

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

L. Ryan Baugh¹, Andrew A. Hill², Donna K. Slonim², Eugene L. Brown² and Craig P. Hunter^{1,*}

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

²Department 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 descendants 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

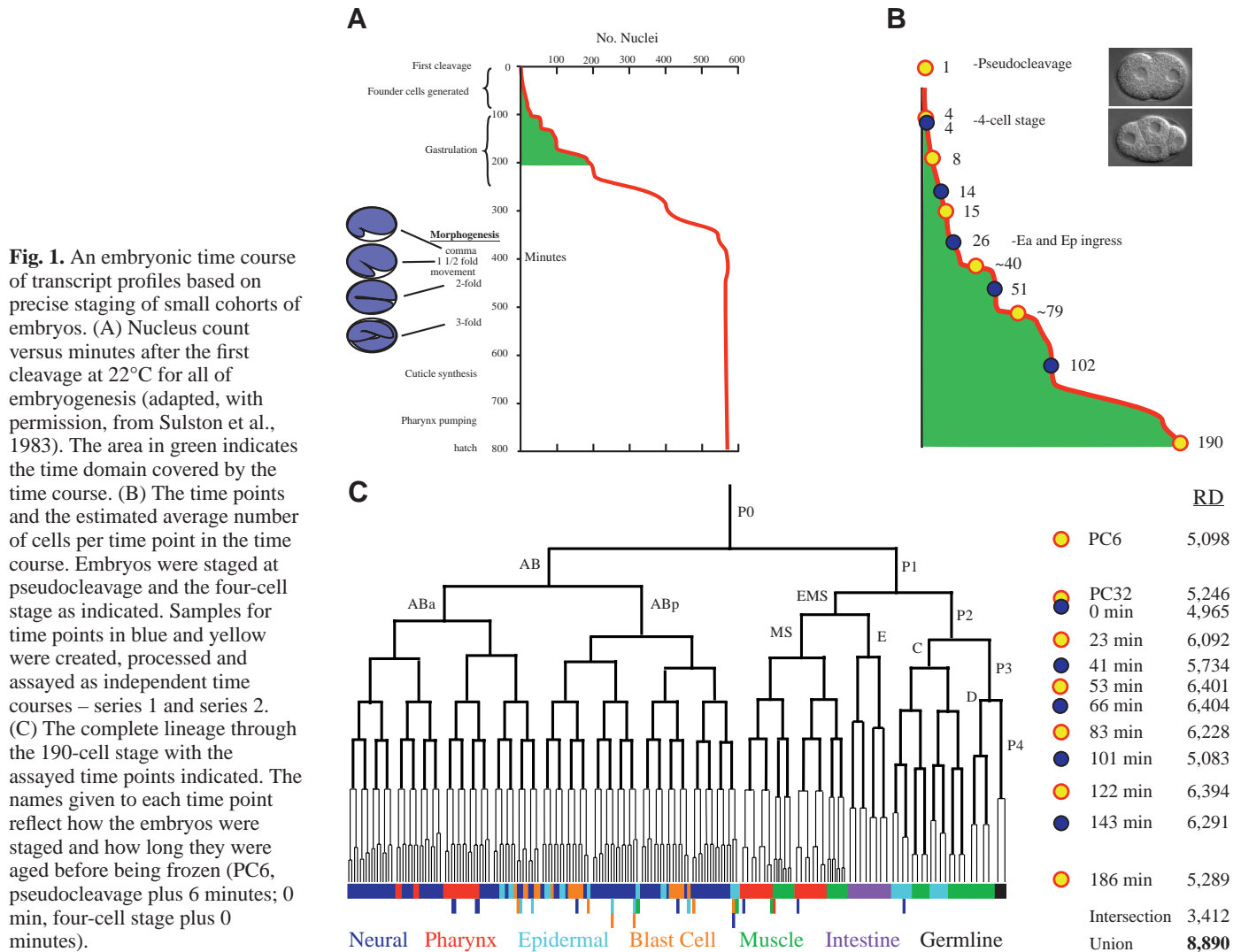
Microarrays were custom manufactured by Affymetrix (Hill et al., 2000). Amplified biotinylated RNA (1 µg of) was used in each hybridization. Data was normalized and converted to average difference values using the dChip software (β-test version 2001) (Li and Wong, 2001). Average difference values were 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



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 $24\pm 3.5\%$. 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

