Developmental Cell
Resource

Genomic Analysis of Gastrulation and Organogenesis in the Mouse

Nesanet Mitiku[1] and Julie C. Baker[1,*]
[1]Department of Genetics, Stanford University, 300 Pasteur Drive, MC 5120, Stanford, CA 94305, USA
*Correspondence: jbaker@stanford.edu
DOI 10.1016/j.devcel.2007.10.004

SUMMARY

We developed a comprehensive dataset that samples the mouse transcriptome every 6 hr, from gastrulation through organogenesis. We observe an abrupt increase in overall transcript diversity at the onset of organogenesis (e8.0); the genes that comprise these changes are preferentially clustered along chromosome 7 and contain a significant enrichment of Gli binding sites. Furthermore, we identify seven dominant patterns of gene expression during gastrulation and organogenesis. Genes clustered according to these seven patterns constitute distinct functional classes, including a cluster enriched for gastrulation and pluripotency genes, two clusters differentially regulating localization and ion metabolism, and three clusters involved in discrete aspects of organogenesis. The last cluster is defined by a dramatic transient decrease in the expression of genes that regulate RNA processing and the cell cycle. *Drosophila* homologs of these genes are also coordinately downregulated following gastrulation, suggesting that the combined function of these genes has been conserved during metazoan evolution.

INTRODUCTION

During mammalian embryogenesis, an intricate series of morphological and molecular changes occur to establish the early body plan. Just after implantation, the mouse embryo is a hollow cylinder containing only a few cell types. Through gastrulation, this simple embryonic shape becomes organized and patterned. Germ layers emerge, including the mesoderm and definitive endoderm, and neural tissue is formed (Beddington and Robertson, 1998, 1999; Hogan et al., 1986). Gastrulation is followed by organogenesis, whereby the pattern established during previous stages is more intricately defined into whole organ systems. The transition from a cylindrical cup to a more distinctly embryonic form containing a head, heart, limbs, and spinal cord is rapid, occurring over the short span of three days (Hogan et al., 1986).

Proliferation and differentiation are the key components in establishing early cellular fates within the embryo. During early development in most species, cellular proliferation and differentiation are separated embryologically: rapid cycling during cleavage stages and differentiation during gastrula stages (O'Farrell et al., 2004). The mouse gastrula is an anomaly, containing both rapid cell cycles and differentiation. Gastrulation in the mouse is a particularly important period of cell cycle activities. The cell cycle accelerates during postimplantation development, becoming as short as 2–3 hr in some tissues and lacks gap phases (Snow, 1977, 1981; Snow and Bennett, 1978; Snow and Tam, 1980). It has been proposed that these mammalian cycles are similar to the cleavage cycles in most other species, including frogs and flies (O'Farrell et al., 2004). Decades of work have also shown that gastrulation in the mouse is a period of intense patterning and differentiation activities. Signaling pathways such as TGFβ, Wnt, and FGF, are critical in driving the proper movement of cells, the establishment of germ layers, and the formation of the body plan (Beddington and Robertson, 1998, 1999; Hogan et al., 1986). How rapid proliferation and differentiation coexist in the mouse gastrula is yet to be understood, although recently it has been proposed that, like other species, proliferation and differentiation are not truly contemporaries but rather closely juxtaposed during gastrulation so that the rapid cycling precedes a wave of intense differentiation (O'Farrell et al., 2004).

A global perspective is necessary to understand the complex processes that regulate proper morphogenesis. Genomic approaches have proven valuable in deciphering functional relationships among similarly expressed transcripts. In *C. elegans*, the analysis of the transcriptome of muscle cells provided compelling evidence that genes expressed together in muscle or coregulated in microarray experiments are spatially clustered along chromosomes (Roy et al., 2002; Wang et al., 2004). Work in human tissues and in yeast has also shown coexpressed genes to segregate in clusters along chromosomes (Caron et al., 2001; Cohen et al., 2000). Analysis of the transcriptome can also provide a glimpse into how the genome regulates these events. Global analysis of the preimplantation mouse embryo demonstrated two phases of transcription: one at the onset of zygotic transcription and the other at

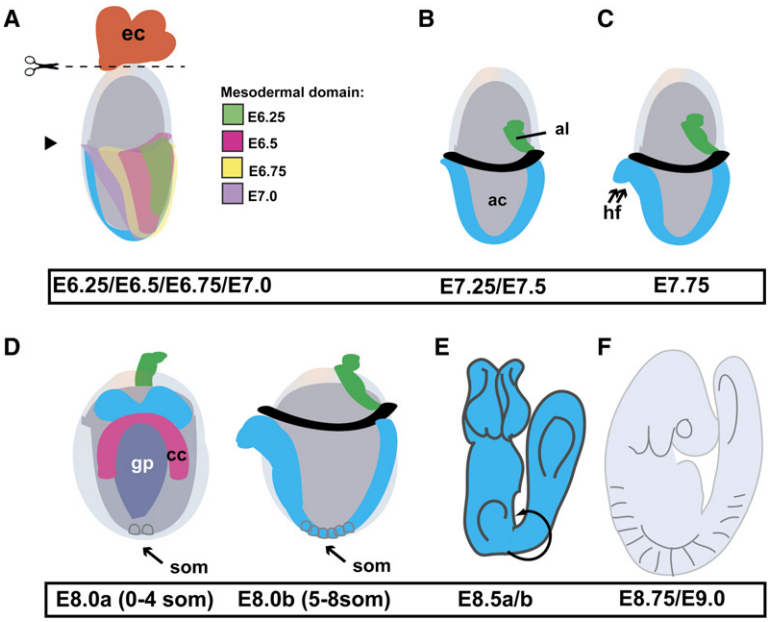


Figure 1. Criteria for Staging Embryos for Rigorous Analysis of Gastrulation and Organogenesis in the Mouse

(A) E6.25–E7.0 embryos were primarily distinguished by the extent of mesodermal migration toward the anterior (arrowhead) of the embryo. The mesodermal domain corresponding to each stage is marked in a different color (E6.25, green; E6.5, pink; E6.75, yellow; E7.0, purple). E7.0 was marked by the arrival of the mesoderm at the anterior of the embryo.

(B) At E7.25, the proamniotic canal narrows. By E7.5, the proamniotic canal closes, sealing off the amniotic cavity (ac), and the allantois (al) is visible posteriorly.

(C) The appearance of prominent head folds (double arrows, hf) was the hallmark of E7.75.

(D) At E8.0a (anterior view shown) the following are apparent: the gut pocket (gp), cardiac crescent (cc), and 0–4 somite pairs (som). E8.0b (lateral view shown) is morphologically similar to E8.0a, except for the increase in somite number to 5–8 somite pairs (som).

(E) E8.5 was marked by the turning of the embryo along its length. E8.5a embryos have 9–11 somite pairs and are in the process of turning. E8.5b embryos have 11–13 somite pairs and have completed turning.

