pend on Hec1 may signal checkpoint activation through diffusible Mad2 complexes. In Hec1-depleted cells, this activity could be generated through CENP-E or BubR1. Because kinetochores were not stretched in Hec1-depleted cells (*30*), it is plausible that persistent checkpoint activity was caused by lack of tension.

Injection of antibodies to Hec1 into bladder carcinoma cells was reported to cause aberrant mitotic progression and cell death but no checkpoint arrest (*23*). This result could be explained if these tumor cells were checkpoint-deficient or if the injected antibodies interfered with checkpoint signaling. In *Saccharomyces cerevisiae*, mutations in the *Hec1* homolog *Ndc80* caused chromosome segregation defects without activating the checkpoint (*24*, *26*). This may relate to the fact that kinetochores in budding yeast bind only a single MT, whereas those in vertebrate cells capture multiple MTs (*8*, *9*). Furthermore, kinetochore–MT interactions and checkpoint signaling in vertebrates may involve two distinct pathways: one centered on Hec1 interacting with Mad1/Mad2 and the other on CENP-E interacting with CENP-F and BubR1, both pathways converging onto APC/C (*35*, *36*). Yeast has a clear counterpart of Hec1 but lacks an obvious homolog of CENP-E.

The human kinetochore protein Hec1 was required, together with Mps1, for recruiting the Mad1/Mad2 complex to kinetochores. Moreover, Hec1-depleted cells displayed persistent spindle checkpoint activity although they lacked significant amounts of Mad1 or Mad2 at kinetochores. This latter observation contrasts with models emphasizing the importance of high steady-state levels of kinetochore-associated Mad1/Mad2 complexes in checkpoint signaling and instead suggests that some protein that does not depend on Hec1 for kinetochore localization is able to communicate with diffusible Mad2 complexes. Many tumor cells are thought to be defective in the spindle checkpoint (*37*). Any interference with Hec1 function in checkpoint-deficient cells, be it through siRNA or other specific inhibitors, is predicted to result in mitotic catastrophe, thereby causing the demise of most progeny. In contrast, normal checkpoint-proficient cells may arrest transiently in response to reversible Hec1 inhibition. Thus, Hec1 may be an attractive target for therapeutic intervention in cancer and other hyperproliferative diseases.

References and Notes

1. R. B. Nicklas, *Science* **275**, 632 (1997).
2. K. Nasmyth, J. M. Peters, F. Uhlmann, *Science* **288**, 1379 (2000).
3. A. Musacchio, K. Hardwick, *Nature Rev. Mol. Cell Biol.*, in press.
4. J. V. Shah, D. W. Cleveland, *Cell* **103**, 997 (2000).
5. C. L. Rieder, R. W. Cole, A. Khodjakov, G. Sluder, *J. Cell Biol.* **130**, 941 (1995).
6. M. A. Hoyt, L. Totis, B. T. Roberts, *Cell* **66**, 507 (1991).
7. R. Li, A. W. Murray, *Cell* **66**, 519 (1991).
8. K. Kitagawa, P. Hieter, *Nature Rev. Mol. Cell Biol.* **2**, 678 (2001).
9. C. L. Rieder, E. D. Salmon, *Trends Cell Biol.* **8**, 310 (1998).
10. V. M. Stucke, H. H. Sillje, L. Arnaud, E. A. Nigg, *EMBO J.* **21**, 1723 (2002).
11. S. S. Taylor, E. Ha, F. McKeon, *J. Cell Biol.* **142**, 1 (1998).
12. X. Luo, Z. Tang, J. Rizo, H. Yu, *Mol. Cell* **9**, 59 (2002).
13. L. Sironi *et al.*, *EMBO J.* **20**, 6371 (2001).
14. G. Fang, H. Yu, M. W. Kirschner, *Genes Dev.* **12**, 1871 (1998).
15. M. Kallio, J. Weinstein, J. R. Daum, D. J. Burke, G. J. Gorbsky, *J. Cell Biol.* **141**, 1393 (1998).
16. L. H. Hwang *et al.*, *Science* **279**, 1041 (1998).
17. V. Sudakin, G. K. Chan, T. J. Yen, *J. Cell Biol.* **154**, 925 (2001).
18. Z. Tang, R. Bharadwaj, B. Li, H. Yu, *Dev. Cell* **1**, 227 (2001).
19. B. J. Howell, D. B. Hoffman, G. Fang, A. W. Murray, E. D. Salmon, *J. Cell Biol.* **150**, 1233 (2000).
20. D. A. Skoufias, P. R. Andreassen, F. B. Lacroix, L. Wilson, R. L. Margolis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4492 (2001).
21. J. C. Waters, R. H. Chen, A. W. Murray, E. D. Salmon, *J. Cell Biol.* **141**, 1181 (1998).
22. Materials and methods are available as supporting material on *Science* Online.
23. Y. Chen, D. J. Riley, P. L. Chen, W. H. Lee, *Mol. Cell. Biol.* **17**, 6049 (1997).
24. P. A. Wigge *et al.*, *J. Cell Biol.* **141**, 967 (1998).
25. P. A. Wigge, J. V. Kilmartin, *J. Cell Biol.* **152**, 349 (2001).
26. X. He, D. R. Rines, C. W. Espelin, P. K. Sorger, *Cell* **106**, 195 (2001).
27. C. Janke *et al.*, *EMBO J.* **20**, 777 (2001).
28. M. S. Campbell, G. K. T. Chan, T. J. Yen, *J. Cell Sci.* **114**, 953 (2001).
29. S. M. Elbashir *et al.*, *Nature* **411**, 494 (2001).
30. S. Martin-Lluesma, E. A. Nigg, unpublished observations.
31. A. Abrieu *et al.*, *Cell* **106**, 83 (2001).
32. V. M. Stucke, E. A. Nigg, unpublished observations.
33. T. M. Kapoor, T. U. Mayer, M. L. Coughlin, T. J. Mitchison, *J. Cell Biol.* **150**, 975 (2000).
34. X. B. Yao, A. Abrieu, Y. Zheng, K. F. Sullivan, D. W. Cleveland, *Nature Cell Biol.* **2**, 484 (2000).
35. G. K. Chan, S. A. Jablonski, V. Sudakin, J. C. Hittle, T. J. Yen, *J. Cell Biol.* **146**, 941 (1999).
36. B. F. McEwen *et al.*, *Mol. Biol. Cell* **12**, 2776 (2001).
37. C. Lengauer, K. W. Kinzler, B. Vogelstein, *Nature* **396**, 643 (1998).