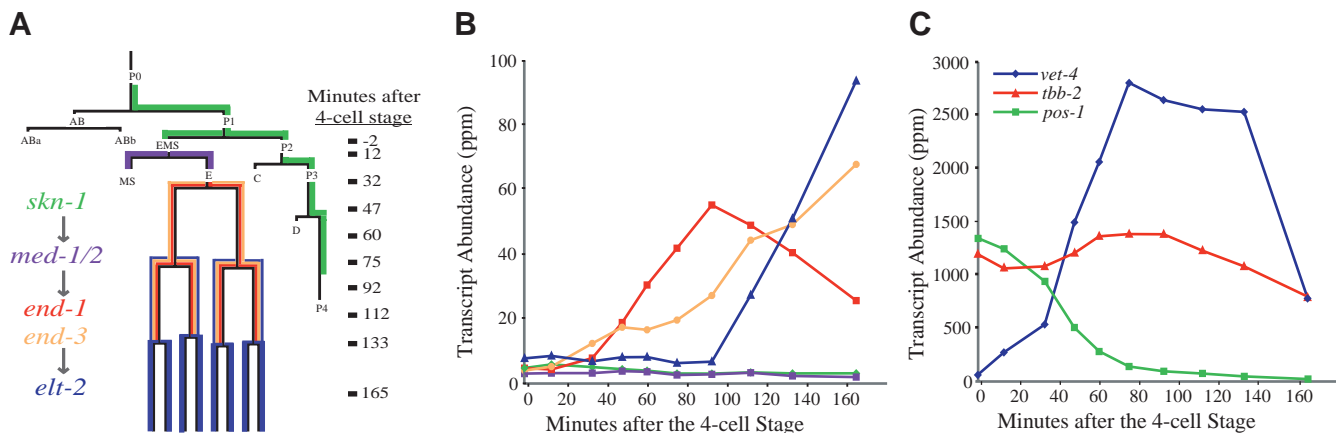


Fig. 2. Genes from a known transcriptional cascade and from three previously characterized expression classes are all detected. (A) Published localization patterns for five transcription factors involved in specification of intestinal fate are depicted on a partial lineage diagram. *skn-1*, *end-1* and *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