(F) E8.75 and E9.0 were distinguished on the basis of somite number, bearing 13–16 and 17–20 somite pairs, respectively.

the 4 cell stage. This second transition was previously unappreciated as it is not observable morphologically (Hamatani et al., 2004; Wang et al., 2004), demonstrating that significant and unexpected developmental insights can be obtained through scrutiny of gene expression events in whole organisms.

The mammalian gastrula provides a challenge due to its small size and relative inaccessibility within the mother. This difficulty has not precluded the analysis of the mouse gastrula using high throughput approaches. Genome-scale screening projects, including the generation of mutants by ENU or enhancer trap mutagenesis, have been very effective in clarifying the role of single genes during gastrulation and organogenesis in the mouse (Eggenschwiler et al., 2001; Garcia-Garcia et al., 2005; Justice, 2000; Kile et al., 2003; Mitchell et al., 2001; Nord et al., 2006; Skarnes, 2005; Skarnes et al., 2004). In situ screens and expression cloning approaches have also been adapted for gene discovery in the mouse gastrula (Chiao et al., 2005; Sousa-Nunes et al., 2003). None of these techniques, however, offers a global perspective on the molecules expressed and regulated during this critical moment in embryonic time.

To gain insight into the global networks functioning during postimplantation development, we examined the transcriptome of the mouse embryo from early gastrulation through organogenesis taking time points every 6 hr beginning at e6.25 and concluding at e9.0. These time points provide high resolution that allows an unprecedented view of transcript diversity during very defined and short periods of gastrulation and organogenesis.

RESULTS

Transcript Diversity Varies Substantially across Developmental Stages

Extensive morphological change occurs within the embryo throughout gastrulation and early organogenesis. We hypothesized that these morphological changes are driven by dramatic changes within the transcriptome. Therefore, to analyze the pattern of the transcriptome during these stages, we generated microarray-based gene expression data representing thirteen different stages of postimplantation mouse embryogenesis. Samples were harvested at approximately 0.25 day intervals between e6.25 and e9.0. As developmental stages can vary significantly even within a single litter, embryonic stage was carefully determined, not simply by plug date, but by coupling estimated gestational age with morphological landmarks (Figure 1, Table S1 and Supplemental Experimental Procedures, see the Supplemental Data available with this article online). As RNA quantities were limiting for the earliest embryonic stages in our developmental time course, all samples were amplified (Eberwine, 1996) and a common reference-design hybridization approach was used (see Experimental Procedures). Given an estimated 25,613 protein coding genes in the mouse genome (Ensembl v32, July 2005, http://www.ensembl.org), approximately 74% (19,000/25,613) of the mouse genome was surveyed in this study. Data are available at GEO accession code GSE9046.

Surprisingly, relatively few genes change more than 2-fold between gastrulation and early neurula stages

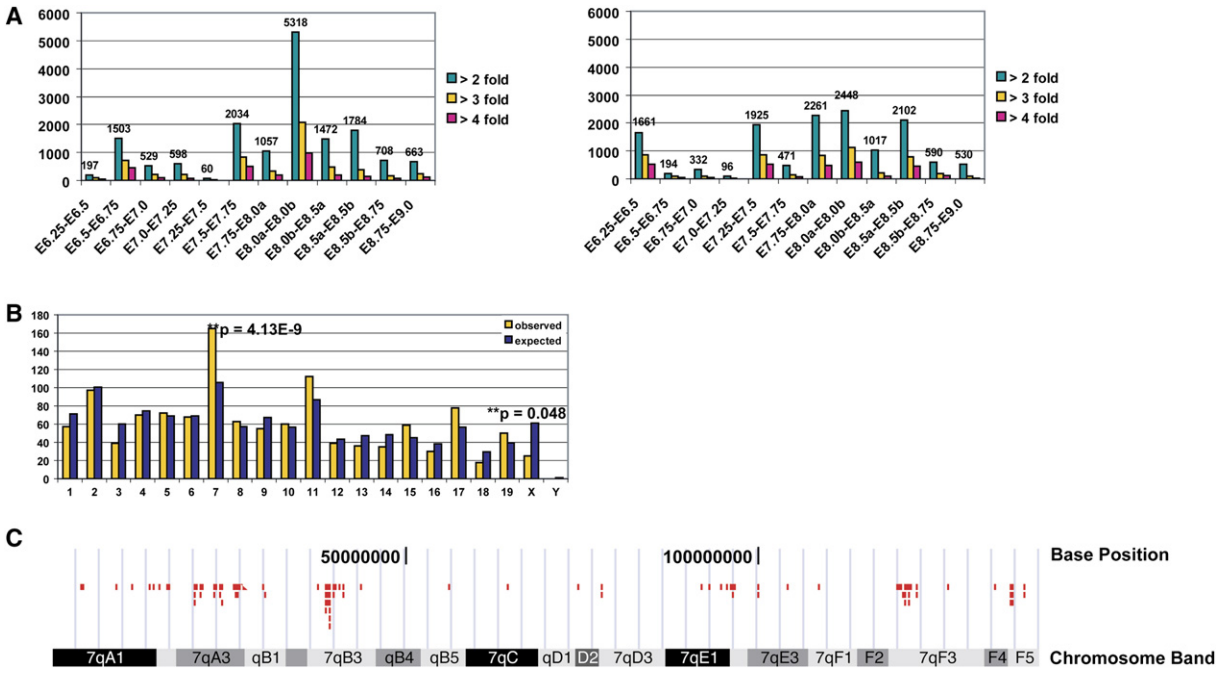


Figure 2. Somitogenesis Is Enriched for Transcripts that Are Located on Chromosome 7

(A) Genes increasing (left) or decreasing (right) over each developmental window either 2-, 3-, or 4-fold.

(B) One interval, e8.0a–e8.0b, has a significant distribution on chromosome 7. The observed distribution is shown in yellow while the expected distribution is shown in blue.

