Class ME	Isomerase	14	2705	19	1.92	3.8E-04
Class ME	Translation factor	10	2705	13	2.00	8.8E-04
Class ME	Abnormal vulva	23	2705	36	1.66	5.5E-04
Class ME	Dark intestine	2	2705	2	2.60	0.0E+00
Class ME	Germline maintenance is defective	6	2705	6	2.60	0.0E+00
Class ME	Lethal larval	37	2705	55	1.75	3.7E-06
Class ME	Ribosome associated	17	2705	20	2.21	2.5E-06
Class ME	Glycosylation unknown type	2	2705	2	2.60	0.0E+00
Class ME	N terminus unmodified	3	2705	3	2.60	0.0E+00
Class ME	Basolateral plasma membrane	2	2705	2	2.60	0.0E+00
Class ME	Mitochondrial matrix	3	2705	3	2.60	0.0E+00
Class ME	Nuclear	118	2705	244	1.26	5.4E-04
Class ME	Emb	274	1025	476	1.31	6.1E-12
Class ME	Lva	76	1025	117	1.48	8.4E-07
Class ME	Lvl	26	1025	38	1.56	6.1E-04
Class ME	Pvl	20	1025	27	1.69	3.4E-04
Class ME	Chromosome I	558	2705	1275	1.14	7.9E-06
Class ME	Chromosome III	543	2705	1261	1.12	9.6E-05

Table C. Gene annotations of developmental classes and subclasses

Expression group	Gene class	Class in group	Total in group	Class in background	Fold enrichment	*P*
Class MET	Protein conjugation factor	6	915	12	3.84	2.7E-04
Class MET	Abnormal alae	4	915	6	5.13	2.0E-04
Class MET	Dark intestine	2	915	2	7.69	0.0E+00
Class MET	Methylation	6	915	10	4.61	5.2E-05
Class MET	Emb	100	353	476	1.39	3.8E-05
Class MET	Pvl	12	353	27	2.94	4.7E-05
Class MET	Chromosome I	212	915	1275	1.28	1.4E-05
Class SE pi(41 minutes)	Chromosome II	25	61	1300	2.22	1.0E-05
Class SE pi(53 minutes)	N linked glycosylation	4	342	16	5.14	7.4E-04
Class SE pi(53 minutes)	WT	68	75	1739	1.22	9.2E-05
Class SE pi(53 minutes)	Chromosome II	87	342	1300	1.38	4.1E-04
Class SE pi(53 minutes)	Chromosome V	87	342	1212	1.48	3.2E-05
Class SE pi(66 minutes)	Pol II transcription	4	55	92	5.56	7.1E-04
Class SE pi(66 minutes)	Transcription factor	5	55	99	6.46	1.1E-04
Class SE pi(83 minutes)	ATP binding cassette	2	123	6	19.07	1.0E-04
Class SE pi(83 minutes)	Chromosome X	30	123	865	1.98	5.7E-05
Class SE pi(101 minutes)	Pol II transcription	6	97	92	4.73	2.6E-04
Class SE pi(101 minutes)	Transcription factor	6	97	99	4.40	4.0E-04
Class SE pi(101 minutes)	Cell fate lineage defects	5	97	44	8.24	2.7E-05
Class SE pi(122 minutes)	Chromosome X	23	93	865	2.01	2.6E-04
Class SE pi(143 minutes)	8 introns	4	67	30	14.00	8.0E-06
Class SE pi(186 minutes)	Oxidoreductase	2	53	23	11.54	6.4E-04
Class SE pi(186 minutes)	Cell migration defects	2	53	26	10.21	9.3E-04
Class SE pi(186 minutes)	Chromosome X	21	53	865	3.22	7.9E-08
Class SE	Pol II transcription	24	973	92	1.89	5.2E-04
Class SE	Transcription factor	30	973	99	2.19	5.4E-06
Class SE	Cell migration defects	10	973	26	2.78	3.5E-04
Class SE	Touch sensation is defective	4	973	5	5.78	5.0E-05
Class SE	N linked glycosylation	7	973	16	3.16	6.0E-04
Class SE	Prl	2	243	2	9.60	0.0E+00
Class SE	WT	213	243	1739	1.18	3.7E-08
Class SE	Chromosome X	170	973	865	1.42	1.5E-07
Class SET	18 introns	2	441	4	7.98	9.3E-04
Class SET	WT	95	109	1739	1.17	3.0E-04
Class SET	Chromosome II	114	441	1300	1.40	2.9E-05
Class SM	Pol III transcription	2	3357	2	2.10	0.0E+00
Class SM	Cyclin	2	3357	2	2.10	0.0E+00
Class SM	GTPase activating protein	3	3357	3	2.10	0.0E+00
Class SM	Major Facilitator Superfamily	2	3357	2	2.10	0.0E+00
Class SM	Proteasome subunit	23	3357	25	1.93	2.5E-07
Class SM	Topoisomerase	2	3357	2	2.10	0.0E+00
Class SM	tRNA synthetase	3	3357	3	2.10	0.0E+00
Class SM	20 introns	2	3357	2	2.10	0.0E+00
Class SM	22 introns	2	3357	2	2.10	0.0E+00
Class SM	Defects in neurotransmitter metabolism	2	3357	2	2.10	0.0E+00
Class SM	Hermaphrodite germline is feminized	3	3357	3	2.10	0.0E+00
Class SM	Lysine methylation	2	3357	2	2.10	0.0E+00
Class SM	O linked glycosylation	2	3357	2	2.10	0.0E+00
Class SM	WT	844	1065	1739	1.06	5.9E-07
Class SMD	Protein degradation	18	953	49	2.71	9.1E-06
Class SMD	Recombination	4	953	7	4.22	7.5E-04
Class SMD	Conserved ATPase domain	3	953	4	5.54	3.3E-04
Class SMD	Proteasome subunit	16	953	25	4.72	5.7E-10
Class SMD	Topoisomerase	2	953	2	7.38	0.0E+00
Class SMD	Cytoplasmic	34	953	142	1.77	2.4E-04
Class SMD	Endoplasmic reticulum	11	953	32	2.54	5.9E-04
Class SMD	Ocs	6	384	8	4.56	2.1E-05
Class SMD	Chromosome I	215	953	1275	1.24	7.8E-05
Depletions						
Class E pi(66 minutes)	WT	55	102	1739	0.72	1.00000
Class E pi(83 minutes)	WT	85	137	1739	0.83	0.99941
Class E pi(101 minutes)	Chromosome II	39	335	1300	0.63	0.99965
Class E	Proteasome subunit	2	3678	25	0.15	1.00000
Class E	WT	895	1268	1739	0.95	1.00000
Class ET max(66 minutes)	Chromosome X	8	181	865	0.36	0.99978
Class M	Transcription factor	69	6062	99	0.81	0.99998
Class M	WT	1526	2090	1739	0.98	1.00000
Class M	Chromosome X	695	6062	865	0.93	1.00000
Class MD pd(23 minutes)	Chromosome X	24	551	865	0.35	1.00000
Class MD pd(53 minutes)	Chromosome X	38	733	865	0.42	1.00000
Class MD pd(66 minutes)	Chromosome X	7	159	865	0.36	0.99946
Class MD	Pol II transcription	6	1764	92	0.26	1.00000
Class MD	Signal transduction	18	1764	152	0.47	0.99997
Class MD	DNA binding protein	9	1764	90	0.40	0.99974
Class MD	Receptor signalling	3	1764	73	0.16	1.00000