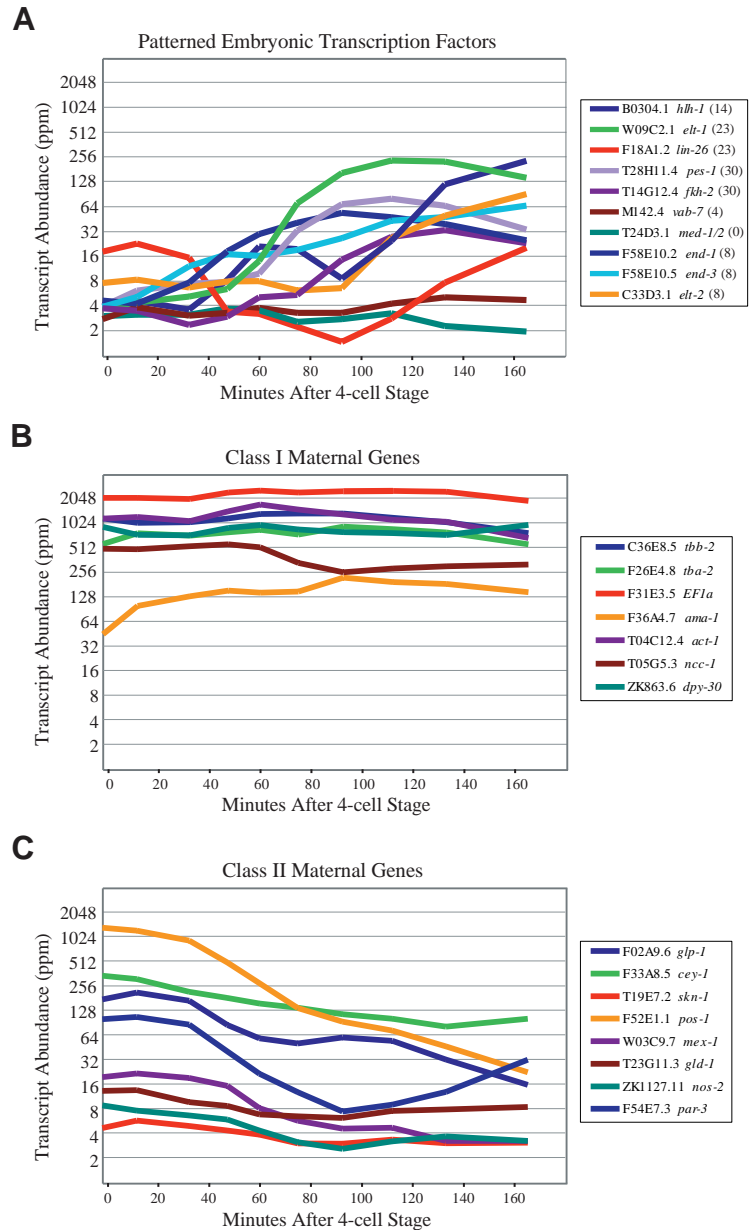


Fig. 3. Twenty-five genes with known expression patterns are detected. (A) Gene expression profiles for 10 embryonic transcription factors characterized by specific developmental phenotypes resulting from disruption of their function. Transcript abundance is plotted on a log₂ scale. The key includes 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).

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

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.

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

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.

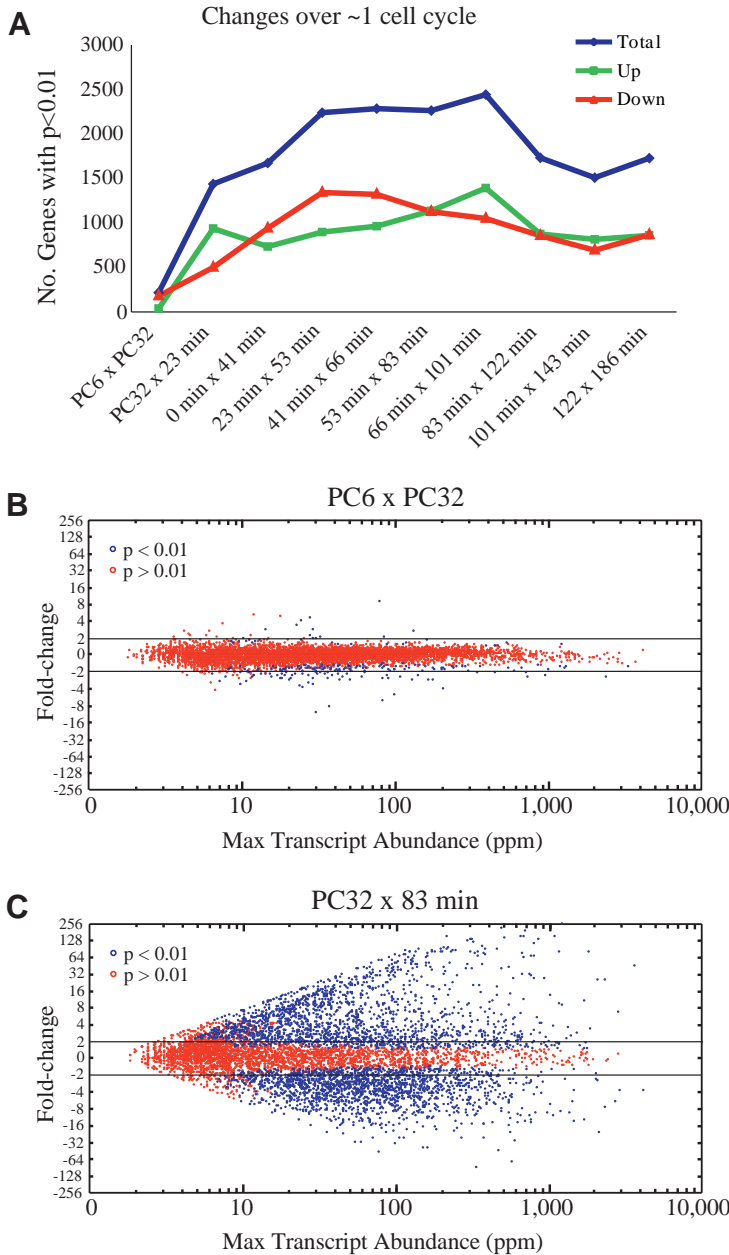


Fig. 4. The transcriptome is stable up to the four-cell stage and changes dramatically thereafter. (A) The x-axis shows ten pairs of time points analyzed by paired-timepoint ANOVA. The y-axis shows the number of RD genes with $P < 0.01$. The number of genes making the cut-off is also split according 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.

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.