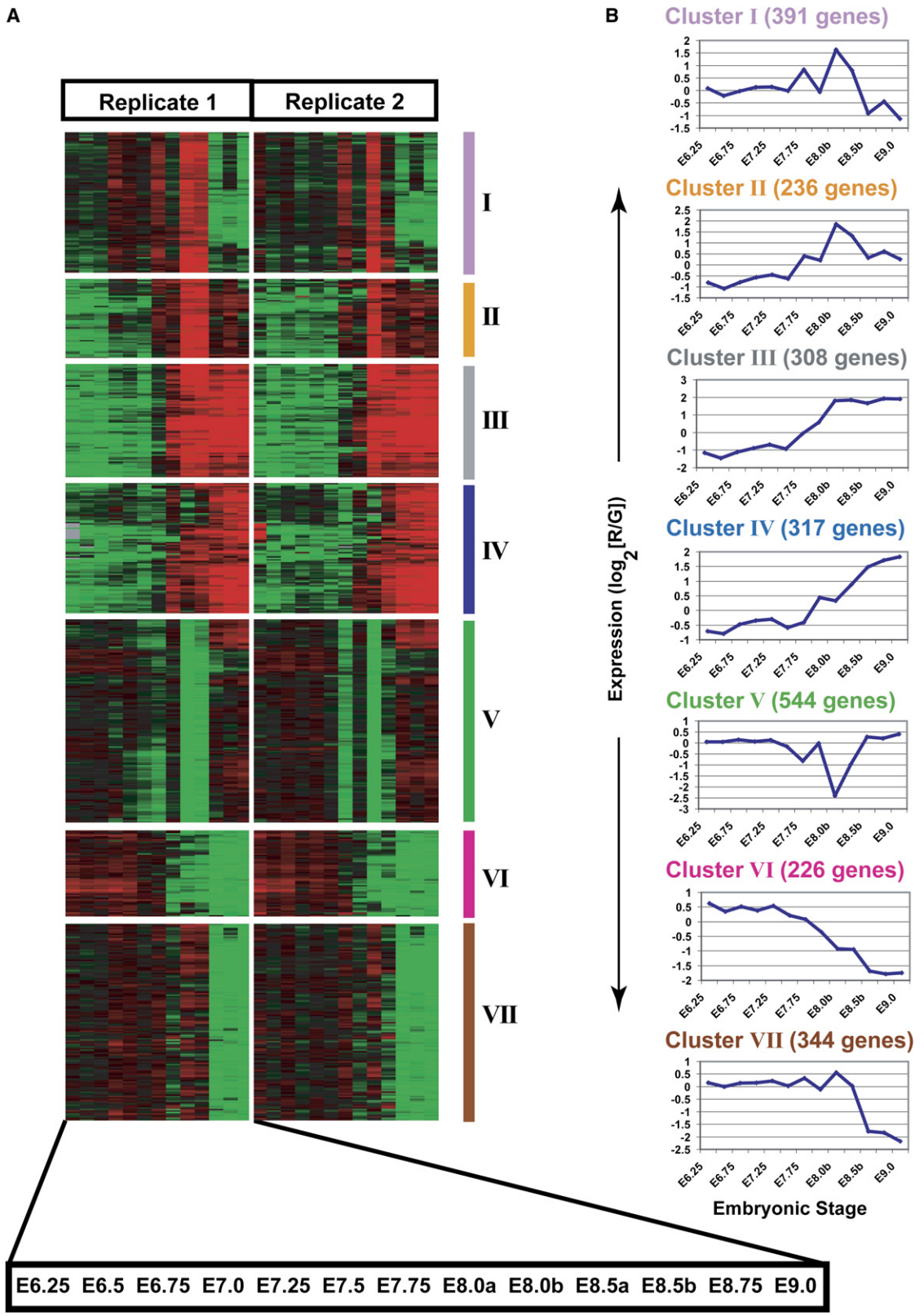
(C) Chromosomal location of genes is clustered in discrete regions on chromosome 7.

(Figure 2A and Tables S2 and S3). At both the beginning of gastrulation (e6.25–e6.5) and the end of gastrulation (e7.25–e7.5) the absolute extent of transcript change is minimal. The stability of transcript use during this time is puzzling since the embryo is undergoing significant morphological changes. This is particularly notable in the earliest window (e6.25–e6.5) when mesoderm and definitive endoderm emerge, migrate, and become patterned and suggests that either posttranscriptional processing or transcripts amassed at earlier stages drive these morphological changes. In the next developmental interval (e6.5–e6.75) more change is detected within the transcriptome. Consistent with the need during this window of time to promote massive cellular migration and cell fate changes, genes in this window are significantly annotated to cell communication (95 genes, $p = 3.03^{-5}$) and establishment of localization (82 genes, $p = 0.037$).

In contrast to the few changes in transcript abundance during gastrulation, the onset of organogenesis is marked by a burst in transcript diversity (Figure 2A). By far the greatest interval increase in transcript use occurs between the formation of 0 and 8 somites, a window of time approximately 8 hr in length (Figure 2A). This magnitude of change is not observed in any interval preceding or following, suggesting that the expansion and diversification of transcripts is a developmentally coordinated phenomenon. To better understand this molecular event, we analyzed the 5318 genes induced in this window for common functions or processes. Indeed these transcripts are significantly involved in cell differentiation (115 genes, $p = 1.82^{-4}$), organ development (205 genes, $p = 1.55^{-10}$), and morphogenesis (216 genes, $p = 2.96^{-10}$). Therefore, by e8.0 the mouse embryo is already heavily invested in organogenic processes.

Transcripts Induced at the Onset of Somitogenesis Cluster along Chromosome 7

The vast increase in transcript diversity occurring during the formation of only a few somites (e8.0a and e8.0b), suggests an almost operon-like response, whereby certain regions of the genome are efficiently unwound in order to generate coordinated transcription. Therefore, we assessed whether there was a chromosomal bias in the localization of genes induced throughout our time course. To assess differential chromosome use during all intervals examined, we compared the chromosomal distribution of genes that increase more than 3-fold to that expected for a similarly sized sample from the genome at large (NCBI build 34.1) using a χ^2 test. Three intervals (e7.5–e7.75, $p = 0.00087$; e8.0a–e8.0b, $p = 1.23^{-13}$; and e8.0b–e8.5, $p = 0.00025$) exhibited significant distributions across all chromosomes (data not shown). Although a number of chromosomes were overrepresented, by far the strongest enrichment was found on chromosome 7 between e8.0a–e8.0b ($p = 4.13^{-9}$; Figure 2B). Closer examination of overrepresented chromosome 7 genes (e8.0a–e8.0b) reveals that they are organized in clusters along the length of the chromosome, suggesting a potential shared local


A
Replicate 1
Replicate 2
I
II
III
IV
V
VI
VII
E6.25 E6.5 E6.75 E7.0 E7.25 E7.5 E7.75 E8.0a E8.0b E8.5a E8.5b E8.75 E9.0
B
Cluster I (391 genes)
Cluster II (236 genes)
Cluster III (308 genes)
Cluster IV (317 genes)
Cluster V (544 genes)
Cluster VI (226 genes)
Cluster VII (344 genes)
Expression (log2[R/G])
E6.25 E6.75 E7.25 E7.75 E8.0b E8.5b E9.0
Embryonic Stage

mechanism of induction (Figure 2C). This enrichment along chromosome 7 occurs in the same 8 hr window of time when transcript diversity as a whole increases dramatically, suggesting that chromatin changes rapidly occur to facilitate transcriptional change.

