

RESEARCH ARTICLE

A Gene Expression Map for the Euchromatic Genome of *Drosophila melanogaster*

Viktor Stolc,^{1,5*} Zareen Gauhar,^{1,2*} Christopher Mason,^{2*}
Gabor Halasz,⁷ Marinus F. van Batenburg,^{7,9} Scott A. Rifkin,^{2,3}
Sujun Hua,² Tine Herreman,² Waraporn Tongprasit,⁶
Paolo Emilio Barbano,^{2,4} Harmen J. Bussemaker,^{7,8}
Kevin P. White^{2,3,†}

We used a maskless photolithography method to produce DNA oligonucleotide microarrays with unique probe sequences tiled throughout the genome of *Drosophila melanogaster* and across predicted splice junctions. RNA expression of protein coding and nonprotein coding sequences was determined for each major stage of the life cycle, including adult males and females. We detected transcriptional activity for 93% of annotated genes and RNA expression for 41% of the probes in intronic and intergenic sequences. Comparison to genome-wide RNA interference data and to gene annotations revealed distinguishable levels of expression for different classes of genes and higher levels of expression for genes with essential cellular functions. Differential splicing was observed in about 40% of predicted genes, and 5440 previously unknown splice forms were detected. Genes within conserved regions of synteny with *D. pseudoobscura* had highly correlated expression; these regions ranged in length from 10 to 900 kilobase pairs. The expressed intergenic and intronic sequences are more likely to be evolutionarily conserved than nonexpressed ones, and about 15% of them appear to be developmentally regulated. Our results provide a draft expression map for the entire nonrepetitive genome, which reveals a much more extensive and diverse set of expressed sequences than was previously predicted.

Characterization of the complete expressed set of RNA sequences is central to the functional interpretation of each genome. For almost 3 decades, the analysis of the *Drosophila* genome has served as an important model for studying the relationship between gene expression and development. In recent years, *Drosophila* provided the initial demonstration that DNA microarrays could be used to study gene expression during development (1), and subsequent large-scale studies of gene expression in this and other developmental model organisms have given new insights into how

organisms implement their developmental plan (2). Additional studies in *Drosophila* have addressed questions in diverse biological processes, including metamorphosis, aging, innate immune response, and sexual dimorphism, and in evolutionary patterns of gene expression (3–9). The cumulative data from such microarray studies are valuable for functional annotation of the genome. However, all microarray studies of gene expression in *Drosophila* have been restricted to fragments of genes from predicted gene sets and thus are subject to the limitations of gene prediction algorithms. Large-scale expressed sequence tag (EST) sequencing has also been extremely valuable for annotation of the *Drosophila* genome (10), but this approach is limited by biases due to 5' or 3' end DNA sequencing, the extent to which transcripts are represented in cDNA libraries, and the number of EST sequences generated. A nonbiased approach is necessary for the determination of the entire catalog of expressed sequences in the genome. Genomic "tiling" DNA arrays, which use oligonucleotides or polymerase chain reaction (PCR) products corresponding to chromosomal sequences as probes, have been used to create transcript activity catalogs for portions

of the human genome and for *Arabidopsis* (11–13). Microarrays have also recently been used to characterize the great diversity of RNA transcripts brought about by differential splicing in human tissues (14). We used both types of approaches to characterize the *Drosophila* genome.

Experimental design. To determine the expressed portion of the *Drosophila* genome, we designed high-density oligonucleotide microarrays with probes for each predicted exon and probes tiled throughout the predicted intronic and intergenic regions of the genome. We used maskless array synthesizer (MAS) technology (15, 16) to synthesize custom microarrays containing 179,972 unique 36-nucleotide (nt) probes (17). Of these, 61,371 exon probes (EPs) assayed 52,888 exons from 13,197 predicted genes, 87,814 nonexon probes (NEPs) assayed expression from intronic and intergenic regions, and 30,787 splice junction probes (SJPs) assayed potential exon junctions for a test subset of 3955 genes. For the SJPs, we used 36-nt probes spanning each predicted splice junction, with 18 nt corresponding to each exon (14). RNA from six developmental stages during the *Drosophila* life cycle (early embryos, late embryos, larvae, pupae, and male and female adults) was isolated and reverse-transcribed in the presence of oligodeoxythymidine and random hexamers, and the labeled cDNA was hybridized to these arrays. The stages were chosen to maximize the number of transcripts that would be differentially expressed between samples on the basis of previous results (3, 7). Each sample was hybridized four times, twice with Cy5 labeling and twice with Cy3 labeling (fig. S1).

Genomic and chromosomal expression patterns. We determined which exon or nonexon probes correspond to genomic regions that are transcribed at any stage during development (18). We used a negative control probe (NCP) distribution (fig. S3) to score the statistical significance of the EP or NEP signal intensities for each of the 24 unique combinations of stage, dye, and array, correcting for probe sequence bias (17, 19). These results were combined into a single expression-level estimate (19), a threshold for which was determined by requiring a false discovery rate of 5% (20). This threshold shows 47,419 of 61,371 EPs (77%) and 35,985 out of 87,814 NEPs (41%) were significantly expressed at some point during the fly life cycle. Significantly expressed EPs correspond to 79% (41,559/52,888) of all exons probed and 93% (12,305/13,197) of all probed gene annotations. Our results confirmed 2426 annotated genes not yet validated through an EST sequence (Fig. 1A). Out of 10,280 genes represented by EST sequences,

¹Department of Molecular, Cellular, and Developmental Biology, ²Department of Genetics, ³Department of Ecology and Evolutionary Biology, ⁴Department of Mathematics, Yale University, New Haven, CT 06520, USA. ⁵Genome Research Facility, NASA Ames Research Center, Mail Stop 239-11, Moffett Field, CA 94035, USA. ⁶Eloret Corporation, Sunnyvale, CA 94087, USA. ⁷Department of Biological Sciences, ⁸Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10027, USA. ⁹Bioinformatics Laboratory, Academic Medical Center, University of Amsterdam, 1100 DE Amsterdam, Netherlands.

