Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 9716–9717, August 1998

Commentary

New advances in *Drosophila* provide opportunities to study gene functions

*Norbert Perrimon**

Howard Hughes Medical Institute, Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115

Since the discovery that animal development is under genetic control, one of the major challenges for developmental biologists has been to decipher the functions of specific genes in patterning. The most successful and widely accepted approach to address these challenges in the past 15–20 years has been to characterize the defects associated with specific gene mutations after either random mutagenesis or targeted gene knock outs. Studies in *Drosophila melanogaster* and *Caenorhabditis elegans* have validated the use of systematic genetic approaches to dissect developmental pathways. In particular, the seminal screens for both maternal and zygotic mutations affecting embryonic pattern in *Drosophila* have led to a comprehensive model encompassing a finite set of mechanisms underlying the establishment of the embryonic axes and subsequent segmentation of the embryo (1, 2). In theory, however, one needs to reflect on whether the knowledge gained through one methodology can provide a thorough understanding of development. Perhaps many critical developmental steps have been missed because the genetic screens that were conducted do not take into account gene pleiotropy or redundancy. In other words, is it possible that the genetic approach has unraveled only part of the logic of development? A paper in this issue of *Proceedings* (3) illustrates that many of the genes that have not yet been analyzed in *Drosophila* encode potentially important developmental functions. This analysis underscores the need to address their biological functions because they may identify new genetic networks involved in patterning.

The genetic approach used to dissect developmental pathways relies on the identification of mutations that affect specific processes. Because this approach relies on single mutational events, it only detects genes, which, when mutated, result in readily obvious mutant phenotypes. This may not be the case, however, if the gene is pleiotropic or if multiple gene activities act in a redundant manner. In the case of a pleiotropic gene, that is, a gene expressed and playing a role at multiple times during development, it might be difficult to detect its function in a specific pathway because the activity of the same molecule in an earlier acting pathway will prevent the analysis of its role at a later time. For example, if a mutation in a specific gene is associated with embryonic lethality, it is not possible, by looking at whole mutant animals, to analyze the function of this gene in formation of adult structures. In *Drosophila*, most genes are pleiotropic, which makes the development of methodologies critical to study their multiple functions (4, 5).

Over the years, two genetic approaches have been used widely to facilitate the analysis of pleiotropic genes. The first approach relies on the creation of mosaic animals whereby the genotype varies in a cell- or tissue-specific manner. Although various techniques to generate mosaics have been developed, almost all of those in current use utilize the FLP-FRT recombination system (6) to promote chromosomal site-specific exchange. This system allows the efficient recovery of homozygous patches in an otherwise heterozygous animal and thus permits a phenotypic analysis of mutant tissues. The second approach relies on the design of sensitized genetic screens whereby mutations that either enhance or suppress a sensitized genetic background are isolated. Such sensitized screens have been particularly powerful in dissecting various signal transduction pathways (7).

The issue of redundancy, however, is more difficult to address because mutations, either in entire animals or mosaic clones, do not readily demonstrate the function of a specific gene in a developmental process. Studies on cell adhesion molecules, in particular, have demonstrated nicely the problem of genetic redundancy. Single mutations in many cell adhesion molecules expressed in the central nervous system exhibit subtle mutant phenotypes in axonal fasciculation and/or pathfinding; however, removal of two or more cell adhesion molecules can reveal more severe axonal defects (8). The *Drosophila* genome is estimated to contain 12,000 genes, and of these, mutations in over two-thirds are estimated to show no obvious loss-of-function phenotypes (5). Thus, new methodologies are needed to gain insights into the functions of these ≈8,000 genes.

A critical issue relevant to the genes that are refractory to genetic analysis is the nature of the information that they encode. How many of these genes are likely to be developmentally interesting? A study presented in this issue of *Proceedings* (3) provides a partial answer to this question and suggests, based on spatial expression profiles, that a large fraction of these genes encode potentially important developmental genes. Kopczynski *et al.* (3) have examined the expression patterns during embryonic development of a large number of cDNAs that preferentially encode secreted and transmembrane proteins. Reasoning that most membrane and secreted proteins are encoded by mRNAs that are bound to the rough endoplasmic reticulum, they prepared a cDNA library from this population of messages. To increase the chances of identifying genes that encode low abundance mRNAs, the library was normalized by using a procedure based on hybridizing a large excess of single-stranded cDNAs to a limiting amount of genomic DNA. Additionally, to reflect copy number rather than mRNA abundance, the cDNA library was prepared from cDNAs that hybridized to the genomic DNA. The authors have characterized the expression patterns during embryogenesis of these cDNAs as well as determined portions of their sequences. They found that, among the 2,518 individual cDNAs screened by *in situ* hybridization, 917 showed differential expression patterns during embryonic development. These cDNAs are expressed in a wide variety of temporal and spatial expression domains that, in some instances, are reminiscent of previously characterized genes known to play critical roles in developmental decisions. Partial sequence analysis of 1,001 of these cDNAs revealed that 811 represented novel genes. Information on the expression pat-

The companion to this commentary is published on pages 9973–9978.