38. We thank D. Brown, W. Earnshaw, J. Kilmartin, and T. U. Mayer for antibodies; F. Barr, S. Elledge, and K. T. Jeang for plasmids; F. Barr for advice on yeast work; and all members of the laboratory for helpful discussions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/297/5590/2267/DC1

Materials and Methods
Figs. S1 to S4
References and Notes

1 July 2001; accepted 13 August 2002

Gene Expression During the Life Cycle of *Drosophila melanogaster*

Michelle N. Arbeitman,[1*] Eileen E. M. Furlong,[2,3,5] Farhad Imam,[3,4*] Eric Johnson,[3,4,6*] Brian H. Null,[2,7*] Bruce S. Baker,[1] Mark A. Krasnow,[3,4] Matthew P. Scott,[2,4] Ronald W. Davis,[3,7] Kevin P. White[7,8†]

Molecular genetic studies of *Drosophila melanogaster* have led to profound advances in understanding the regulation of development. Here we report gene expression patterns for nearly one-third of all *Drosophila* genes during a complete time course of development. Mutations that eliminate eye or germline tissue were used to further analyze tissue-specific gene expression programs. These studies define major characteristics of the transcriptional programs that underlie the life cycle, compare development in males and females, and show that large-scale gene expression data collected from whole animals can be used to identify genes expressed in particular tissues and organs or genes involved in specific biological and biochemical processes.

Molecular studies of development in multicellular organisms have gone through two major phases during the past three decades. Initially, solution hybridization studies quantitated transcript abundance and showed that large-scale changes in gene expression accompany development (*1*). In *Drosophila*, such studies suggested that 5000 to 7000 different polyadenylated RNA species are produced at each stage of the life cycle and that the composition of this set of RNAs shifted during development (*1*). These analyses gave an overview of genome activity during development, but they could not follow the expression of individual genes or reveal their identities. Later, when it became possible to clone individual genes (*2*, *3*), RNA blots and in situ hybridization revealed when and where individual genes were active. This second phase of analysis allowed

27 September 2002

Science

Vol. 297 No. 5590
Pages 2157–2332 $9

American Association for the Advancement of Science

an initial determination of the links between molecules and developmental functions. This gene-by-gene approach has dominated developmental biology for the past two decades.

DNA microarrays extend the single-gene approach to the genome level by measuring the transcript levels of thousands of genes simultaneously (*4*–*6*). Here we present the transcriptional profiles for about one-third of all predicted *Drosophila* genes (*7*) throughout the life cycle, from fertilization to aging adults. cDNA microarrays were used to analyze the RNA expression levels of 4028 genes in wild-type flies examined during 66 sequential time periods beginning at fertilization and spanning the embryonic, larval, and pupal periods and the first 30 days of adulthood, when males and females were sampled separately (Fig. 1A). Early embryos change rapidly, so overlapping 1-hour periods were sampled; adults were sampled at multiday intervals (Fig. 1A) (*8*). We compared each experimental sample to a common reference sample made from pooled mRNA representing all stages of the life cycle, allowing us to measure each transcript's relative abundance (*8*). We refer to this relative abundance at each time as a gene's transcript or expression level, and to each gene's overall pattern of expression during development as its transcript or expression profile.

Expression of most genes assayed (3483 out of 4028, 86%) changed significantly [$P <$ 0.001, analysis of variance (ANOVA)] during the 40-day period surveyed (*8*). Of these, 3219 genes exhibited at least a fourfold difference between their highest and lowest levels of expression (Fig. 1B and table S1). The vast majority of these developmentally modulated genes (>88%) are expressed during the first 20 hours of development, before the end of embryogenesis (Fig. 1, B and C). To identify patterns of gene reexpression during development, we applied a peak-finding algorithm (*8*) to each gene's expression profile. We found that 36.3% of the genes (1169 genes) showed a single major peak of expression (Fig. 1D, left panels), whereas 40.3% (1298) showed two peaks (Fig. 1D, right panels) and 23.4% (752) showed three or more peaks (fig. S1 and tables S2 to S6).

Many genes are expressed in two waves during development, with embryonic expression patterns recapitulated in pupae and larval patterns recapitulated in adults. Genes with a first peak in their transcript level at the beginning (0 to 2.5 hours) of embryogenesis commonly have their second peaks during the larva-pupa transition, whereas genes with a first peak of expression at the end of embryogenesis (10 to 21 hours) commonly have their second peak during the late pupal period (Fig. 1E). When overall similarities in somatic gene expression between different developmental stages were compared by hierarchical clustering (*8*, *9*), expression patterns