*These authors contributed equally to this work.

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

only 401 (3.0%) were not detected in these microarray experiments. Our finding that a large fraction of intergenic and intronic regions (NEPs) is expressed in *D. melanogaster* mirrors similar observations for chromosomes 21 and 22 in humans (16) and for *Arabidopsis* (14). These results support the conclusion that extensive expression of intergenic and intronic sequences occurs in the major evolutionary lineages of animals (deuterostomes and protostomes) and in plants.

We noted that mRNA expression levels for protein-encoding genes varied with the protein function assigned in the *Drosophila* Gene Ontology (fig. S2) (21). For example, genes encoding G protein receptors were expressed at relatively low levels, whereas genes encoding ribosomal proteins were highly expressed. A gene's expression level was also associated with cellular compartmentalization and the biological process it mediates (fig. S2). For example, genes encoding cytosolic and cytoskeletal factors were more highly expressed than those predicted to function within organelles such as the endoplasmic reticulum, Golgi, and peroxisome. To determine whether a high level of gene expression was associated with essential genetic functions, we examined the expression levels of genes recently shown to be required for cell viability (Fig. 1B) in a genome-wide RNA interference (RNAi) screen in *Drosophila* (22). Compared to the rest of the genome, the genes identified as essential by RNAi showed a significant increase in expression during all stages of development ($P = 0.0009$, t test), even when the highly expressed ribosomal protein genes were omitted ($P = 0.0005$, t test). This result is also consistent with the observation that genes with mutant phenotypes from the 3-Mbase Adh genomic region are overrepresented in EST libraries (23). High levels of essential gene expression may in part reflect widespread expression in cells throughout the animal, and the relative RNA expression level may serve as a rough predictor of essential cellular function.

We also examined changes in gene expression during the fly life cycle to determine what fraction of the entire genome is differentially expressed between developmental stages. Figure 2A shows the expression signal intensities of transcripts from a typical 50-kilobase pair (kbp) region of the *Drosophila* genome during each major developmental stage. Stage-specific variation in expression is observed not only for exon probes, as expected, but also for intergenic and intronic probes. We used analysis of variance (ANOVA) (24) to systematically identify probes as differentially expressed at a false discovery rate of 5% (16). As expected, the majority of probes detecting differentially expressed sequences are also expressed above background noise level (89% of EPs and 81% of NEPs) (17) (Table 1). We found 27,176

EPs to be differentially expressed, corresponding to 76% of annotated genes, and even more when we applied a less conservative background model (fig. S4). The fact that the majority of genes are developmentally regulated is consistent with previous results obtained with cDNA microarrays (3). In intergenic and intronic regions, we detected differential expression for 15% (5508/35,985) of significantly expressed NEPs, indicating that many of the putative noncoding expressed transcripts are also developmentally regulated.

By examining gene expression levels along chromosomes, we discovered that genes that have remained linked during evolution are correlated with one another in expression levels. We considered the genes inside 827 syntenic blocks identified between *D. melanogaster* and *D. pseudoobscura* (25), whose

expression varies substantially during development (Fig. 2A). We computed the average pairwise expression correlation between genes within each syntenic block and compared with the correlations among nonsyntenic genes. Average correlations among genes within 729 (88%) of these syntenic blocks showed a positive value (Fig. 2B), and genes within 369 syntenic blocks (45%) across all chromosomes showed positive correlations that were significant ($P < 0.05$) (Fig. 2C). Larger blocks show higher significance of correlated gene expression levels than smaller ones (Fig. 2D). Expression correlations for 104 out of 108 syntenic blocks greater than 300 kbp in length were statistically significant, and all expression level correlations among genes in blocks greater than 650 kbp were significant (fig. S5). These results indicate that chromatin domains