Table C. Gene annotations of developmental classes and subclasses

Expression group	Gene class	Class in group	Total in group	Class in background	Fold enrichment	*P*
Class MD	Transcription factor	5	1764	99	0.20	1.00000
Class MD	DNA associated direct or indirect	19	1764	168	0.45	0.99999
Class MD	Nuclear	40	1764	244	0.65	0.99939
Class MD	Chromosome X	96	1764	865	0.44	1.00000
Class MDE	Receptor signalling	0	643	73	0.00	0.99912
Class MDE	Nuclear	9	643	244	0.40	0.99929
Class MDE	Chromosome X	44	643	865	0.56	1.00000
Class ME	Proteasome subunit	2	2705	25	0.21	0.99929
Class ME	WT	682	1025	1739	0.89	1.00000
Class ME	Chromosome V	403	2705	1212	0.86	0.99998
Class SE pi(53 minutes)	Emb	4	75	476	0.26	0.99980
Class SE pi(53 minutes)	Chromosome III	39	342	1261	0.64	0.99952
Class SE pi(83 minutes)	Chromosome III	9	123	1261	0.41	0.99942
Class SE	Emb	17	243	476	0.34	1.00000
Class SE	Chromosome I	136	973	1275	0.77	0.99988
Class SE	Chromosome III	133	973	1261	0.76	0.99992
Class SET	Emb	8	109	476	0.36	0.99988
Class SET	Chromosome III	50	441	1261	0.63	0.99994
Class SM	Pol II transcription	28	3357	92	0.64	0.99948
Class SM	DNA binding protein	27	3357	90	0.63	0.99956
Class SM	Transcription factor	22	3357	99	0.47	1.00000
Class SM	Cell migration defects	3	3357	26	0.24	0.99990
Class SM	Lethal larval	12	3357	55	0.46	0.99994
Class SM	DNA associated direct or indirect	53	3357	168	0.66	0.99999
Class SM	Nuclear	84	3357	244	0.72	0.99999
Class SM	Emb	185	1065	476	0.85	0.99950
Class SM	Lva	36	1065	117	0.67	0.99945
Class SM	Chromosome X	366	3357	865	0.89	0.99962
Class SMD	Pol II transcription	3	953	92	0.24	0.99913
Class SMD	Transcription factor	1	953	99	0.07	0.99999
Class SMD	Lethal larval	0	953	55	0.00	0.99968
Class SMD	Chromosome X	45	953	865	0.38	1.00000

Enrichments and depletions of gene annotations in all 45 expression classes and subclasses. Enrichments are shown first in alphabetical order by class name, followed by depletions in alphabetical order. See Fig. 10 and Materials and Methods for class names and definitions. Enrichments and depletions are determined by a hypergeometric probablility analysis and are shown where significant ($P<0.001$). Functional categories are from Worm Proteome Database and three-letter abbreviations correspond to RNAi phenotypes from WormBase. For RNAi phenotypes 'Total in group' does not correspond to the class size but rather the number of genes in the class for which an RNAi assay has been published.