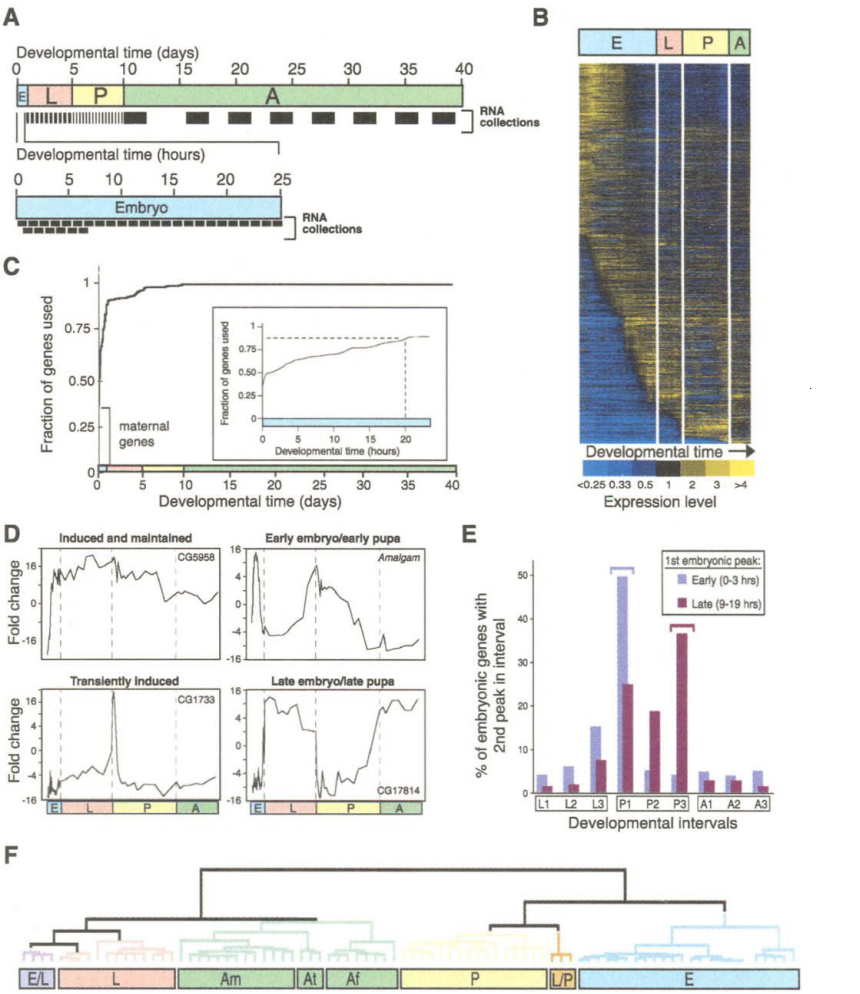

Fig. 1. Patterns of gene expression during development. (**A**) Whole-animal collections were made for embryos (E), larvae (L), pupae (P), and adults (A). Black bars indicate the periods of development that were sampled (*8*); for all stages, independent samples were collected in duplicate. (**B**) Gene expression profiles ordered by onset of their first increase in transcript abundance (*8*). Data for 3219 genes that change expression by more than fourfold during development ($P <$ 0.001, ANOVA) are shown. Each row represents data for one gene, and each column is a developmental time point, as indicated in (A). Expression level relative to the reference sample is indicated with color; blue indicates low levels and yellow indicates high levels. (**C**) Cumulative fraction of genes that exhibited a strong increase in transcript level over time. (**D**) Examples of common gene expression patterns. CG5958 (top left) shows induction in early embryogenesis and is maintained. CG1733 (bottom left) has a short peak of intense expression and is not expressed at other points in development. *Amalgam* (top right) is expressed in early embryogenesis and at the larval/pupal transition, whereas the late reinduced gene CG17814 (bottom right) shows a bimodal pattern in late embryo and late pupa. (**E**) Postembryonic reinduction of genes initially expressed in early and late embryos. Only the second, postembryonic onset of expression is shown. Genes with initial onset of expression in the first 3 hours after fertilization (0 to 3 hours, blue) are often reexpressed in early pupae (blue bracket), and genes with expression onset in the late embryo (9 to 19 hours after fertilization, purple) are often reexpressed in late pupae (purple bracket). (**F**) Hierarchical clustering of developmental time points on the basis of their pattern of somatic gene expression. Time points with highly correlated gene expression patterns are grouped adjacently. Embryo expression pattern group with those of pupae, and larvae expression patterns group with those of adults. Adult *tudor* (At), adult males (Am), adult females (Af), embryonic/larval transition (E/L), larval/pupal transition (L/P).

[1]Department of Biological Sciences, [2]Department of Developmental Biology and Department of Genetics, [3]Department of Biochemistry, [4]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA. [5]Developmental Biology Program, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [6]Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA. [7]Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94306, USA. [8]Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA.

*Co-first authors
†To whom correspondence should be addressed. E-mail: kevin.white@yale.edu