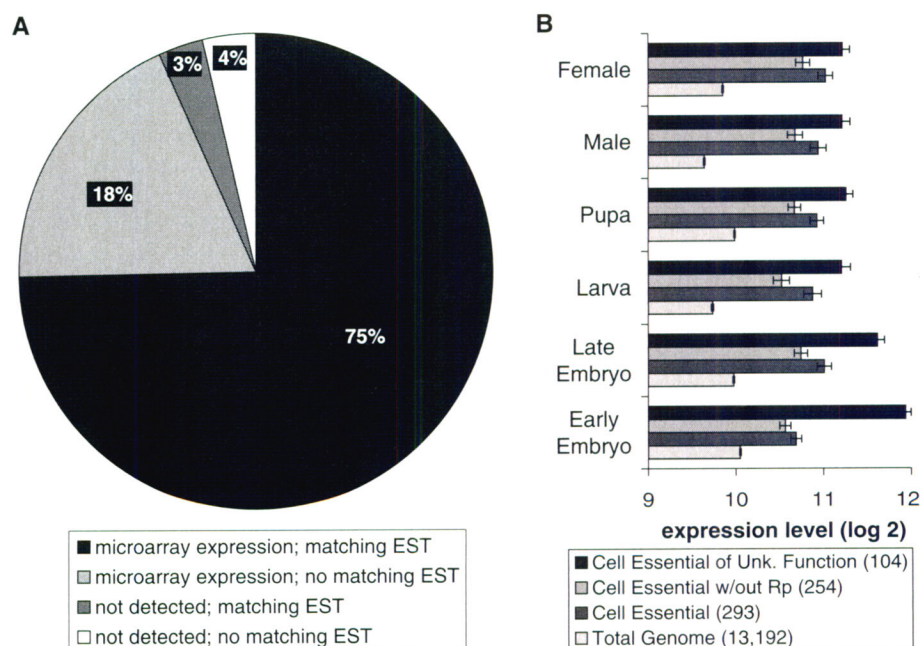


Fig. 1. Characterization of microarray probes by global expression levels can predict biological function. **(A)** Annotation confirmation. Each probe was compared to the *Drosophila* genome annotation version 3.1 (v3.1) and to cDNA/EST sequences produced by the Berkeley *Drosophila* Genome Project. 74.9% of v3.1 genes were confirmed by cDNA/EST sequencing and by this microarray analysis, 18.4% were confirmed solely by this study, 3.0% were confirmed solely by cDNA/EST sequencing, and 3.7% were unconfirmed by either method. **(B)** Cell-essential genes are expressed at higher-than-average levels. Compared to the average genome-wide, the 293 known genes identified as essential by Boutros *et al.* (23) showed significantly higher levels of expression during all stages of development ($P = 0.0009$, t test). A similar result was obtained when the highly expressed ribosomal protein (Rp) genes are omitted ($P = 0.0005$, t test) or when only the 104 cell-essential genes of unknown function ($P = 0.004$, t test) were examined. Error bars indicate ± 1 SE.

Table 1. Genome-wide statistics for expressed probes. We compiled lists of probes that show significant expression on the basis of two distinct criteria: (i) absolute probe expression above background (PEAB) noise level in one or more stages based on comparison with negative control probes and (ii) differential expression between stages based on ANOVA (16). FDR, false discovery rate.

Probe type	Total probes	Significant probes (FDR = 0.05)		Overlap: ANOVA and PEAB
		PEAB	ANOVA	
NEPs	87,814	35,985	6789	5508 (81%)
EPs	61,371	47,419	27,176	24,062 (89%)

or long-range enhancers act across wide genomic regions and that genomic rearrangements in these regions have been constrained during the evolutionary lineages leading to these two species, thereby maintaining these blocks of synteny. Natural selection may therefore act to maintain genes within neighborhoods where expression is coordinately regulated.

Exon expression and splicing. Within each gene, expression levels of exon pairs were typically highly correlated, and the distribution of these correlations was significantly shifted relative to that of exon pairs from different genes (Fig. 3A). A pattern separation algorithm was used to determine the patterns of expression for exons within each gene during development (17). We found three major trends (Fig. 3, B to D): 53% of genes showed uniform or highly correlated ($r > 0.8$) expression (Fig. 3B), 46% of genes showed multiple patterns of exon expression suggesting alternative promoter usage or splicing (Fig. 3C), and 1% of genes showed multiple patterns with at least one exon pair showing strong anticorrelation during the life cycle (Fig. 3D). These strong anticorrelations suggest exclusivity in the use of one exon or another for this small subset of genes. Together, this initial

analysis indicated that a vast amount of gene expression variation is missed in previous microarray studies that have used cDNAs or that assay only a subset of exons from each gene.

However, although these initial analyses allowed identification of exons that are differentially expressed during development, they did not reveal precisely which exons are spliced to one another during posttranscriptional RNA processing. To estimate the extent of splicing in the *Drosophila* genome during development, we used SJPs to directly assay spliced exons and differential use of splice isoforms. We focused on a subset of 3955 predicted genes that included genes with as few as 2 and as many as 54 exons. For these genes, we designed SJPs for all theoretically possible splicing combinations. Because SJPs could potentially hybridize to exons that were not spliced directly together, we used a specialized set of wrong junction probes (WJPs, splice junctions formed from exon segments but that do not match any possible transcript) as negative controls (fig. S6). These WJPs show higher hybridization signal than the NCPs we used as a reference distribution to detect absolute expression because of partial hybridization. We find that 28% (8732) of the 30,787

SJPs are expressed at a level above the background level defined by the WJPs.

By examining the ratio of significantly expressed, sequentially spliced exons and significantly expressed exons that are noncontiguous, we were able to determine the proportion of exons and genes that are alternatively spliced during development. The ratio of noncontiguous to contiguous splicing (NC/C ratio) for exons' use of downstream exons holds near a constant 0.40 (Fig. 4A), indicating an average of 2.5 contiguous splice events for every exon skipping (i.e., alternative) splice form. Also, these SJP data show that 53% (1374 of 2606) of expressed *Drosophila* genes from multi-exon genes exhibit exon skipping (18).

