Review Article

The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms

John S. Mattick* and Michael J. Gagen†[1]

*Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia; and †Department of Mechanical Systems Engineering, Kanazawa University, Kanazawa, Ishikawa, Japan

Eukaryotic phenotypic diversity arises from multitasking of a core proteome of limited size. Multitasking is routine in computers, as well as in other sophisticated information systems, and requires multiple inputs and outputs to control and integrate network activity. Higher eukaryotes have a mosaic gene structure with a dual output, mRNA (protein-coding) sequences and introns, which are released from the pre-mRNA by posttranscriptional processing. Introns have been enormously successful as a class of sequences and comprise up to 95% of the primary transcripts of protein-coding genes in mammals. In addition, many other transcripts (perhaps more than half) do not encode proteins at all, but appear both to be developmentally regulated and to have genetic function. We suggest that these RNAs (eRNAs) have evolved to function as endogenous network control molecules which enable direct gene-gene communication and multitasking of eukaryotic genomes. Analysis of a range of complex genetic phenomena in which RNA is involved or implicated, including co-suppression, transgene silencing, RNA interference, imprinting, methylation, and transvection, suggests that a higher-order regulatory system based on RNA signals operates in the higher eukaryotes and involves chromatin remodeling as well as other RNA-DNA, RNA-RNA, and RNA-protein interactions. The evolution of densely connected gene networks would be expected to result in a relatively stable core proteome due to the multiple reuse of components, implying that cellular differentiation and phenotypic variation in the higher eukaryotes results primarily from variation in the control architecture. Thus, network integration and multitasking using *trans*-acting RNA molecules produced in parallel with protein-coding sequences may underpin both the evolution of developmentally sophisticated multicellular organisms and the rapid expansion of phenotypic complexity into uncontested environments such as those initiated in the Cambrian radiation and those seen after major extinction events.

Introduction

Our understanding of the relationship between genetic information and biological function is rooted in the one gene–one protein hypothesis and in classical studies of the *lac* operon and the "genetic code," i.e., the triplet code specifying amino acids in protein-coding sequences. The concept of DNA as a relatively stable, heritable source of template information for proteins, transduced through a temporary and discrete RNA readout, has become an article of faith and implicitly, but very powerfully, influenced our ideas on the structure of genetic systems. Accordingly, cells and organisms are thought of as being built from a myriad of structural and catalytic proteins whose expression is generally controlled by other regulatory proteins which bind to DNA. This is a biochemical rather than an informatic perspective, which, apart from local analysis of promoter function, gives little thought to the problem of how complex programs of gene activity in the higher organisms might be integrated and regulated in four dimensions.

Genome sequencing projects have shown that the core proteome sizes of *Caenorhabditis elegans* and *Drosophila melanogaster* are similar and that each is only about twice the size of yeast and some bacteria, despite these animals' every appearance of possessing more than twice the complexity of micro-organisms (Chervitz et al. 1998; Rubin et al. 2000), leading to the conclusion that "the evolution of additional complex attributes is essentially an organizational one; a matter of novel interactions that derive from the temporal and spatial segregation of fairly similar components" (Rubin et al. 2000). This conclusion is reinforced by the finding that the human genome has only about 30,000 protein-coding genes (Roest Crollius et al. 2000; International Human Genome Sequencing Consortium 2001; Venter et al. 2001), 99% of which are shared in common with the mouse (J. C. Venter, personal communication). The increased complexity of the higher eukaryotes is related, at least in part, to the production of different protein isoforms from the same gene by alternative splicing (Croft et al. 2000). However, the other striking feature of the evolution of these organisms, largely ignored to date, is the huge increase in the amount of complex non-protein-coding RNAs, which can represent up to 97%–98% of all transcriptional output from the genome. That is, the vast majority of the expressed information in the higher eukaryotes is in RNA, not protein-coding sequences. Moreover, less than 1% of the sequence differences between individual humans occurs in protein-

coding sequences (Venter et al. 2001), which suggests that the majority of phenotypic variation between individuals (and species) results from differences in the control architecture, not the proteins themselves. This is in contrast to bacteria, wherein phenotypic variation is primarily achieved by varying the proteome—different strains of *Escherichia coli* have been found to differ by over 20% in their gene complement (Hayashi et al. 2001).