Transcriptional Regulation of Chromosome 7 Clusters

We hypothesized that the dramatic usage of clusters along chromosome 7 might be caused by the coordinated regulation of shared transcription factors at these upregulated sites along the chromosome. Therefore, we sought to determine whether these gene clusters shared regulatory mechanisms by examining their upstream regions for conserved transcription factor elements. To this end, we utilized rVISTA to determine conserved noncoding elements across chromosome 7 (Frazer et al., 2004). In this method, global sequence alignment is used to identify conserved noncoding elements. The resulting sites are examined for similarity to known transcription factor binding sites, and the identified transcription factor sites are tested for enrichment in the query list relative to the rest of the genome. We found the genes clustering along chromosome 7 to be enriched for several transcription factor binding sites. The most significant of these were binding sites for MAZR ($p = 10^{-15}$), LFA1($p = 10^{-11}$), and GLI ($p = 10^{-11}$). Of these, the most intriguing is the significant enrichment of GLI binding sites. In our dataset, both *Gli2* and *Gli3*, transcription factors involved in SHH signaling, increase in expression just prior to the expansion of the genes clustering along chromsome 7. Thus it is plausible that SHH signaling may play a key role in the coordinated upregulation of these clusters.

Microarray Data Set Recapitulates the Regulation of Known Genes over Time

A successful transcriptional map of early embryonic development should recapitulate previously analyzed gene expression changes within this time period. We have looked closely at known genes during the windows of development analyzed in this paper and find that almost all behave as expected within our time course (Figure S1). Those genes involved in maintaining pluripotency, including *Dppa4*, *Dppa5*, *Nanog*, and *Oct4*, are expressed at the highest levels at the earliest time points and decrease substantially by organogenesis stages (Figure S1A). Likewise, molecules that are known to play functional roles during gastrulation, including *Eomes*, *Brachyury*, *Lhx1*, and *Otx2*, are highly expressed during the gastrulation period, but are significantly downregulated during organogenesis (Figure S1B). Shortly after gastrulation, we observe an increase in the expression of genes involved in aspects of organ development. The induction of the neural markers *Pax6*, *HoxB1*, *Dlx5*, *Ncam*, and *Otx3* are shown in Figure S1E, demonstrating the proper timing for neural induction and patterning. Furthermore, we observe the upregulation of cardiac lineage markers and vascular markers coincidentally at the onset of organogenesis (Figures S1C and S1D). This analysis demonstrates that we are detecting accurate changes in our time course and can recapitulate the known expression of genes involved during these phases of embryogenesis.

Seven Dominant Patterns of Temporal Gene Expression Reveal Function

One of the goals of this study was to identify groups of genes that coordinately participate in processes relevant to early postimplantation mouse development. To this end, we utilized hierarchical clustering of our gene expression data to identify functionally related gene groups (Eisen et al., 1998). Prior to clustering, we selected genes demonstrating consistent expression patterns in replicate time courses using a Rank Products-based approach (Breitling et al., 2004; see Experimental Procedures). The resulting 6440 genes were clustered via the OLC (Bar-Joseph et al., 2001) algorithm and seven dominant temporal patterns were identified (Figure 3 and Table S4). The magnitude of gene expression change within each cluster is best appreciated in the accompanying plot representations (Figure 3B). From this point forward, these clusters will be referred to as Cluster I through Cluster VII.

Somitogenesis stages encompass most of the changes occurring—positively and negatively—within the transcriptome. All clusters (I–VII) demonstrate significant changes in the transcriptome during somitogenesis, perhaps correlated with the extreme number of transcripts changing during these periods (Figure 3). Cluster I, II, V, and VII contain the most extreme changes in transcript usage within the somitogenesis window of development. Cluster I and II demonstrate highest expression levels at the 5–8 somite stage (e8.0b) and are quickly downregulated after e8.5 (the difference in these clusters is a result of moderate expression of transcripts during gastrulation in Cluster I, at which time Cluster II has little or no expression). Conversely, Cluster V shows the lowest level of expression at e8.0b, which then rebounds by e8.5, while the levels of Cluster VII downregulate sharply at e8.5 and show no rebound within this time series.

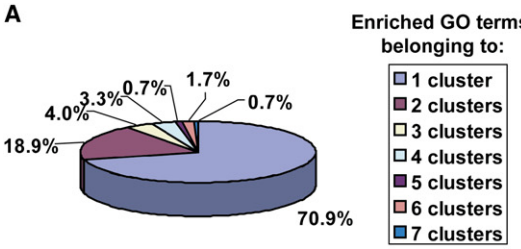
To determine whether there was recognizable biological relevance to expression patterns, we analyzed gene ontology (GO) terms with respect to biological processes. Each cluster contained significantly enriched process annotations (Figure 4 and Table S5). Intriguingly, there was little overlap in enriched biological function between the clusters: 70.9% of all process terms arose in only one cluster whereas 18.9% arose in two clusters. All seven clusters shared only 0.7% of the total process terms (see Figure 4A). These results strongly suggest that the individual

Figure 3. Dynamic Gene Expression Observed after Gastrulation

(A) Seven clusters demonstrate specific expression patterns during gastrulation and organogenesis and show that gastrulation is relatively quiescent compared to the beginning of organogenesis.

(B) Plot representations highlighting the magnitude of change in expression for each cluster.

A



B

Enriched Processes
Cluster I
localization (0.00)
ion Transport (1.11)
ion Homeostasis (1.25)
Cluster II
organogenesis (0.00)
blood vessel branching (0.00)
blood vessel morphogenesis (0.00)
blood vessel development (0.00)
neurogenesis (0.15)
Cluster III
organogenesis (0.00)
axonogenesis (0.00)
neurogenesis (0.00)
muscle development (0.00)
axon guidance (0.00)
Cluster IV
organogenesis(0.00)
neurogenesis (1.45)
muscle development (4.21)
kidney development (4.88)
urogenital development (5.0)
Cluster V
RNA processing (0.00)
RNA splicing/metabolism (0.00)
chromosome segregation (0.00)
cell cycle (2.5)
mitosis (1.0)
Cluster VI
Embryonic development (2.90)
regulation of metabolism (1.80)
regulation of transcription (2.38)
Cluster VII
localization (0.00)
transport (0.00)
ion homeostasis (0.25)
Wound healing (1.62)

Figure 4. Coexpressed Gene Groupings Have Specific Functions

Enriched GO terms were determined for each of the seven clusters.

(A) The fraction of enriched GO terms arising from one or more clusters, demonstrating selective functions for each cluster.

(B) Enriched process terms are shown for each cluster (see Table S5 for complete analysis).

clusters represent specific biological functions. This level of uniqueness is surprising, especially for the clusters with the most similar patterns of expression.