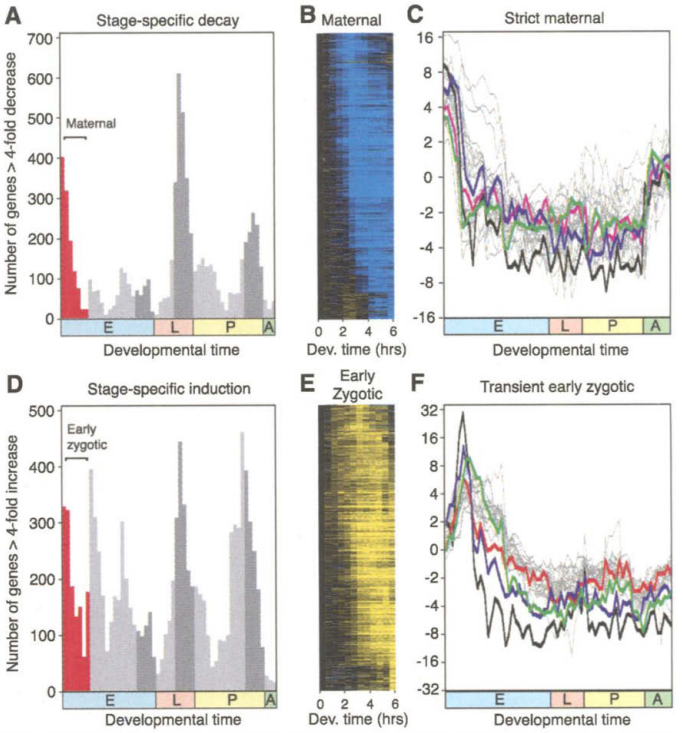

Fig. 2. Stage-specific changes in gene expression. (**A**) Patterns of stage-specific transcript level decline. Each bar represents the number of genes decreasing by more than fourfold within the four following time points when compared to their average in the previous two time points. Red bars correspond to the developmental interval shown in (B). Dark gray bars indicate intervals spanning major developmental transitions between stages. (**B**) Early expression profiles of the 322 maternal genes that decrease expression by more than threefold during the first 0 to 6.5 hours of embryonic development, arranged by one-dimensional SOM analysis (table S16). (**C**) Full expression profiles of the 27 strictly maternal genes identified using criteria optimized on a training set of known maternal genes and with a SOM analysis (*8*) (table S12). Selected genes are highlighted: *swallow* (blue), *fs(1)Ya* (pink), *cyclinJ* (green), and CG18543 (black), which has the most dramatic reduction in expression. (**D**) Patterns of stage-specific transcript level increase. Analysis as in (A), showing the number of genes induced above a fourfold threshold. Red bars correspond to the developmental interval shown in (E). (**E**) One-dimensional SOM analysis of 534 genes induced over 0 to 6.5 hours of embryonic development (table S18). (**F**) Early expression profiles of 21 transiently expressed zygotic genes identified using criteria optimized on a training set of known maternal and zygotic genes and by a SOM analysis (*8*) (table S20). Previously identified genes included *blastoderm-specific gene 25D* (red), CG9506 [*slam*, a gene required for polarized membrane growth during cellularization (*31*); blue], and *Sep5*, which encodes a septin-like protein (green). Among the 18 newly identified genes in this class is a *CG15634* (black), which displayed the most rapid induction and the highest levels of blastoderm-specific expression.

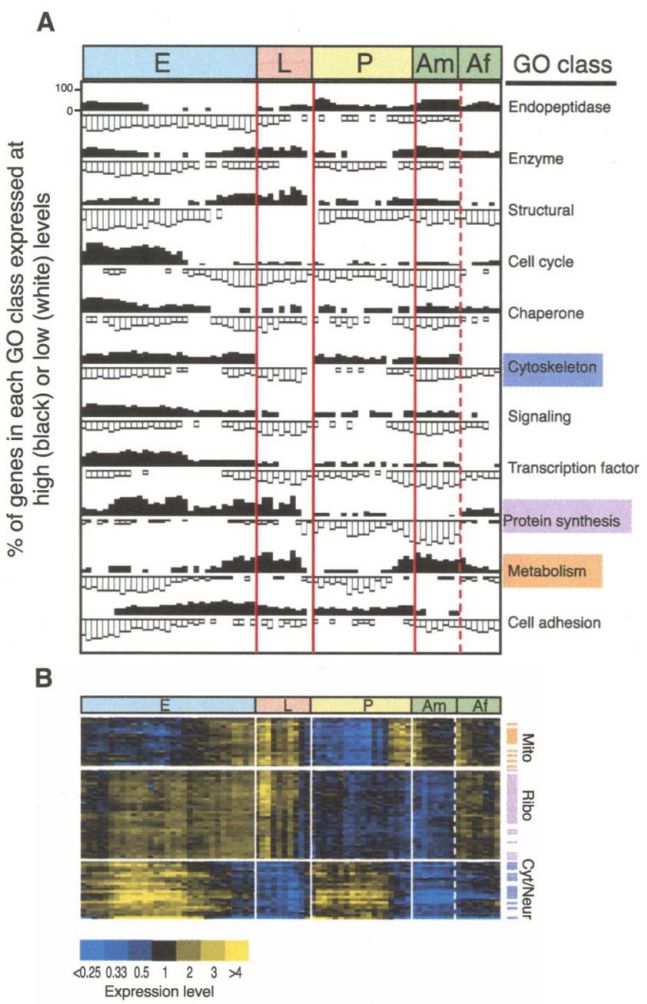

Fig. 3. Coordinate expression of genes encoding components of macromolecular complexes or involved in specific physiological processes. (**A**) For each GO class of protein, open bars below each line indicate the percentage of genes with low expression (bottom 25% of a gene's expression range during development), and filled bars above each line indicate the percentage of genes with high expression (top 25% of a gene's expression range). Colored GO classes correspond to clusters shown in (B). The scale (100% equals all genes in the GO class) is indicated for the endothelial class. (**B**) Three selected clusters of genes with similar expression profiles and related biological functions: components of mitochondria (Mito), ribosome (Ribo), and cytoskeletal/neural genes (Cyt/Neur). Genes within each cluster that are known to share a common biological function are indicated by a colored bar. Developmental stages as indicated in Fig. 1.

during embryonic stages were most similar to those of pupal time periods, and expression patterns during larval time periods were most similar to those of adult (Fig. 1F). Thus, despite morphological differences between developmental stages, disparate life stages share molecular commonalities.

We analyzed changes in gene expression during each major stage of development. The transcript levels of 2103 genes changed significantly ($P < 0.001$, ANOVA) during embryogenesis (table S7). A total of 445 genes changed during larval life (table S8), 646 during the pupal stage (table S9), and 118 during adult life (table S10) (*8*). The transcript levels of only 16 genes changed significantly ($P < 0.001$, ANOVA) between 5- and 30-day-old adults (table S11) (*8*). The transcript levels of hundreds of genes changed at least fourfold during five developmental periods that correspond to major morphological changes (the beginning, middle, and end of embryogenesis; the larval-pupal transition; and the end of the pupal period) (Fig. 2, A and D). Transcript levels changed much less during "morphologically quiescent" periods of early larval and adult life.

