Hidden Messages in Genetic Maps

Christine J. Farr and Peter N. Goodfellow

In viruses and prokaryotes, the positioning of genes is important and is often used for regulating gene expression. In higher organisms, the meaning of gene order is less obvious. Nevertheless, mapping information generated by the Genome Project is beginning to provide new clues.

Genomes become increasingly complex by duplication of DNA. This evolution can occur by duplication of the whole genome, duplication of individual chromosomes, gene clusters, genes, or parts of genes. In vertebrate evolution, the genome has undergone at least two complete duplications and many duplications of its parts. Once new genetic material is created, gene order can be disrupted by translocations and other chromosomal rearrangements. The order of particular sets of genes can be identified as functionally significant by comparing gene order between species (*1*). This type of comparative gene mapping will also help to elucidate the evolutionary relations among different species.

The best example of a conserved association of genes in the mammalian genome is the X chromosome, although gene order is not conserved in this case. The chromosomal basis of sex determination in mammals results in females having two X chromosomes and males having only one. The gene dosage inequality between sexes is obviated by the random inactivation of one of the X chromosomes in female somatic cells. The requirement to function optimally with only a single copy per cell and to be subject to X-inactivation selects against both acquiring new X-linked genes and against losing genes from the X chromosome. X-inactivation may be a special case of the wider phenomenon of genomic imprinting; if this is so, other blocks of imprinted genes may exhibit conserved gene associations. The Y chromosome is also subject to special constraints because of its role as a sex chromosome; all mammalian Y chromosomes must carry the testes-determining gene *SRY* (*2*).

The conservation of gene associations on the X chromosome contrasts with the pseudoautosomal region, which is shared by the X and Y chromosomes. Genes in this region escape X-inactivation and function at two doses; the only functional constraint is that the X and Y sequences are the same. The human and mouse pseudoautosomal regions may be unrelated: although *CSF2RA* is pseudoautosomal in humans, homologous sequences map to the telomeric end of mouse chromosome 19 (*3*).

Bacteria exhibit many examples of ordered gene clusters and the genes encoding enzymes of a metabolic pathway are often tightly linked as operons, with expression under coordinate control. In the genomes of higher organisms, genes are usually separated by thousands of base pairs of DNA. Occasionally, however, sets of functionally or structurally related genes are found clustered and gene order is conserved. The genes within these clusters may be located within one broad unit of transcriptional control. For example, it has been suggested that the HOX genes descended from a single ancestral homeobox gene through successive duplication events, which would account for the clusters. The HOX genes exert their function by differential expression along the anterior-posterior axis of the developing embryo. There is a strict correspondence between the chromosomal ordering of the genes within a cluster and their expression domains, suggesting that these two phenomena are mechanistically linked (*4*).

The implication that gene order is an important parameter for correct developmental expression has also emerged from studies of the human β-globin gene cluster. The different developmentally expressed β-type globin genes of mammals can be traced back to a single progenitor gene, which duplicated 150 to 200 million years ago in the early mammals. Further duplications led to a genomic domain of five developmentally regulated loci (5′-ε-γ^G-γ^A-δ-β-3′). In the primate β-globin gene cluster, sequences from all five loci in the ancestral 5′ to 3′ arrangement have been detected. The entire set of β-like globin genes is controlled by the locus control region (LCR) situated 5′ of the ε-globin gene and more than 50 kilobases away from the β-globin gene. Transgenic studies indicate that expression of the various genes within the β-globin locus is influenced by their position relative to the LCR (*5*).

Gene clusters can contain genes that are not similar by sequence analysis but that are related by function. The best studied example is the major histocompatibility complex duplication, for example, the class 1 and class 2 genes, and the nearby unrelated genes involved in antigen presentation and other aspects of immune function (*6*). What advantages may be gained by keeping these genes together? Linkage of the closely related class 1 genes may facilitate the generation of diversity by gene conversion between the different members. The clustering of the disparate genes involved in a single function may help coordinate induction and control. What selective forces shaped the major histocompatibility locus? Interaction between the polymorphic proteins produced by different genes may have resulted in the selection of favorable combinations. Loss of the optimal combinations of genes by recombination could be decreased by reducing recombination, leading to condensation of the genome (*7*).

Gene order can also be fixed as the result of structural constraints. In several cases, overlapping genes are transcribed on opposite DNA strands. In the course of efforts to identify the neurofibromatosis gene, *NF1*, three active genes (*OMGP*, *EVI2A*, and *EVI2B*) were located on the antisense strand of an intron of *NF1* (*8*). Homologs of *erbA* have been found that use coding information from both strands of the same DNA sequence in both rats and humans (*9*). The mouse *surfeit* locus contains at least six housekeeping genes, which are unrelated by sequence homology and apparently unrelated by function. In the tightest clustering of genes so far described in the mammalian genome, the *Surf-1* and *Surf-2* genes share a bi-directional promoter, while *Surf-2* and *Surf-4* overlap at their 3′ ends. The tight cluster of *Surf-1* to *Surf-5* genes has been conserved over 600 million years of divergent evolution (*10*). Once such tight gene associations have been established they are effectively locked into the genome, unable to separate because of the small distances between the genes.

Although it is clear that selection and functional constraints can modify gene order, it is also obvious that not all gene orders are selected. Sydney Brenner has likened genome mappers to astronomers boldly mapping the heavens (*11*). Seeking meaning in gene order may be the equivalent of astrophysics—or it might be astrology.

The authors are in the Genetics Department, University of Cambridge, Cambridge, CB2 3EH United Kingdom.

REFERENCES AND NOTES

1. Human Gene Mapping 11, *Cytogenet. Cell Genet.* **58**, 1 (1991).
2. J. W. Foster *et al.*, *Nature*, in press.
3. C. M. Disteche *et al.*, *Nat. Genet.* 1, 333 (1992).
4. S. J. Gaunt, *Bioessays* **13**, 505 (1991).
5. O. Hanscombe *et al.*, *Genes Dev.* **5**, 1387 (1991).
6. J. Trowsdale and S. H. Pois, *Curr. Opin. Genet. Dev.* **2**, 492 (1992).
7. R. A. Fisher, *The Genetical Theory of Natural Selection* (Dover, New York, ed. 2, 1958).
8. D. Viskochil *et al.*, *Mol. Cell. Biol.* **11**, 906 (1991).
9. N. Miyajima *et al.*, *Cell* **57**, 31 (1989).
10. P. Colombo, J. Yon, K. Garson, M. Fried, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6358 (1992).
11. S. Brenner, *Nature* **356**, 27 (1992).