Organogenesis Clusters Have Distinct Functions

The distinct processes that define each cluster provides an opportunity to explore the roles of these gene groups during embryogenesis. Clusters II, III, and IV are the most biologically similar, being heavily annotated to processes involved in organogenesis and morphogenesis (Figure 4B; for a complete breakdown of biological process, molecular function, and cellular compartment annotations see Table S5). However, since the organogenesis and morphogenesis terms cast a very large umbrella, by looking at the smaller "nodes" we can deduce more specialized functions for each of these three clusters. Cluster II is heavily annotated toward vasculogenesis and angiogenesis, including significant enrichment in the following terms: branching morphogenesis (FDR = 0.00), blood vessel morphogenesis (FDR = 0.00), vascular development (FDR = 0.00), blood vessel development (FDR = 0.00), and angiogenesis (FDR = 0.21). This cluster also has an enrichment for the development of specialized organ systems, including formation of the gonads (FDR = 0.93), heart (FDR = 4.24), and kidney (FDR = 1.33). Cluster III is heavily annotated toward neural development, cell migration, and muscle movements. These enrichments include the following terms: axon guidance (FDR = 0.00), axonogenesis (FDR = 0.00), muscle contraction (FDR = 0.00), cell migration (FDR = 0.00), cell motility (FDR = 0.00), locomotion (FDR = 0.00), and neurogenesis (FDR = 0.00). Like Cluster II, this cluster also includes an enrichment in angiogenesis (FDR = 0.25) and vascular development (FDR = 0.26), which is certainly a case where Cluster II and III have overlapping functions. Cluster IV is less clearly annotated to any particular organ system, but does include annotations for neurogenesis (FDR = 1.45), muscle development (FDR = 4.21), kidney development (FDR = 4.88), and urogenital development (FDR = 5.0). Unlike Clusters II and III, Cluster IV is not annotated for any process involved in angiogenesis or vasculogenesis. Furthermore, there is little overlap with any of the annotations of Cluster II, suggesting that Cluster II and Cluster IV have very different functions in mediating organogenesis processes. Overall, these organogenesis clusters (II, III, and IV) are clearly involved in establishing organ morphogenesis within the embryo and their pattern of expression, increasing after gastrulation, is consist with this function.

Cluster I, although the expression changes are very similar to those of Cluster II, has a very distinct biological role and is not annotated for organogenesis or morphogenesis (Figure 4B). In this cluster the significant biological processes include localization (FDR = 0.00), transport (FDR = 1.11), and ion homeostasis (FDR = 1.25). Molecular

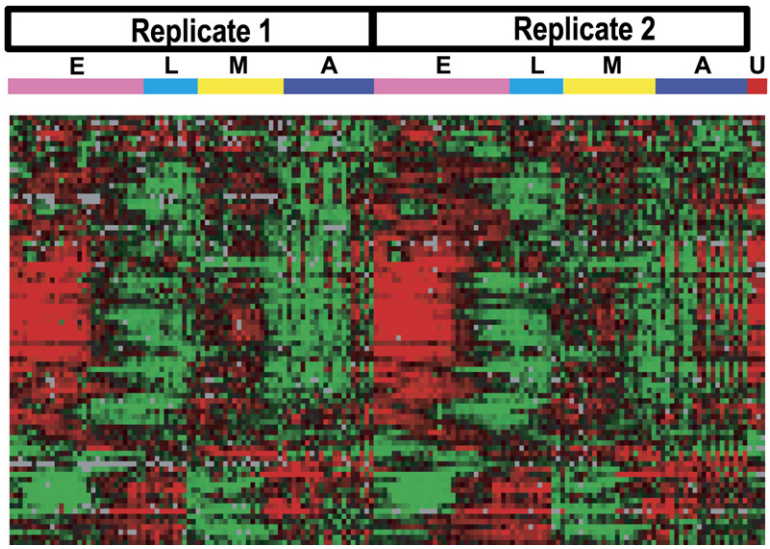


Figure 5. Cluster V Is Functionally Conserved in the Fly

Expression of the 65 *Drosophila* homologs of Cluster V genes throughout the *Drosophila* life cycle. Two replicate series are present. The stages of *Drosophila* development are as follows: E = embryo, L = larva, M = metamorphosis, A = adult, and U = unfertilized egg.

process annotations suggest that most of these biological roles are mediated by protein binding and the cellular compartment annotation indicates that this occurs extracellularly. This cluster is intriguing since although its expression changes are very similar to those of Cluster II, the annotations are distinctly different.

Gastrulation Clusters Involved in Pluripotency and Ion Metabolism

Cluster VI and VII represent groups of genes that have their highest expression during gastrulation and then are downregulated at the beginning of (VI) or during (VII) organogenesis. Although the expression profiles for these clusters have a similar trend, their functions are distinct, having no overlapping annotations. Cluster VI we label the "pluripotency" cluster since a close inspection of the genes expressed within this cluster reveals most of the known pluripotency molecules, including *Oct4*, *Nanog*, *Rex1*, *Dppa4*, *Dppa5*, *Dzl*, and *Utf1* (Chambers, 2004; Chambers et al., 2003; Loh et al., 2006; Mitsui et al., 2003; Pan et al., 2002). Known regulators of gastrulation are also expressed within this cluster, including *brachyury*, *follistatin*, *Lim1*, *Mesp2*, and *Eomes* (Beddington and Robertson, 1998, 1999; Hogan et al., 1986). Consistent with this finding, this cluster is enriched for embryonic development terms (FDR = 2.90). The highest enrichment of function in Cluster VI, however, is with regulation of cellular processes, including regulation of metabolism (FDR = 1.80) and regulation of transcription (FDR = 2.38) and unlike many of the other clusters, this "pluripotency cluster" is enriched in molecules that play a role within the cell, suggesting that during gastrulation cellular functions are autonomous (Table S5).

Cluster VII shares mores functional similarity to Cluster I than any other, even though their patterns of expression are very distinct (Figure 4B). Cluster VII is significantly annotated to transport (FDR = 0.00), localization (FDR = 0.00), ion homeostasis (FDR = 0.25), and wound healing (FDR = 1.62). Like Cluster I, the cellular compartment regulating these processes is likely to be extracellular (Table S5). The similarity in annotations within these clusters is intriguing, suggesting that they serve similar functions during two distinct phases of embryonic development.

Cluster V Is Enriched for RNA Processing and Cell Cycle Regulation

Genes in Cluster V exhibit a particularly striking and severe dip in expression pattern beginning at 0–4 somites (E8.0a) and ending at e8.5 (Figure 3B and Table S6). Unlike any of the other clusters, Cluster V is heavily enriched for RNA processing (FDR = 0.00), RNA splicing (FDR = 0.00), RNA metabolism (FDR = 0.00), chromosome segregation (FDR = 0.00), cell cycle (FDR = 0.25), and mitosis (FDR = 1.00). The regulation of this cluster coincides with the dramatic increase in transcripts observed at e8.0b, suggesting a possible correlation between the downregulation of Cluster V RNA processing and cell cycle genes and the increase in transcript complexity.