To determine the extent of splicing that is captured by large-scale EST sequencing, we used BLASTn to align sequences from the junctions' probes to known EST and cDNA databases (Fig. 4B). We identified 3292 splice junctions in the databases that matched our positive hybridization results, whereas 3464 junction probes did not detect hybridization ("EST only" sequences) but did match the expression databases, providing a very high false negative rate of ~46%. Nevertheless, 5440 new splice variants were identified

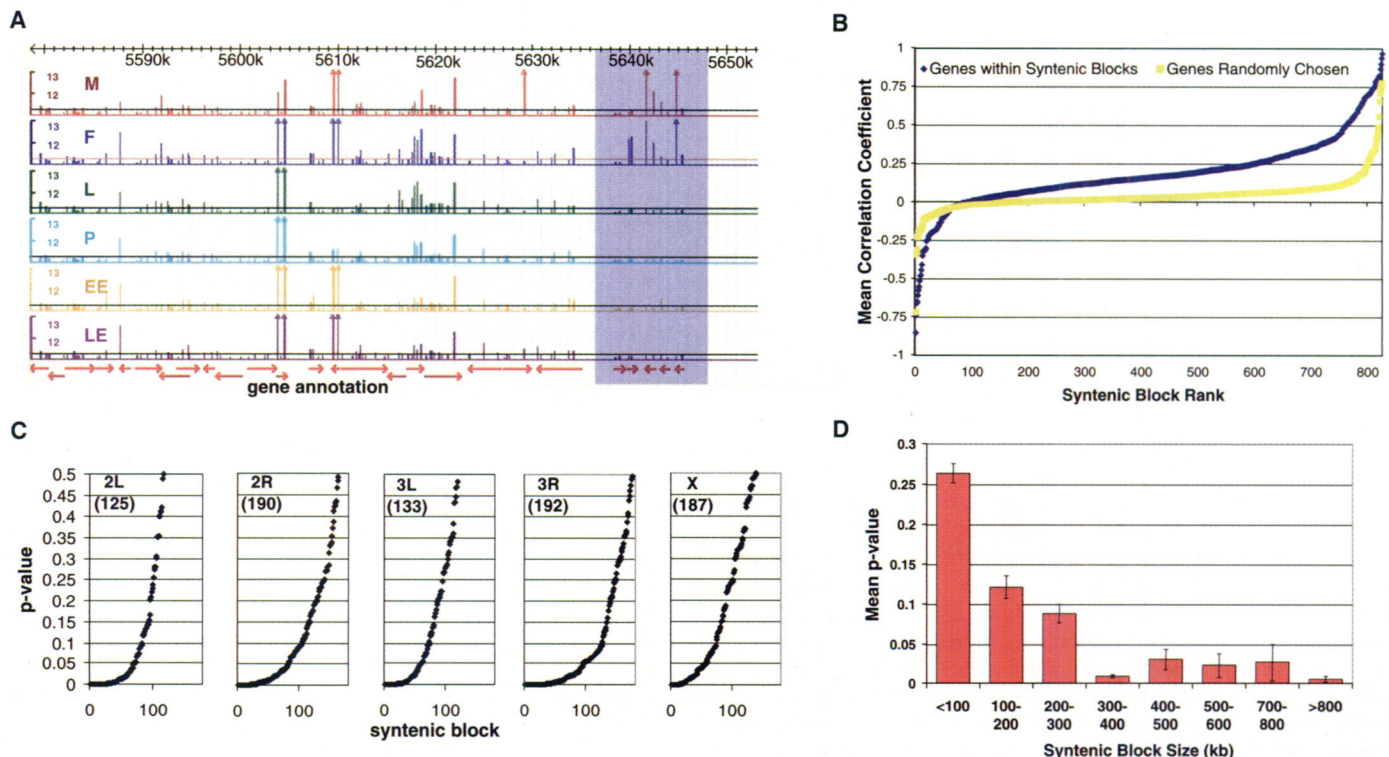
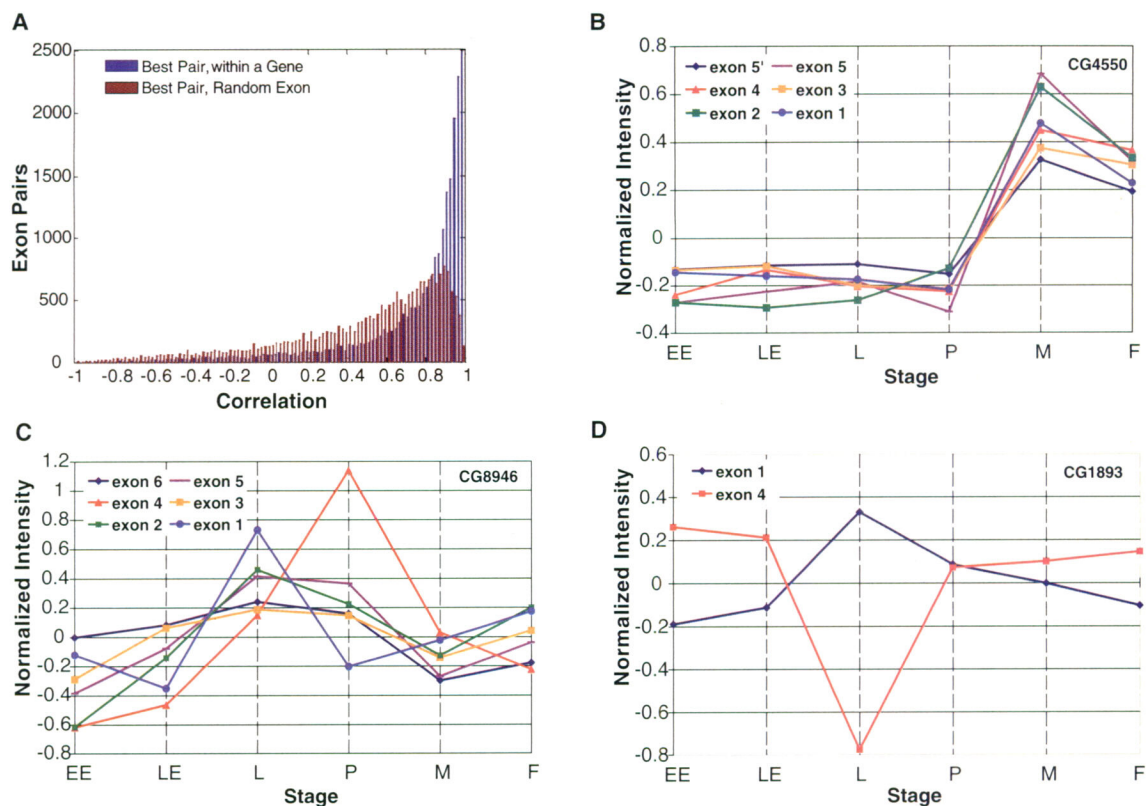