In the first hours of embryonic development between fertilization and gastrulation, gene expression is highly dynamic. Two broad categories of transcripts are present at this time: those deposited into the egg during oogenesis (produced by maternal genes) and those that are expressed only after fertilization (produced by zygotic genes). The expression profiles of 1212 genes were similar to those of known maternal genes (*8*), indi-

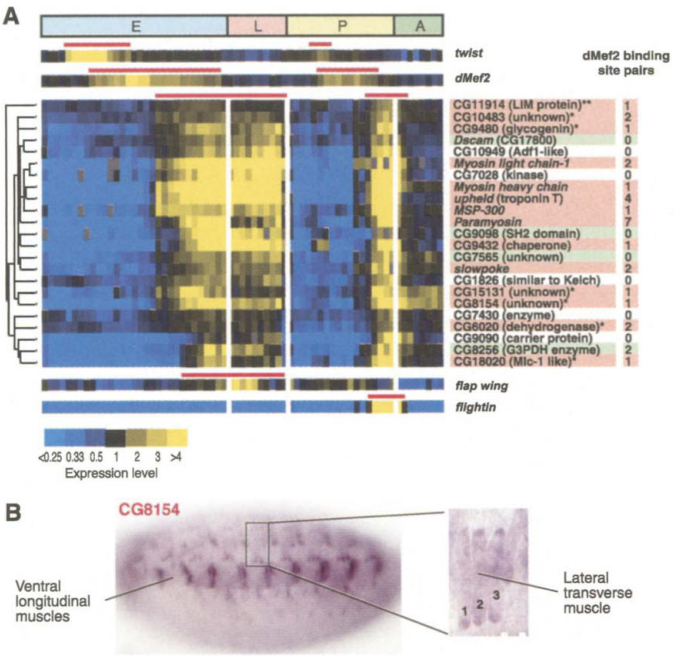

Fig. 4. Muscle differentiation. (**A**) A cluster enriched for genes expressed in terminally differentiated muscle (correlation coefficient of 0.862). Pink shading indicates genes that were either previously shown or shown here to be expressed in muscle (*confirmed by whole-mount in situ hybridization,**CG11914 was not tested but is predicted to be expressed in muscle on the basis of homology to muscle LIM proteins). Green shading indicates that in situ hybridization showed neuronal expression. The number of dMEF2 consensus binding site pairs in the vicinity of each gene is shown (*8*). Red bars highlight the sequential expression of the muscle gene regulatory hierarchy (*twist* > *dMef2* > terminal differentiation genes) during the embryonic development of larval muscles and again during the pupal development of adult muscle. Several known muscle differentiation genes on the array did not group with this cluster, but showed an expression pattern consistent with higher expression during the development of larval (e.g., *flap wing*) or adult (e.g., *flightin*) muscle (see also fig. S4). (Note: male and female adult data were averaged after clustering for display purposes.) (**B**) In situ hybridization showing expression of *CG8154* in ventral and lateral muscle fibers. Developmental stages as indicated in Fig. 1. Lateral transverse muscles are labeled 1, 2, and 3.

cating that at least 30% of the transcripts analyzed (1212 of 4028) are maternally deposited (tables S12 to S17). Although many maternal transcripts persisted during embryogenesis, 322 (27%) of the 1212 maternal gene transcripts decreased by at least threefold (Fig. 2B), and 36 (3%) decreased by 10-fold or more during the 6.5 hours after egg deposition (fig. S2) (*8*). A self-organizing map (SOM) algorithm (*10*), applied to the data from all 1212 maternally deposited genes, identified a cluster of 27 "strictly" maternal genes. Transcripts from almost all 27 of these genes were degraded after fertilization and were not subsequently expressed at high levels until they appeared in the female germ line during oogenesis (Fig. 2C). Of these, 5 were previously known "strictly" maternal genes and 22 were new (table S12).

Early zygotic genes were identified in a similar manner. A total of 534 genes have expression profiles similar to those of known early zygotic genes (Fig. 2E; tables S18 to S22 for zygotic gene lists) (*8*). Among these genes, 53 increased expression by at least 10-fold in the first 6.5 hours of development, 26 of which were previously characterized (fig. S2). Sixteen of these 26 genes are known to play critical roles in embryonic development and patterning. These include eight transcription factor genes (*invected*, *odd-paired*, *Antennapedia*, *tailless*, *bagpipe*, *prospero*, *ribbon*, and *grainyhead*), and genes encoding two signaling molecules (*wingless* and *decapentaplegic*), a signal transduction protein (*stumps*), a cell adhesion molecule (*neurotactin*), and a channel protein (*big brain*). The early developmental gene-regulatory hierarchy, including gap, pair-rule, segment polarity, and homeotic gene induction (*11*), was recapitulated in the microarray data. Sequence similarities suggest that the 27 uncharacterized, rapidly induced zygotic genes encode cell adhesion molecules (6 genes), channels and transporters (6 genes), metabolic and biosynthetic enzymes (5 genes), or kinases and phosphatases (4 genes). None of these newly identified genes have sequence similarity to transcription factors. Transient early zygotic ("blastoderm-specific") genes are expressed at high levels only during the critical period of development when cellularization of the syncitial blastoderm embryo occurs. SOM analysis of the expression patterns of early zygotic genes identified 21 such genes, including 3 previously known genes and 18 previously unknown ones (Fig. 2F, table S20).

We investigated whether genes with related biochemical functions are coordinately expressed during development. Genes encoding functionally related proteins were identified by gene ontology (GO) annotations, which classify genes according to the functions of their encoded proteins (*8*, *12*). Genes within a functional group tend to be expressed at similar times (Fig. 3A). For example, most cell cycle genes are expressed at high levels during the first 12 hours of development, when cell division is rapid, and few are expressed at high levels thereafter. In contrast, most metabolic genes are expressed at their highest levels only immediately before and during larval and adult life.