The Pattern of Cluster V Transcripts Is Conserved in Evolution between Mice and Flies

Given the unusual expression pattern embodied by Cluster V, its basic molecular annotation, and its putative relationship to the expansion of the transcriptome and the onset of organogenesis, we sought to evaluate its evolutionary importance. We hypothesized that these genes would be expressed in a similar pattern in other organisms. In order to identify the *Drosophila* homologs of these the 544 mouse genes comprising Cluster V, we queried the Ensembl (http://www.ensembl.org) database. Of the 152 *Drosophila* homologs retrieved via Ensembl, 65 *Drosophila* homologs were measured for their expression patterns during the entire *Drosophila* life cycle (Arbeitman et al., 2002). After hierarchical clustering, we show that, as in mice, these 65 homologs are coordinately regulated during *Drosophila* development (Figure 5). The majority of the *Drosophila* Cluster V genes have highest expression during embryogenesis and begin their downregulation at the 14th

cell cycle, just following gastrulation. The expression of this cluster is lowest during larval stages and begins to rebound during metamorphosis. In order to evaluate the chance probability of making such an observation, we ran 10,000 simulated draws of *Drosophila* genes in the Arbeitman dataset (see Experimental Procedures) and assessed the extent to which these randomly selected groups recapitulated the observed pattern. In 10,000 simulations, we observed 80 such occurrences indicating that the probability of making such an observation by chance is exceedingly small ($p = 0.008$). Thus, the expression-based association of Cluster V genes is evolutionarily conserved. Furthermore, the *Drosophila* Cluster V genes are also significantly annotated toward "cell cycle" and "RNA processing," suggesting not only a conserved pattern of expression, but conserved function as well.

DISCUSSION

In this report, we provide a comprehensive analysis of the transcriptome during mouse gastrulation and organogenesis. This analysis is a critical step in the assessment of the global molecular activity occurring during early embryogenesis: a window that sets the tone for all ensuing developmental events. To this end, we provide insight into the regulation of genes during mouse embryogenesis and define groups of genes that have distinct functions, including those with roles in pluripotency, vasculogenesis, RNA processing, and cell cycle. We further show that the transcriptional landscape is largely invariant until the onset of somitogenesis when a significant change in transcript diversity occurs. Genes induced at this time are largely clustered in regions along chromosome 7 and their coordinated expression may be mediated by a handful of transcription factors whose binding sites lie within the promoter regions of these genes.

The beginning of somitogenesis is also categorized by an abrupt decrease in molecules involved in RNA processing and the cell cycle, leading us to hypothesize that these activities may provide a controlled mechanism of entry into differentiation. This pattern of expression is conserved in *Drosophila* and—more significantly—this grouping marks the same embryological window across two developmentally distinct species, strongly suggesting that this block of genes indeed acts together toward a functional goal during the entrance into organogenesis and differentiation. What function might these gene groupings reflect? The cell cycle during mouse gastrulation is rapid, with some cells dividing every 2–3 hr. These rapid cell cycles slow down as gastrulation commences and as such it is intriguing that Cluster V genes are also enriched for cell cycle annotations. A closer analysis of the genes annotated for cell cycle show that one third are involved in the progression through M phase and suggest that the decrease of these genes may provide a resistance to mitosis between 0 and 8 somites (e8.0a–e8.0b). Further examination of these Cluster V genes reveals members that can bridge distinct Cluster V processes. For example, *Xpo1* (*Crm1*; annotated as an RNA processing gene) connects mRNA export and mitosis by recruiting *Ranbp2*, another Cluster V gene, to mitotic kinetochores (Arnaoutov et al., 2005). FUSIP1 (SRp38) couples mRNA splicing to M phase of the cell cycle by mediating M-phase-specific splicing inhibition (Shin et al., 2004). Cells lacking FUSIP1 also have a prolonged G2/M phase indicating an important role in mitotic progression (Shin et al., 2004). SUZ12, a member of the polycomb group of proteins, links chromatin remodeling to cell proliferation; mutants for SUZ12 display histone methylation deficits, fail to proliferate adequately, and exhibit a complete absence of organogenesis (Pasini et al., 2004). Furthermore, Cluster V RNA processing genes are functionally diverse within the realm of posttranscriptional regulation with roles in capping, splicing, exosomal activity, and nucleocytoplasmic trafficking. This places Cluster V RNA processing genes in a position to guide cellular transcript composition at multiple steps. Since RNA abundance is simply the net balance of synthesis and destruction, one could envision a model in which RNA processing steps are used to clear the gene expression palate in order to accommodate new expression programs. Certainly, this would be an efficient means to effect a developmental transition, particularly one leading to a process as complex and varied as organogenesis. Thus, a synthesis of Cluster V gene activities suggests an intimate association of RNA processing, chromosomal state, and the mitotic phase of the cell cycle. Together these processes may suggest a multifaceted approach to the critical coordination of proliferation and differentiation during mammalian development. More specifically, given the Cluster V pattern, this group of genes is well situated to coordinate proliferation and differentiation at the entrance to organogenesis.

In this report, we have provided a glimpse into the regulation of the transcriptome throughout the postimplantation mouse embryo. The number of time points and the statistical rank products analysis method greatly bolsters the quality of the data. Our hope is that these data will be useful to the scientific community and that they will be mined for further insights into gastrulation and early organogenesis in the mouse.

EXPERIMENTAL PROCEDURES

Embryo Dissection, RNA Isolation, and RNA Amplification

All mice used in this study were outbred Swiss Websters. Embryos were dissected in cold PBS containing 10% Bovine Calf Serum (GIBCO). At each stage from e6.25–e7.5 more than 200 embryos were dissected to establish each biological replicate and ectoplacental cone was removed to increase homogeneity of the fetal samples. At each stage from e7.75–e9, each replicate ranged from 90 embryos to 10, respectively. Again, these samples were dissected free from all extraembryonic material.

Total RNA was prepared using a variation on the acid phenol method of Chomczynski and Sacchi, 1987. After staging, embryos were thoroughly homogenized in solution D (4M guanidinium thiocyanate, 100mM NaOAc, 0.5% Sarkosyl, 0.1M beta-mercaptoethanol) and stored at −80°C until RNA isolation. Following isolation and precipitation in the presence of linear acrylamide (Ambion), the resulting RNA pellet was further purified using the RNAqueous Mini Kit (Ambion).

RNA was quantified using Ribogreen reagent (Molecular Probes). Five hundred nanograms of total RNA were used in each amplification reaction (mMessage Amp, Ambion). Amplified RNA is heretofore referred to as aRNA.

Reference RNA Composition and Isolation

A type II (common reference) experimental approach was enlisted in this study. The common reference was comprised of a mixture of E17.5 embryo, E17.5 placenta, CGR8 ES cells, and adult female brain aRNA. Briefly, all tissues used in the reference RNA pool were ground via mortar and pestle in the presence of liquid nitrogen, put through one acid phenol extraction, two Trizol (Invitrogen) extractions, DNase I (Ambion) treated, and then repurified over RNAqueous Midi columns (Ambion). The ES cell component of the reference pool was purified similarly beginning with the acid phenol extraction. In lieu of mortar and pestle grinding, ES cells were spun down and directly resuspended in solution D. Total RNA from all four sources was combined at a proportion of 1:1.5:2:2.5 (brain:placenta:ES cells:E17.5 embryos) and then amplified (mMessageAmp, Ambion). All reference amplification reactions were pooled. This reference aRNA pool was used for all hybridizations in this study.