Fig. 2. Gene activity in conserved chromosomal domains. (A) Expression plots along chromosome 2L in the region: 5,600,000 to 5,650,000 show signal intensity of probes along the chromosome arm in six developmental stages [early embryo (EE), late embryo (LE), larva (L), pupa (P), adult male (M), and adult female (F)]. The shaded region shows a syntenic block where expression levels of transcripts encoded by genes within the block are highly correlated across development. (B) Distribution of Pearson correlation coefficients. Most syntenic blocks show positive correlations among expression levels for genes within them (blue), and the distribution

of correlations differs from a control set of randomly selected gene blocks (yellow). Most correlations in the control set are near zero, as expected. (C) Significance of gene expression correlations within syntenic blocks. The distributions of *P* values are shown by chromosome, showing the significance of the correlations compared to the control set, for each syntenic block. (D) Significance of gene expression correlations increases with syntenic block size. The larger the size of syntenic regions, the stronger the bias for genes within the block to be significantly correlated in expression. Error bars indicate ± 1 SE.

Fig. 3. Exon expression within and between genes. **(A)** Histogram of exon pairs' strongest correlation within a gene. There is a strong bias for exons expression levels within a gene (blue) to be positively correlated, compared with exons chosen at random (red). **(B)** Exon plot of the *ninaE* gene (CG4550). All exons are expressed similarly within this gene. Exon 5' is a second probe placed within exon 5, because it is a large exon. To normalize the data from the different EPs, we show the absolute expression values minus their mean and divided by the sum of the squares. **(C)** Exon plot of gene CG8946, an example of a gene with differential exon expression during development. Exon 1 peaks in larva, whereas exon 4 peaks in pupa, but all other exons plateau at those same stages at a much lower relative level. **(D)** Two anticorrelated exons within gene CG1893.



(Fig. 4B). Most of these (4564) were alternative splice junctions, indicating that EST sequencing missed the vast majority of alternatively spliced transcripts. Taken together, the exon-specific expression patterns and the splice junction expression patterns significantly extend the functional annotation of the predicted genes in the *Drosophila* genome.

Intergenic and intronic expression.

Lastly, we further examined the patterns of RNA expression from nonexonic sequences. There are several reasons to expect that a significant fraction of the annotated non-coding genome is expressed. First, previous studies performed in *Drosophila* using reverse Northern methodologies on chromosomal walks have identified multiple noncoding RNA transcripts at certain loci, such as in the bithorax complex (26–28). Second, the algorithms and experimental methods used for gene prediction and annotation may have not exhaustively identified the entire gene complement or all of the correct gene structures. Third, transcriptional analysis of a fraction of the human genome (13) and the *Arabidopsis* genome (12) has shown that about 50% of the predicted noncoding genome is expressed. Thus, we might expect this to be true for other organisms as well, although it is difficult to predict to what extent.

We examined the 41% of noncoding regions of the genome for which we detected transcriptional activity above background.

Some of the NEP probes' expression may have been due to previously unannotated genes or to additional exons of already annotated genes from the current *Drosophila* genome annotation (29, 30). To explore this possibility, we examined whether the expressed NEP probes corroborated computational exon predictions by Genscan (31). Exon predictions not in Flybase annotation (version 3.1) were divided into two groups: those bordering annotated genes and "unique" exons that did not. The latter set includes potential previously unknown genes. Genscan "unique" exons together encompassed 2045 NEP probes, 1221 (60%) of which were expressed. This represents a significant enrichment for expressed probes (compared with 41% of all NEP probes), strongly suggesting that many of these predictions represent previously unknown genes.