All 4028 genes were grouped by similarity of expression profile with a hierarchical clustering algorithm (*9*), and clusters of genes with similar expression profiles were examined for genes with related biochemical and cellular functions. Many examples of coexpressed genes that encode components of biochemical pathways or subunits of protein complexes were apparent, including genes not previously known to be developmentally regulated. Distinct clusters were enriched for genes encoding mitochondrial proteins, ribosomal proteins, cytoskeletal/neuronal factors, components of the 26*S* proteasome complex, the TCP-1 ring chaperonin complex, coatamer complex, vacuolar adenosine triphosphatases, and antimicrobial peptides (Fig. 3B and fig. S3). These results suggest that new components of biochemical complexes and cellular pathways in *Drosophila* can be identified by virtue of their similar expression profiles.

Clusters of coexpressed genes enriched for tissue-specific genes were also identified. One such cluster includes 23 genes, 8 of which were known to be expressed in terminally differentiated muscle (Fig. 4A). The genes in this group have a two-peak expression pattern that coincides with larval and adult muscle development (*13*). Larval muscle development is initiated in the embryo by the basic helix-loop-helix (bHLH) transcription factor Twist (*13*), which triggers transcription of *dMef2*, a gene encoding a MADS box transcription factor. dMef2 regulates the expression of muscle differentiation genes (*14*). This muscle regulatory hierarchy was recapitulated in the microarray data: The embryonic peak of *twist* transcript preceded that of *dMef2*, which preceded expression of the genes in a muscle differentiation cluster (Fig. 4A). The same sequence was repeated in the pupal period, indicating that the same regulatory hierarchy controls formation of adult muscle.

Fifteen of the 23 genes in this cluster (65%) contained pairs of predicted dMEF2-binding sites (*8*) (Fig. 4A). Only 5% of other genes on the array contain such pairs (*8*), so many of the genes in the cluster are likely to be direct targets of dMef2. Six of the seven previously uncharacterized genes in the cluster, all with dMef2-binding sites, were expressed in differentiated muscle (Fig. 4B). The seventh gene, and the two genes without

dMef2-binding sites that we tested, were expressed in the central nervous system (table S23). These three neural genes together with one previously known neural gene, *Down Syndrome Cell Adhesion Molecule* (*DSCAM*) (*15*), were activated synchronously with muscle genes and may be involved in neural events that are coordinated with muscle development, such as neuromuscular junction formation.

Hierarchical clustering analysis also revealed two large groups of coexpressed genes that encode either female- or male-enriched transcripts. These genes appear to be sex-specifically expressed in the germ line. When RNA from mutants lacking germline tissue [the adult progeny of *tudor* mothers, referred to as *tudor* mutants (*16*)] was analyzed, expression of nearly all genes in the putative male and female germline clusters was substantially reduced (Fig. 5A), demonstrating that these genes are expressed in the germ line or are dependent on the germ line for their expression (*8*). Indeed, nearly all of the male germline genes identified in the *tudor* mutant experiment were highly expressed in isolated testes (Fig. 5A). Increased expression of genes in the male cluster (249 genes) (Fig. 5A and table S25) began at the larva-pupa transition and remained high thereafter (Fig. 5A), coincident with meiosis and spermatogenesis in the male germ line (*17*, *8*). Increased expression of genes in the female cluster (1245 genes) (Fig. 5A and table S24) began in 0- to 24-hour adults and continued thereafter (Fig. 5A), coincident with oogenesis (*18*). Transcripts of most (77%) of the genes in this cluster were present at high levels in early embryos before zygotic transcription began (Fig. 5A), implying that they are maternally provided. RNA blot analysis confirmed sex-specific germline expression of two selected genes in each class (fig. S5).

Analysis of the *tudor* data also led to the identification of 111 genes that were expressed in both male and female germ lines, because they were expressed in wild-type adults of both sexes but markedly reduced in *tudor* mutants (Fig. 5A and tables S26 to S28) (*8*). Among these 111 genes are known germline factors common to both sexes such as *exu* (*19*) and *benign gonial neoplasm* (*20*), whereas dozens of others remain to be characterized. Together, these analyses increase the number of male and female germline genes by an order of magnitude or more and demonstrate a previously unrecognized temporal coordination of germline genes in both sexes.

We identified sex-specific somatic genes by comparing transcript levels in female and male *tudor* adults. We found that 31 genes had significantly higher expression in the soma of adult females compared with 37 genes in males (*8*). The male and female somatic gene sets (Fig. 5B) include the previously identified sex-specific *Yolk protein 1* gene [female (*21*)] and an accessory gland protein gene *Acp 36DE* [male (*22*)]. The rest of the genes in these sets are likely also to be involved in sex-specific adult physiology or function (tables S29 and S30).

Hierarchical clustering identified a small adult-specific set of genes, some of which encode known eye-specific proteins. Using RNA from *eyes absent* mutants, we refined this set to 33 genes that included 11 known eye differentiation genes, many of which function in phototransduction (Fig. 5C) (*8*). Some of the newly identified eye genes may also function in phototransduction, based on the inferred biochemical functions of the encoded proteins. For example, CG10233 and CG3573 encode a putative phosphatidylinositol-4 phosphate 5-kinase and a putative inositol 1,4,5-trisphosphate 5-phosphatase, respectively, and thus may regulate the level of $PtdIns(4,5)P_2$, a key second messenger in invertebrate phototransduction (*23*).