Labeling and Hybridization

Each labeling reaction was carried out using 4 μg of aRNA, 10 μg of pDN6 (Amersham), 0.6 μl 50X dNTPs (Invitrogen), 3 μl 0.1M DTT, and 2 μl Superscript II (Invitrogen) in the supplied buffer. Labeled aRNA samples were purified over Microcon YM-30 columns (Millipore). All samples were hybridized to mouse 42K cDNA arrays produced by the Stanford Functional Genomics Facility. These arrays are printed on Ultra GAPS or GAPSII coated glass slides (Corning) and are comprised of the NIA 15K clone set, NIA 7K clone set, and the RIKEN 21K clone set; all spots on array are PCR products from plasmid clones (Kargul et al., 2001; Sharov et al., 2003; Tanaka et al., 2000; VanBuren et al., 2002). The number of genes present on the 42K array, as determined by Unigene cluster, is 19,000. Final hybridization composition per 40 μl sample was as follows: 10.14 μl purified sample (Cy3 and Cy5 combined), 20 μl 2X formamide hybridization buffer (Genisphere), 5.2 μl deionized formamide (Ambion), 10 μg polyA DNA (Sigma), 10 μg yeast tRNA (Sigma), 0.4 μl 0.1M DTT, 1.5 μl 50X Denhardts solution (Sigma), 20 μg mouse Cot I DNA (Invitrogen), and 0.76 μl H_20. Hybridization was carried out at 42°C for 20 hr. After washing, arrays were spun dry and immediately coated with DyeSaver (Genisphere) in order to prevent Cy5 oxidation.

Image Processing

Arrays were scanned using a single Axon 4000B scanner. The power settings for scanning were optimized to generate a 1:1 ratio of overall Cy3 to Cy5 signal for self-self hybridization experiments using our common reference. Once optimal PMT settings were determined (Cy5 channel = 660; Cy3 channel = 540), they were kept constant for all arrays in this dataset. Numerical values were extracted from scanned images using a combination of Gridmaker/Spotreader (Niles Scientific) and Genepix5.0 (Axon Instruments, Inc). Spots were identified and checked for adequate spot quality using Gridmaker/Spotreader. Results were extracted using Genepix5.0 in the standard Genepix results format. All data are available from the Stanford Microarray Database (SMD) (Sherlock et al., 2001).

Data Processing and Selection

Array elements were considered for further analysis if they met the following requirements: (1) adequate spot quality after visual inspection, (2) a mean fluorescence intensity at least 1.5 times the median local background reference signal in the reference (Cy3) channel, and (3) the presence of a fluorescence measurement in at least 70% of all experiments. Total intensity normalization was performed within SMD (Sherlock et al., 2001). Array elements representing identical clones were collapsed to a single representation by averaging the data across all identical clones. After these filters were applied, 29,420 clones remained. The data were median transformed prior to further analyses. In order to enrich for genes consistently and dramatically changing in the time course, a second level of filtration was conducted using a modification on the Rank Product (RP) method (Breitling et al., 2004). Within each replicated dataset, differences between pairs of experimental time points for each gene/clone were calculated, resulting in an interval representation for all possible time point combinations. Each interval representation was then compared across replicates. The correspondence across replicates was assessed using the RP method (FDR = 0.20) with the following modifications: (1) the substitution of MersenneTwister PRNG (MTP) (Matsumoto and Nishimura, 1998) for Perl's rand, (2) the seeding of MTP with numbers retrieved from http://www.random.org, and (3) the assignment of equal ranks for genes with equal differentials between time points. A master clonelist/genelist was constructed by taking the union of each comparison across replicates and collapsing the resulting list for duplicate assignments. After this step, 6440 clones/genes remained for further analysis.

Data Analysis

Experiments were ordered with respect to temporal progression and the data were clustered by gene using the Optimal Leaf Clustering (OLC) method (Bar-Joseph et al., 2001). Clustered data were visualized with Java TreeView (Saldanha, 2004) (http://jtreeview.sourceforge.net/). Seven clusters were identified in the resulting clustergram. We designed a set of Perl scripts to examine GO term enrichment in each of the seven dominant clusters. These scripts utilize GO::TermFinder (Version 0.7) (Boyle et al., 2004), a set of Perl modules freely available from the Comprehensive Perl Archive Network (CPAN). The ontology and gene association files were from the July 2005 distribution available from http://www.geneontology.org. GO terms were designated as enriched for a particular cluster at an FDR cutoff of 0.10. In order to visualize the representation of GO terms across clusters, a matrix-format Boolean representation of GO term enrichment was developed. In this representation, enriched GO terms are represented by rows and clusters are represented as columns. Clustering of the GO terms was done via OLC and color-coding was done through Java TreeView (Saldanha, 2004). A cluster enriched for a particular GO term is indicated by a purple rectangle at the intersection of its column and the row corresponding to the relevant GO term.

The mouse genes in Cluster V were mapped to their *Drosophila* homologs (Dm-ClusterV) using homology data downloaded via the Ensembl BioMart functionality (Harris et al., 2004). The Arbeitman (Arbeitman et al., 2002) *Drosophila* life cycle dataset was downloaded through the SMD. Dm-ClusterV genes were extracted from this dataset. The population from which all simulations were derived was defined as those genes in the *Drosophila* dataset with mouse homologs. The centroid standard for the Pearson R-squared calculation was generated using the expression data for 30% of Dm-ClusterV genes exhibiting the dominant expression pattern; the centroid was defined as the mean expression curve. Centroid-generating genes were filtered from both the Dm-ClusterV and simulation population genelists prior to the calculation of R-squared distributions. R-squared distributions were compared using the Mann-Whitney Rank Sum test. The probability of observing Dm-ClusterV-like expression was estimated by the number of simulations resulting in an R-squared distribution right-shifted relative to that of Dm-ClusterV divided by the total number of simulations (10,000).

Supplemental Data

Supplemental Data include six tables, one figure, and Supplemental Experimental Procedures and are available at http://www.developmentalcell.com/cgi/content/full/13/6/897/DC1/.

ACKNOWLEDGMENTS

We thank L. Lazzeroni for statistical advice, R. Pesich, C.S. April, and M. Alizadeh for sharing micorarray expertise, O. Troyanskaya, T. Kawli,

and D.S. Manoli for discussion and comments on this manuscript, P.B. Menage for programming discussions, and R.M. Myers, D. Vollrath, and M.W. Tan for the use of essential equipment. This work was supported by grants from the NIH (RO1 HD 41557) and from the Medical Scientist Training Program (GM07365).