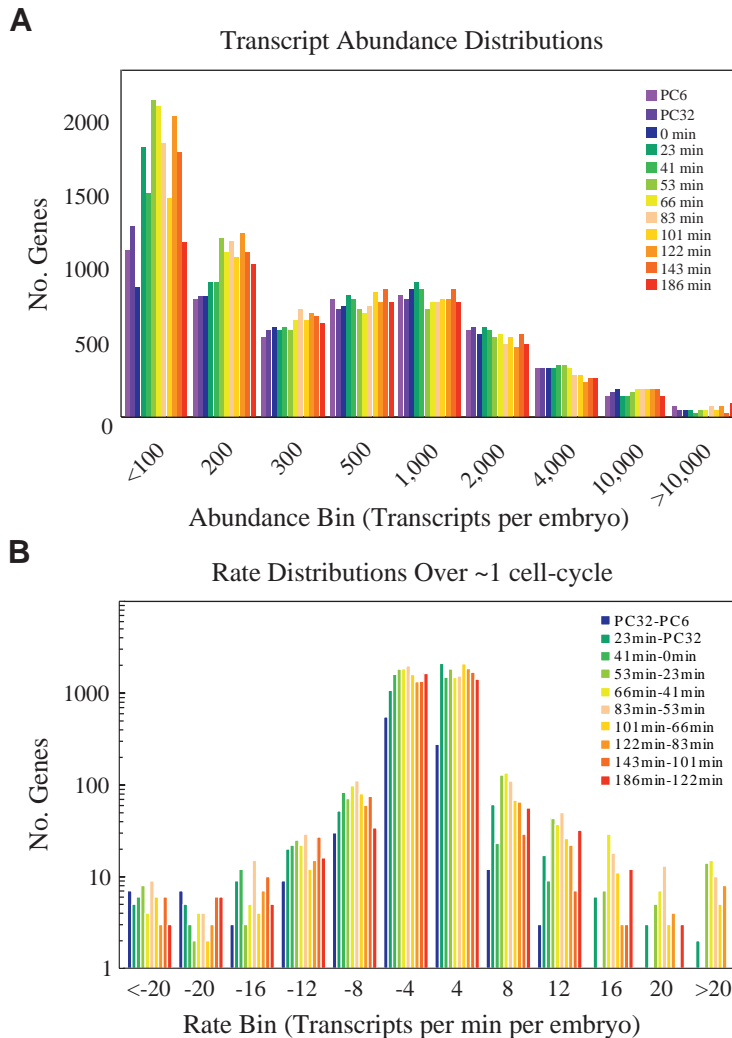


Fig. 6. The transcriptome maintains a steady-state distribution of transcript abundances during early embryogenesis. (A) A histogram plotting the distribution of transcript abundances among the RD genes for each of twelve time points assayed. Binned units along the x-axis are transcripts per embryo and 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.