Hierarchical clusters were examined for biases in the proportion of genes with highly conserved human homologs or for fly-specific genes (*8*). Sixteen of the 20 largest clusters had no significant bias ($P > 0.01$) in the relative proportions of conserved or fly-specific genes (fig. S6). Two clusters were significantly enriched ($P < 0.001$) for fly-specific genes: a cluster of male germline genes and a cluster of genes expressed in larvae that encode peptide hormones, peptidases, and peritrophins. Two other clusters were significantly enriched ($P < 0.001$) for conserved genes. One of these contained many ribosomal genes (Fig. 3B) and the other included a group of 35 zygotically activated genes, 24 of which are highly conserved. This latter cluster includes *Hox* genes, *wingless*, *dpp*, and several other factors involved in developmental processes shared among metazoans.

Genes that encode homologs of human disease proteins were analyzed to determine whether any disease gene homologs were co-expressed with other genes of related function. More than three-quarters of human disease genes have *Drosophila* homologs (*25*, *26*); 240 were present in this data set (*27*). These homologs were dispersed throughout many clusters. One example cluster containing 21 co-expressed genes, including *dPresenilin* and *dNicastrin*, homologs of two subunits of a proteolytic processing complex im-

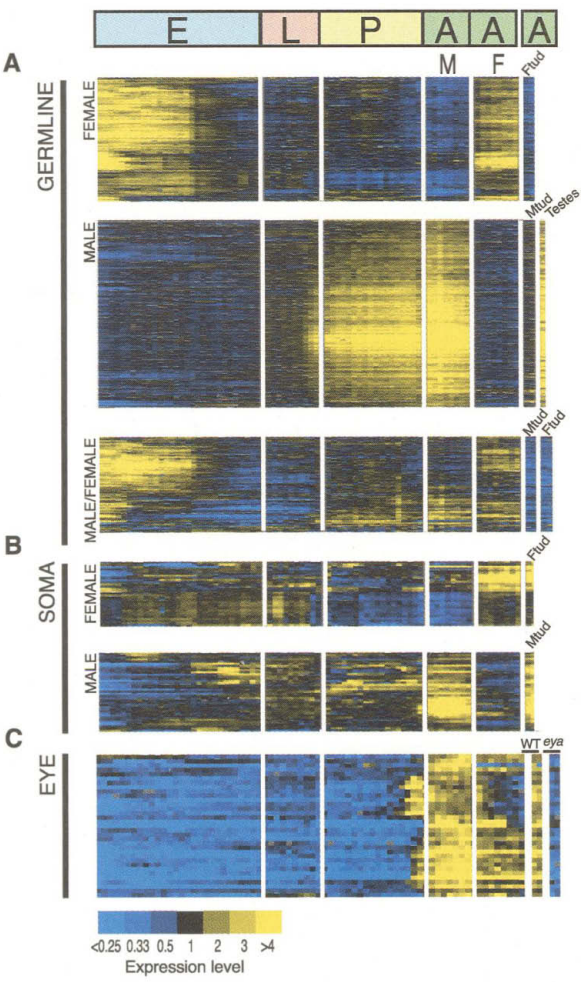

Fig. 5. Sex-enriched germline and somatic genes, and eye differentiation genes. **(A)** Expression profiles of clusters of genes enriched in the female or male germ line, or both (*8*). Female (144) and male (215) germline genes were identified in the hierarchical cluster of the full data set (fig. S7); those with a three-fold or greater difference in expression between adult males and females are shown. Developmental stages are as indicated in Fig. 1. M, adult male; F, adult female; Mtud, adult male *tudor* (0- to 24-hour and 5-day adult time points); Ftud, adult female *tudor* (0- to 24-hour and 5-day adult time points); testes were dissected from adults. **(B)** Clusters of genes enriched in female and male somatic tissue (*8*). **(C)** Eye differentiation genes. Hierarchical cluster of the 33 adult-enriched genes whose expression diminished in *eya* mutants (*8*).

plicated in Alzheimer's disease (Fig. 3B, cytoskeletal/neuronal cluster). Most of the other known genes in this cluster are implicated in neuronal pathfinding and cell adhesion, including *E-cadherin*, which encodes a protein associated with the presenilin complex (*28*), and *Notch*, which encodes a substrate of the presenilin complex (*29*, *30*). The cluster of 21 genes is enriched for components and substrates of the presenilin complex.

These data (*24*) provide an overview of gene expression profiles during *Drosophila* development. An unusually high proportion of the genes are developmentally regulated, but of 4028 genes analyzed, only 903 are previously named *Drosophila* genes with a known mutant phenotype, biochemical function, or protein homology. Fifty-one percent of the genes fall into 50 clusters with correlation coefficients greater than 0.80 (for an annotated hierarchical cluster, see fig. S7, green bars). Virtually all the clusters contain genes with known or predicted roles in development or physiology, and genes to which a biochemical or cellular function has been assigned by the GO project (*12*) [all genes in these clusters are listed in the online database (*24*)]. A large number of the clusters contain genes that are used together in specific developmental or biochemical processes. On the basis of their developmental expression patterns, we have tentatively assigned 53% of the genes to a developmental or biological functional category (for example, male germ line, female germ line, eye, muscle, early zygotic, biochemical complex, or cell biology function).

In addition to providing functional annotation of the *Drosophila* genome, these studies are a step toward a complete description of the genetic networks that control development.