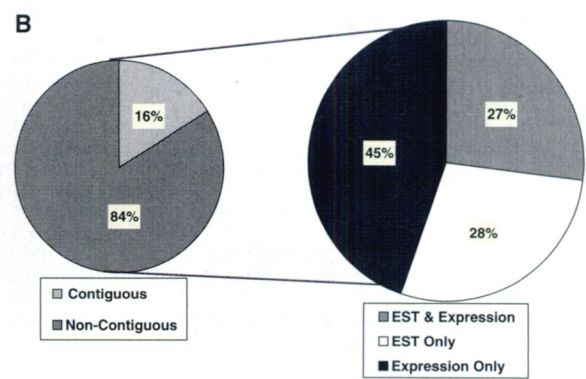
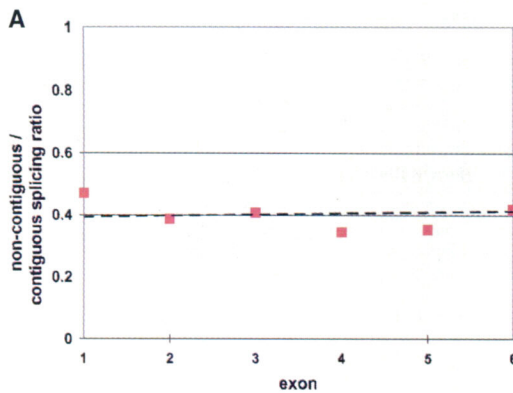
The overlapping NEP probes give supporting expression data for 1155 of Genscan's predicted exons. Of these, 369 represent additional exons bordering existing genes, whereas the remaining 786 exons belong to 716 putative novel genes (18). There were also several instances where Genscan predicted a longer upstream or downstream exon boundary relative to the current annotation. NEP probes overlapping these regions of disagreement confirmed expression of 64% (38/59) of 5' predictions and 81% (30/37) of 3' ones. Lastly, comparing to a recent study that identified several hundred expressed

exons predicted with the Fgenesh algorithm (32, 33), we found a considerable overlap of expressed NEPs within these predicted exons (61% or 477/777).

We next considered intronic expression and found that 43% of introns (6717/15,770) were expressed. There is also a relationship between the activity of an EP and its nearest intergenic NEP (and vice versa). We examined the correlations between these two classes of probes by means of a G test and a Dunn-Sidak correction for multiple testing. The nearest NEP was located on the same or opposite DNA strand in both the 5' and 3' direction, and for each of the 15 stage comparisons the differential expression levels of the probe pair were recorded. The stage or sex bias (e.g., transcript enrichment in males versus females) of NEPs at the 3' end of exons is highly correlated with the stage or sex bias of the exon; similarly, the bias of exons at the 5' end of NEPs is highly correlated with the bias of the NEP (fig. S7). These results indicate that either these expressed sequences are contained within genes whose end points have been misannotated or that they encode noncoding RNAs that are expressed in concert with nearby genes because of local chromatin or cis-regulatory effects. At least some of these transcripts are likely due to the second hypothesis, because extensive EST sequencing has failed to reveal the transcribed sequences within se-

Fig. 4. Alternative and contiguous splicing patterns of the genome. (A) Ratios of noncontiguous to contiguous splicing. For each exon, the percentage of positives from noncontiguous or contiguous SJPs was calculated. The ratio of the percentage of contiguous SJP positives to the percentage of noncontiguous SJP positives to the percentage of noncontiguous SJP positives remains near 0.4, even in larger genes

(with six exons). Dotted line shows a linear regression across the data ($y = 0.0039x + 0.3904$). (B) Exon splicing activity above background (16) (fig. S6). Many of the expressed SJPs matched sequences in *Drosophila* EST databases (27%), yet a similar fraction of SJPs found



within EST databases did not pass our intensity threshold criterion (28%). A substantial fraction of previously unknown splice junctions were detected (45%) as well as noncontiguous (i.e., alternative) exon splicing (84%).

quenced exons and because Northern blot analysis (fig. S8) shows that the noncoding sequences we have identified are often contained within several major and many minor polyA-RNA products that are inconsistent with expected protein-coding gene sizes. However, our analysis of the Genscan predictions suggests that many of the expressed sequences that correlate in expression with nearby annotated exons do in fact simply correspond to unannotated exons.

The extent to which these putative noncoding RNAs are functionally relevant awaits strategies for systematic characterization, but this genome-wide scan indicates that they are both abundant and developmentally regulated. The function of such extensively regulated noncoding gene expression during development is unknown. To determine whether these expressed sequences are functionally constrained at the sequence level, we used an alignment of the genomic sequences of *D. melanogaster* and *D. pseudoobscura* (25). Sequences corresponding to expressed NEPs were indeed more likely to be conserved than those corresponding to nonexpressed NEPs: The fraction of aligned base pairs was 68.7% for expressed probes and 59.0% for nonexpressed probes (Fig. 5) ($P = 10^{-192}$, t test), whereas 63.7% (expressed) and 60.9% (nonexpressed) of the base pairs in fully aligned probes were conserved between these species (34) ($P = 10^{-44}$, t test). This small but highly significant difference in conservation between expressed and nonexpressed NEPs indicates that there are evolutionary constraints on the expressed noncoding portion of the genome. The functional relevance of these sequences could potentially be further tested through methods such as large-scale RNAi screening (22) and systematic mutational analysis or through comparative expression analysis with additional species of *Drosophila*.