Received: May 25, 2007
Revised: August 23, 2007
Accepted: October 9, 2007
Published: December 3, 2007

REFERENCES

Arbeitman, M.N., Furlong, E.E., 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.

Arnaoutov, A., Azuma, Y., Ribbeck, K., Joseph, J., Boyarchuk, Y., Karpova, T., McNally, J., and Dasso, M. (2005). Crm1 is a mitotic effector of Ran-GTP in somatic cells. Nat. Cell Biol. *7*, 626–632.

Bar-Joseph, Z., Gifford, D.K., and Jaakkola, T.S. (2001). Fast optimal leaf ordering for hierarchical clustering. Bioinformatics *17* (*Suppl 1*), S22–S29.

Beddington, R.S., and Robertson, E.J. (1998). Anterior patterning in mouse. Trends Genet. *14*, 277–284.

Beddington, R.S., and Robertson, E.J. (1999). Axis development and early asymmetry in mammals. Cell *96*, 195–209.

Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO:TermFinder–open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics *20*, 3710–3715.

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. *573*, 83–92.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., et al. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science *291*, 1289–1292.

Chambers, I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells. Cloning Stem Cells *6*, 386–391.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643–655.

Chiao, E., Leonard, J., Dickinson, K., and Baker, J.C. (2005). High-throughput functional screen of mouse gastrula cDNA libraries reveals new components of endoderm and mesoderm specification. Genome Res. *15*, 44–53.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. *162*, 156–159.

Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat. Genet. *26*, 183–186.

Eberwine, J. (1996). Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. Biotechniques *20*, 584–591.

Eggenschwiler, J.T., Espinoza, E., and Anderson, K.V. (2001). Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. Nature *412*, 194–198.

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

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res. *32*, W273–W279.

Garcia-Garcia, M.J., Eggenschwiler, J.T., Caspary, T., Alcorn, H.L., Wyler, M.R., Huangfu, D., Rakeman, A.S., Lee, J.D., Feinberg, E.H., Timmer, J.R., et al. (2005). Analysis of mouse embryonic patterning and morphogenesis by forward genetics. Proc. Natl. Acad. Sci. USA *102*, 5913–5919.

Hamatani, T., Carter, M.G., Sharov, A.A., and Ko, M.S. (2004). Dynamics of global gene expression changes during mouse preimplantation development. Dev. Cell *6*, 117–131.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, D258–D261.

Hogan, B., Costantini, F., and Lacy, E. (1986). Manipulating the Mouse Embryo: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Justice, M.J. (2000). Capitalizing on large-scale mouse mutagenesis screens. Nat. Rev. Genet. *1*, 109–115.

Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka, T.S., Carter, M.G., and Ko, M.S. (2001). Verification and initial annotation of the NIA mouse 15K cDNA clone set. Nat. Genet. *28*, 17–18.

Kile, B.T., Hentges, K.E., Clark, A.T., Nakamura, H., Salinger, A.P., Liu, B., Box, N., Stockton, D.W., Johnson, R.L., Behringer, R.R., et al. (2003). Functional genetic analysis of mouse chromosome 11. Nature *425*, 81–86.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Matsumoto, M., and Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Trans. Model. Comput. Simul. *8*, 3–30.

Mitchell, K.J., Pinson, K.I., Kelly, O.G., Brennan, J., Zupicich, J., Scherz, P., Leighton, P.A., Goodrich, L.V., Lu, X., Avery, B.J., et al. (2001). Functional analysis of secreted and transmembrane proteins critical to mouse development. Nat. Genet. *28*, 241–249.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Nord, A.S., Chang, P.J., Conklin, B.R., Cox, A.V., Harper, C.A., Hicks, G.G., Huang, C.C., Johns, S.J., Kawamoto, M., Liu, S., et al. (2006). The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. Nucleic Acids Res. *34*, D642–D648.

O'Farrell, P.H., Stumpff, J., and Su, T.T. (2004). Embryonic cleavage cycles: how is a mouse like a fly? Curr. Biol. *14*, R35–R45.

Pan, G.J., Chang, Z.Y., Scholer, H.R., and Pei, D. (2002). Stem cell pluripotency and transcription factor Oct4. Cell Res. *12*, 321–329.

Pasini, D., Bracken, A.P., Jensen, M.R., Denchi, E.L., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J. *23*, 4061–4071.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature *418*, 975–979.

Saldanha, A.J. (2004). Java Treeview–extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

Sharov, A.A., Piao, Y., Matoba, R., Dudekula, D.B., Qian, Y., VanBuren, V., Falco, G., Martin, P.R., Stagg, C.A., Bassey, U.C., et al. (2003). Transcriptome analysis of mouse stem cells and early embryos. PLoS Biol. *1*, E74.

Sherlock, G., Hernandez-Boussard, T., Kasarskis, A., Binkley, G., Matese, J.C., Dwight, S.S., Kaloper, M., Weng, S., Jin, H., Ball, C.A., et al. (2001). The Stanford Microarray Database. Nucleic Acids Res. *29*, 152–155.

Shin, C., Feng, Y., and Manley, J.L. (2004). Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature *427*, 553–558.

Skarnes, W.C. (2005). Two ways to trap a gene in mice. Proc. Natl. Acad. Sci. USA *102*, 13001–13002.

Skarnes, W.C., von Melchner, H., Wurst, W., Hicks, G., Nord, A.S., Cox, T., Young, S.G., Ruiz, P., Soriano, P., Tessier-Lavigne, M., et al. (2004). A public gene trap resource for mouse functional genomics. Nat. Genet. *36*, 543–544.

Snow, M.H. (1981). Growth and its control in early mammalian development. Br. Med. Bull. *37*, 221–226.

Snow, M.H., and Bennett, D. (1978). Gastrulation in the mouse: assessment of cell populations in the epiblast of tw18/tw18 embryos. J. Embryol. Exp. Morphol. *47*, 39–52.

Snow, M.H., and Tam, P.P. (1980). Timing in embryological development. Nature *286*, 107.

Snow, M.H.L. (1977). Gastrulation in the mouse: Growth and regionalization of the epiblast. J. Embryol. Exp. Morphol. *42*, 293–303.

Sousa-Nunes, R., Rana, A.A., Kettleborough, R., Brickman, J.M., Clements, M., Forrest, A., Grimmond, S., Avner, P., Smith, J.C., Dunwoodie, S.L., et al. (2003). Characterizing embryonic gene expression patterns in the mouse using nonredundant sequence-based selection. Genome Res. *13*, 2609–2620.

Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., et al. (2000). Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. Proc. Natl. Acad. Sci. USA *97*, 9127–9132.

VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin, P.R., Stagg, C.A., Bassey, U.C., Aiba, K., Hamatani, T., et al. (2002). Assembly, verification, and initial annotation of the NIA mouse 7.4K cDNA clone set. Genome Res. *12*, 1999–2003.

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., and Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev. Cell *6*, 133–144.