References and Notes

1. L. S. Levy, J. E. Manning, *Dev. Biol.* **85**, 141 (1981).
2. M. Grunstein, D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961 (1975).
3. W. Bender, P. Spierer, D. S. Hogness, *J. Mol. Biol.* **168**, 17 (1983).
4. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
5. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995).
6. K. P. White, S. A. Rifkin, P. Hurban, D. S. Hogness, *Science* **286**, 2179 (1999).
7. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
8. Materials and methods are available as supporting material on *Science* Online.
9. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
10. P. Tamayo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2907 (1999).
11. M. Akam, *Development* **101**, 1 (1987).
12. M. Ashburner *et al.*, *Nature Genet.* **25**, 25 (2000).
13. M. Bate, in *The Development of* Drosophila melanogaster, A. M. A. Michael Bate, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1993), vol. II, pp. 1013–1090.
14. B. L. Black, E. N. Olson, *Annu. Rev. Cell. Dev. Biol.* **14**, 167 (1998).
15. D. Schmucker *et al.*, *Cell* **101**, 671 (2000).
16. R. E. Boswell, A. P. Mahowald, *Cell* **43**, 97 (1985).
17. D. Lindsley, K. T. Tokuyasu, in *Genetics and Biology of* Drosophila, M. Ashburner, T. R. Wright, Eds. (Academic Press, New York, 1980), pp. 225–294.
18. A. C. Spradling, in *The Development of* Drosophila melanogaster, A. M. A. Michael Bate, Ed. (Cold Harbor Laboratory, Cold Spring Harbor, NY, 1993), pp. 1–70.
19. T. Hazelrigg *et al.*, *Genetics* **126**, 607 (1990).
20. E. Gateff, *Prog. Clin. Biol. Res.* **85** (part B), 621 (1982).
21. T. Barnett, C. Pachl, J. P. Gergen, P. C. Wensink, *Cell* **21**, 729 (1980).
22. M. J. Bertram, D. M. Neubaum, M. F. Wolfner, *Insect Biochem. Mol. Biol.* **26**, 971 (1996).
23. C. S. Zuker, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 571 (1996).
24. The data are available at http://flygenome.yale.edu/Lifecycle.
25. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
26. L. T. Reiter, L. Potocki, S. Chien, M. Gribskov, E. Bier, *Genome Res.* **11**, 1114 (2001).
27. A complete annotated listing of the transcript profiles of human disease gene homologs is available in the database supplement (*24*).
28. L Baki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2381 (2001).
29. G. Struhl, I. Greenwald, *Nature* **398**, 522 (1999).
30. Y. Ye, N. Lukinova, M. E. Fortini, *Nature* **398**, 525 (1999).
31. T. Lecuit, R. Samanta, E. Wieschaus, *Dev. Cell* **2**, 425 (2002).

32. We thank Berkeley *Drosophila* Genome Project/Howard Hughes Medical Institute (HHMI) expressed sequence tag (EST) sequencing project and Research Genetics for providing the EST library, P. Lem and G. Gibson for resequencing the EST library, C. Fan for technical assistance, the Minx Fuller lab for testes RNA, T. Jones for assistance in EST data management and analysis, and I. SanGil for database support. M.P.S., E.E.M.F., and B.H.N. were supported by a Defense Advanced Research Projects grant. M.A.K. and M.P.S. are investigators of the HHMI. F.I. was supported by an NIH Medical Scientist Training Program fellowship. B.S.B. and M.A.K. acknowledge support from National Institute of General Medical Studes (NIGMS) and National Institute of Neurological Disorders and Stroke. M.N.A. acknowledges support from NIGMS. R.W.D. is supported by the National Human Genome Research Institute (NHGRI). Supported by a grant from the NHGRI to K.P.W.

Supporting Online Material

www.sciencemag.org/cgi/content/full/297/5590/2270/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S30

29 April 2002; accepted 8 August 2002

Structural Basis for Gluten Intolerance in Celiac Sprue

Lu Shan,[1] Øyvind Molberg,[5] Isabelle Parrot,[1] Felix Hausch,[1] Ferda Filiz,[1] Gary M. Gray,[2] Ludvig M. Sollid,[5] Chaitan Khosla[1,3,4*]

Celiac Sprue, a widely prevalent autoimmune disease of the small intestine, is induced in genetically susceptible individuals by exposure to dietary gluten. A 33-mer peptide was identified that has several characteristics suggesting it is the primary initiator of the inflammatory response to gluten in Celiac Sprue patients. In vitro and in vivo studies in rats and humans demonstrated that it is stable toward breakdown by all gastric, pancreatic, and intestinal brush-border membrane proteases. The peptide reacted with tissue transglutaminase, the major autoantigen in Celiac Sprue, with substantially greater selectivity than known natural substrates of this extracellular enzyme. It was a potent inducer of gut-derived human T cell lines from 14 of 14 Celiac Sprue patients. Homologs of this peptide were found in all food grains that are toxic to Celiac Sprue patients but are absent from all nontoxic food grains. The peptide could be detoxified in in vitro and in vivo assays by exposure to a bacterial prolyl endopeptidase, suggesting a strategy for oral peptidase supplement therapy for Celiac Sprue.

Celiac Sprue (also known as Celiac disease or gluten-sensitive enteropathy) is an autoimmune disease of the small intestine caused by the ingestion of gluten proteins from widely prevalent food sources such as wheat, rye, and barley. In many human leukocyte antigen (HLA) DQ2 (or DQ8)–positive individuals, exposure of the small intestine to gluten induces an inflammatory response, leading to destruction of the villous structure of the intestine (*1*–*3*). It commonly appears in early childhood, with severe symptoms including chronic diarrhea, abdominal distension, and failure to thrive. In many patients, symptoms may not develop until later in life, when the disease symptoms include fatigue, diarrhea, and weight loss due to malabsorption, anemia, and neurological symptoms. Celiac Sprue is a lifelong disease, and if untreated it is associated with increased morbidity and mortality (*4*, *5*). Despite its high prevalence in most population groups (>1:200) and serious manifestations, the only effective therapy is strict dietary abstinence from these food grains.

[1]Department of Chemical Engineering, [2]Department of Medicine, [3]Department of Chemistry, and [4]Department of Biochemistry, Stanford University, Stanford, CA 94305–5025, USA. [5]Institute of Immunology, Rikshospitalet, University of Oslo, N-0027 Oslo, Norway.

*To whom correspondence should be addressed. E-mail: ck@chemeng.stanford.edu