Summary. Ideally, to create a finished and fully annotated expression map of the

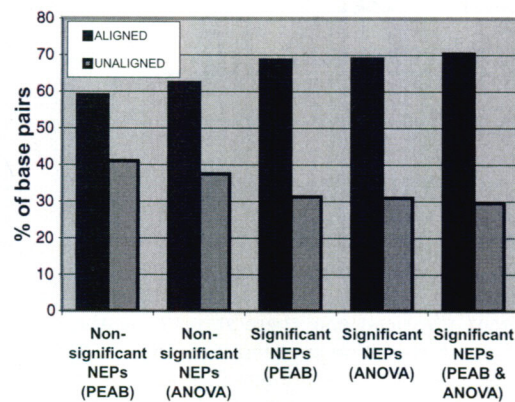


Fig. 5. NEP conservation between *D. melanogaster* and *D. pseudoobscura*. Expressed NEPs show increased sequence conservation when compared to nonexpressed NEP sequences ($P = 10^{-192}$). We used the whole-genome BLASTZ alignment between *D. melanogaster* and *D. pseudoobscura* of (25) to classify each base pair within our NEP probes as aligned (i.e., belonging to a chromosomal region with homology to *D. pseudoobscura*) or unaligned.

genome, each stage and tissue would be assayed by multiple methods. Confidence measures of expression levels can have their basis in negative controls and cross-checking between data sets, such as we have presented here. It is clear that our past understanding of genome-wide RNA transcription has been very limited, because a large proportion of exons show dynamic patterns of differential splicing and noncoding activity is ubiquitous. Taken together, our results also indicate that there are thousands of uncharacterized and unannotated transcripts expressed in a developmentally coordinated manner. Systematic genetic approaches will likely be required to determine the functions of the large class of newly identified noncoding expressed sequences, which are slightly more conserved than other noncoding sequence. Additionally, the existence of evolutionarily conserved chromosomal domains of correlated gene expression indicates that these domains are also functionally important. However, the mechanisms responsible for these expression domains remain to be elucidated. This draft expression map of the *Drosophila* genome shows that there is considerably more complexity in gene and transcript regulation than was previously known, and it represents

an initial step in identifying all the functional elements that ultimately control the developmental program of this organism.

References and Notes

1. K. P. White, S. A. Rifkin, P. Hurban, D. S. Hogness, *Science* **286**, 2179 (1999).
2. V. Reinke, K. P. White, *Annu. Rev. Genom. Hum. Genet.* **3**, 153 (2002).
3. M. N. Arbeitman et al., *Science* **297**, 2270 (2002).
4. E. De Gregorio, P. T. Spellman, G. M. Rubin, B. Lemaître, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12590 (2001).
5. S. A. Rifkin, J. Kim, K. P. White, *Nat. Genet.* **33**, 138 (2003).
6. T. R. Li, K. P. White, *Dev. Cell* **5**, 59 (2003).
7. J. De Vos et al., *Oncogene* **21**, 6848 (2002).
8. M. Parisi et al., *Science* **299**, 697 (2003); published online 2 January 2003 (10.1126/science.1079190).
9. S. D. Pletcher et al., *Curr. Biol.* **12**, 712 (2002).
10. G. M. Rubin et al., *Science* **287**, 2222 (2000).
11. J. L. Rinn et al., *Genes Dev.* **17**, 529 (2003).
12. K. Yamada et al., *Science* **302**, 842 (2003).
13. P. Kapranov et al., *Science* **296**, 916 (2002).
14. J. M. Johnson et al., *Science* **302**, 2141 (2003).
15. E. F. Nuwaysir et al., *Genome Res.* **12**, 1749 (2002).
16. S. Singh-Gasson et al., *Nature Biotechnol.* **17**, 974 (1999).
17. Materials and methods are available on Science Online.
18. All data are available online at <http://flygenome.yale.edu>. GEO submission GSE 1761.
19. R. A. Fisher, in *Statistical Methods for Research Workers*, ed. 11 (Oliver & Boyd, Edinburgh, 1950), section 21.1, pp. 99.
20. Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc.* **57**, 289 (1995).
21. M. Ashburner et al., *Nature Genet.* **25**, 25 (2000).

22. M. Boutros *et al.*, *Science* **303**, 832 (2004).
23. M. Ashburner *et al.*, *Genetics* **153**, 179 (1999).
24. M. K. Kerr, G. A. Churchill, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8961 (2001).
25. S. Richards *et al.*, in preparation.
26. W. Bender, B. Weiffenbach, F. Karch, M. Peifer, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 173 (1985).
27. W. Bender, D. P. Fitzgerald, *Development* **129**, 4923 (2002).
28. H. D. Lipshitz, D. A. Peattie, D. S. Hogness, *Genes Dev.* **1**, 307 (1987).
29. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
30. S. Misra *et al.*, *Genome Biol.* **3**, RESEARCH0083 (2002).
31. C. Burge, S. Karlin, *J. Mol. Biol.* **268**, 78 (1997).
32. A. A. Salamov, V. V. Solov'yev, *Genome Res.* **10**, 516 (2000).
33. M. Hild *et al.*, *Genome Biol.* **5**, R3 (2003).
34. G. Halasz *et al.*, unpublished data.
35. V.S. was supported by NASA Center for Nanotechnology, NASA Fundamental Biology Program, and NASA Biomolecular Systems Research Program (contract NAS2-99092). H.J.B. was partly supported by grant P20LM007276-01 from NIH. M.v.B. was supported by grant BMI 050.50.201 from the Netherlands Organization for Scientific Research (NWO). G.H. was supported by NIH training grant T32GM008224-17. C.M. was supported by an NSF Graduate Student Fellowship. This work was supported by grants to K.P.W. and H.J.B. from the Human Frontier Science Program and to K.P.W.