*To whom reprint requests should be addressed. e-mail: perrimon@rascal.med.harvard.edu.

terns and DNA sequences of these cDNAs are publicly available from the Berkeley Drosophila Genome Project database. This extensive analysis not only establishes that a large number of new genes with developmentally interesting expression patterns remains to be characterized but also underscores the need to address their biological functions. Convincingly, Kopczynski *et al.* (3) argue that the methodology they used could be extended to collect similar data for all *Drosophila* genes.

Because the majority of the genes characterized by Kopczynski *et al.* (3) has not been identified in previous genetic screens, it suggests that they are part of the 8,000 genes of the genome that are refractory to conventional genetic analyses. What approaches should be taken to analyze their functions? Presently, over- and misexpression studies may provide the most efficient means to gain insights into the functions of these genes. Studies in tissue culture cells and with *Xenopus*, in particular, have demonstrated that forced expression of genes can be used successfully to identify their functions. In *Drosophila*, the GAL4-UAS technique (9), a system of conditional gene expression based on the transactivator properties of the GAL4 yeast protein, is used widely to drive the spatially and temporally controlled expression of specific genes. This approach, which requires the cloning of specific cDNAs under control of GAL4 target sequences (UAS) in a P element vector, cannot be used to analyze systematically all expressed genes in *Drosophila* because it would be too labor-intensive. Recently, however, Rorth (10) developed a systematic misexpression screen by combining P element insertional mutagenesis with GAL4 regulated gene expression. This system allows conditional expression of genes that are tagged randomly by insertion of a target P element that carries UAS elements and a basal promoter to direct expression of genomic sequences adjacent to the P element insertion site. When combined with a source of GAL4, the P element will direct expression of any gene that happens to lie next to its insertion site. To demonstrate the feasibility of this approach, Rorth *et al.* (11) generated 2,300 independent lines that were screened for dominant phenotypes in combination with various GAL4 lines. They found that between 2 and 7% of the lines showed dominant phenotypic abnormalities depending on which GAL4 driver was used. These gain-of-function screens appear to provide a valid approach to analyze the functions of *Drosophila* genes that have no obvious loss-of-function phenotypes. To fully evaluate the power of this system, it will be critical to determine precisely how many of the genes identified by these insertions are not associated with loss-of-function phenotypes. Finally, it should be pointed out that, because P element insertions are not random (12), only a fraction of the genome will be amenable to this approach.

To date, the information that we have obtained on genes that function during development has been mainly obtained from the analysis of phenotypes produced by specific mutations. As discussed above, this methodology has targeted only a fraction of the *Drosophila* genome and has left open the likelihood that important developmental pathways and gene functions remain uncovered. The advent of novel methodologies to manipulate the genome, in combination with information generated from systematic expression and sequencing projects, provides us with powerful tools to analyze the information content of that part of the genome that has been refractory to genetic analysis. Indeed, we should look forward to the completion of the sequence of the entire *Drosophila* genome, as well as to a complete catalog of the expression pattern of all 12,000 *Drosophila* genes. In total, this information will provide the basic reagents necessary to fully dissect the genetic networks deployed by individual cells and between cells to build an animal.

1. Nusslein-Volhard, K. C. & Weischaus, E. (1980) *Nature (London)* **379,** 597–600.
2. St. Johnston, D. & Nusslein-Volhard, C. (1992) *Cell* **68,** 201–219.
3. Kopczynski, C. C., Noordermeer, J. N., Serano, T. L., Chen, W.-Y., Pendleton, J. D., Lewis, S., Goodman, C. S. & Rubin, G. M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 9973–9978.
4. Perrimon, N., Engstrom, L. & Mahowald, A. P. (1989) *Genetics* **121,** 333–352.
5. Miklos, G. L. & Rubin, G. M. (1996) *Cell* **86,** 521–529.
6. Golic, K. G. (1991) *Science* **22,** 958–961.
7. Simon, M. A., Bowtell, D. D. L., Dodson, G. S., Laverty, T. R. & Rubin, G. M. (1991) *Cell* **67,** 701–716.
8. Goodman, C. S. (1996) *Annu. Rev. Neurosci.* **19,** 341–377.
9. Brand, A. & Perrimon, N. (1993) *Development* **118,** 401–415.
10. Rorth, P. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12418–12422.
11. Rorth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., Rubin, G. M. Weigmann, K., Milan, M., Benes, V., Ansorge, W. & Cohen, S. M. (1998) *Development* **125,** 1049–1057.
12. Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T. & Rubin, G. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10824–10830.