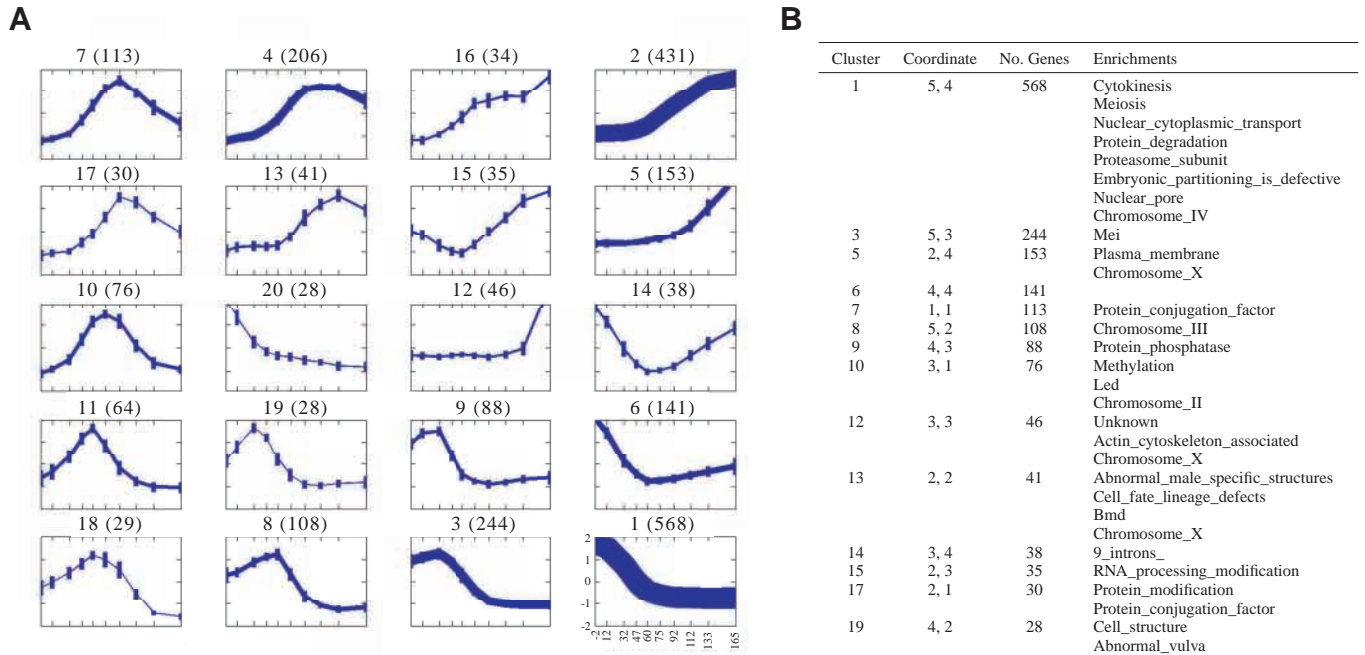


Fig. 7. Cluster analysis of expression profiles reveals the most predominant expression patterns as well as significant associations of gene annotations. (A) Means of the 20 largest of 106 total clusters are presented (including 79% of 3157 genes with $P < 0.001$ in either within-series ANOVA). Clusters 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

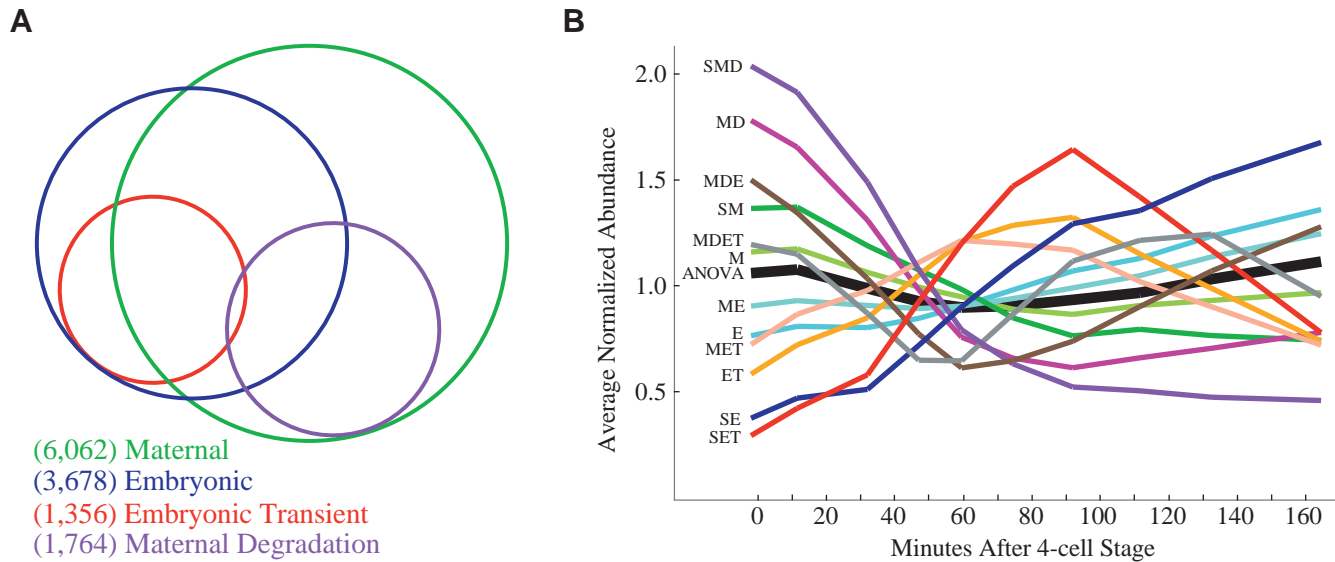


Fig. 8. Defined expression classes based on developmental concepts reveal an inflection point in the transition from maternal to embryonic control. (A) A Venn diagram relating by area the relative sizes and intersections of the four basis classes. The number of genes in each class is in parentheses. (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. phenotypic 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.
- Arbeitsman, 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., Sugang, 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 *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-Kellogg, 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.
- Labouesse, 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-3 β 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.