from the Yale Genomics and Proteomics Initiative, the NIH National Human Genome Research Institute, and the W. M. Keck Foundation. Thanks to S. Schaeffer for providing syntenic block coordinates and F. Aoki for help with the probe design. The GSE accession number is GSE 1761.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5696/655/DC1

Materials and Methods

SOM Text

fig. S1

9 June 2004; accepted 20 September 2004

REPORTS

Quantum-to-Classical Transition with Single-Photon-Added Coherent States of Light

Alessandro Zavatta, Silvia Viciani, Marco Bellini*

Single-photon-added coherent states are the result of the most elementary amplification process of classical light fields by a single quantum of excitation. Being intermediate between a single-photon Fock state (fully quantum-mechanical) and a coherent (classical) one, these states offer the opportunity to closely follow the smooth transition between the particle-like and the wavelike behavior of light. We report the experimental generation of single-photon-added coherent states and their complete characterization by quantum tomography. Besides visualizing the evolution of the quantum-to-classical transition, these states allow one to witness the gradual change from the spontaneous to the stimulated regimes of light emission.

A coherent state $|\alpha\rangle$ is the closest analog to a classical light field and exhibits a Poisson photon number distribution with an average photon number $|\alpha|^2$. Coherent states have relatively well-defined amplitude and phase, with minimal fluctuations permitted by the Heisenberg uncertainty principle. On the contrary, a Fock state $|n\rangle$ is strictly quantum-mechanical and contains a precisely defined number (n) of quanta of field excitation, hence its phase is completely undefined.

Photon-added coherent states (I) are the result of successive elementary one-photon excitations of a classical coherent field, and they occupy an intermediate position between the Fock and the coherent states. They are obtained by repeated (m times) application of the photon creation operator \hat{a}^\dagger on a coherent

state $\langle\alpha, m\rangle = k_{\alpha, m} \hat{a}^{+m}|\alpha\rangle$, $k_{\alpha, m}$ being a normalization constant and m an integer) and reduce to the limit Fock or coherent states for $\alpha \rightarrow 0$ or $m \rightarrow 0$, respectively. Quite differently from the so-called displaced Fock states, where a coherent state is used to displace a number state [for example, by mixing the two fields upon a highly reflecting beam splitter (2)], photon-added coherent states can be roughly viewed as obtained from the displacement of a coherent state operated by a Fock state. Indeed, one easily finds that all the $|n\rangle$ terms with $n < m$ are missing in the expansion of the states $|\alpha, m\rangle$ in the Fock basis, and that all the elements of the corresponding density matrix are essentially displaced toward higher indices $\rho_{ij} \rightarrow \rho_{i+m, j+m}$, leaving all the elements with $i, j < m$ void.

In the case of a single quantum of field excitation ($m = 1$), the single-photon-added coherent states (SPACs) read as

$$|\alpha, 1\rangle = \frac{\hat{a}^\dagger|\alpha\rangle}{\sqrt{1+|\alpha|^2}} \quad (1)$$

Unlike the operation of photon annihilation, which maps a coherent state into another

coherent state (that is, a classical field into another classical field), a single-photon excitation of a coherent state changes it into something quite different. In general, the application of the creation operator \hat{a}^\dagger changes a completely classical coherent state into a quantum state with a varying degree of nonclassicality that becomes more evident the smaller the initial amplitude of the $|\alpha\rangle$ state. In the extreme case of an initial vacuum state $|0\rangle$, a single excitation event transforms it into the very nonclassical single-photon Fock state $|1\rangle$, which exhibits negative values of the Wigner function (3, 4). The Wigner function is a quasi-probability distribution (5–7), which fully describes the state of a quantum system in phase space (either the position-momentum space for an harmonic oscillator or, equivalently, the space spanned by two orthogonal quadratures of the electromagnetic field for a single-mode state of light, as in this case) in the same fashion as a probability distribution (nonnegative by definition) characterizes a classical system. The negativity of the Wigner function is indeed a good indication of the highly nonclassical character of the state (Fig. 1).

We report the experimental generation of SPACs and their complete tomographic analysis, which unveils the nonclassical features associated with the excitation of a classical coherent field by a single light quantum.

Parametric down-conversion in a nonlinear crystal is the basis for the production of the desired states (Fig. 2). Here one high-energy pump photon can annihilate into two photons that obey the global energy and momentum conservation laws and thus have lower energies and are normally emitted into symmetrically oriented directions, also called the signal and idler modes. When no other field is injected in the crystal, spontaneous parametric down-conversion takes place, starting from the input vacuum field, and pairs of entangled photons with random (but mutually correlated) phases are produced. In order to generate SPACs, one has to inject a seed

Istituto Nazionale di Ottica Applicata, Largo Enrico Fermi, 6, I-50125, Florence, Italy; European Laboratory for Nonlinear Spectroscopy (LENS) and Department of Physics, University of Florence, I-50019 Sesto Fiorentino, Florence, Italy.

*To whom correspondence should be addressed. E-mail: bellini@ino.it