The view that phenotypic variation in complex organisms results from the differential use of a set of core components is becoming common (Gerhart and Kirschner 1997; Duboule and Wilkins 1998) and includes such concepts as ''synexpression groups'' (Niehrs and Pollet 1999), ''syntagms'' of interacting genes (Huang 1998) and gene cassettes (Jan and Jan 1993), the reuse of modules in signaling pathways (Pawson 1995; T. Hunter 2000), and enhanced rates of evolution by varying connections between modular network components (Hartwell et al. 1999; Holland 1999). These concepts have been drawn primarily from electrical circuit design and have focused principally on the modules rather than on the interconnecting control architecture of the system.

Particular network models, which range in size from single regulated circuits (Mestl, Plahte, and Omholt 1995; Almeida, Fernandes de Lima, and Infantosi 1998; Mendoza and Alvarez-Buylla 1998; Yuh, Bolouri, and Davidson 1998) to complete genomes (Thieffry et al. 1998), have demonstrated that feedback-subnetworks can exhibit computational behaviors including ''learned behavior'' (Bhalla and Iyengar 1999), that switching networks and transcriptional control networks can exhibit dynamical stability (Wolf and Eeckman 1998; Smolen, Baxter, and Byrne 2000), and that feedback circuits can implement oscillators governing cell cycles and circadian clocks (Dano, Sorensen, and Hynne 1999; Haase and Reed 1999; Shearman et al. 2000). Stochastic noise and time delays allowing feedback, molecular memory, and oscillations can be incorporated into such circuit models (Smolen, Baxter, and Byrne 1999), generating probabilistic phenotypic variation (McAdams and Arkin 1997) and amplification of signals (Hasty et al. 2000). Some of these models have been verified by synthesizing circuits in cells to feature bistability, oscillations, and stochastic destruction of temporal correlations (Becskei and Serrano 2000; Elowitz and Leibler 2000; Gardner, Cantor, and Collins 2000).

However, such models are unsuited to the analysis of global cellular connectivity and dynamics, as they cannot be scaled up to large network sizes, since linear increases in the number of interconnected circuit nodes requires quadratic increases in the number of interconnecting molecules. This leads to an explosive increase in model size which severely constrains numerical simulations using current computing technologies (see, e.g., Weng, Bhalla, and Iyengar 1999). A number of alternate approaches have sought to avoid this size explosion by treating subnetworks as active integrated logic components which are interconnected into larger networks (McAdams and Shapiro 1995), or by exploiting hierar-

chically organized control systems to significantly decrease analytical complexity (van der Gugten and Westerhoff 1997).

We suggest that biology has solved this problem differently. Here we examine first whether the types of network control architecture which are used to integrate and multitask computers (and which implicitly feature in other complex information processing systems) might also be employed by molecular biological networks to generate phenotypic complexity and variability. Second, we examine the proposition and collate the evidence that introns and other nonprotein-coding RNAs may have evolved to function as network control molecules in the higher organisms, freeing such organisms from the constraints of a simple single-output protein-based genetic operating system.

Multitasking by Programmed Network Control

Multitasking is employed in every computer in which control codes (program instructions) of n bits set the central processing circuit to process one of 2^n different operations. Sequences of control codes (a program) can be internally stored in memory, creating a self-contained programmed response network—a computer—as originally defined by von Neumann in 1945 (von Neumann 1982). Prior to the arrival of the von Neumann computing architecture, a computer could only be reprogrammed by laborious rewiring of the central processing unit, while subsequent reprogramming simply required loading new control codes into memory. In all computing networks, processing requires not only stored program instructions, but also communication between nodes to synchronize and integrate network activity. In theory, gene networks could exploit similar technology using internal controls to multitask components and subnetworks to generate a wide range of programmed responses, such as in differentiation and development.

Existing genetic circuit models, although sophisticated, ignore endogenous controlled multitasking and consider each molecular subnetwork (involving a few genes, for instance) to be sparsely interconnected and either off or on to express only one dynamical output (see, e.g., McAdams and Shapiro 1995; Bhalla and Iyengar 1999; Weng, Bhalla, and Iyengar 1999). Such models require more complex genetic programs to be built from many subnetworks encoded by exponentially large numbers of genes, a severe constraint. In contrast, multitasking via n controls (single molecules suffice) can, in theory, achieve exponential (2^n) multitasking of subnetwork dynamical outputs and allow a wide range of programmed responses to be obtained from limited numbers of subnetworks (and genetic coding information). The imbalance between the exponential benefit of controlled multitasking and the small linear cost of control molecules makes it likely that evolution will have explored this option. Indeed, this may be the only feasible way to lift the constraints on the complexity and sophistication of genetic programming.

The relevant output dynamics of complex systems can only be found by a comprehensive search of input parameter space, as nonlinear interactions within the network can have unexpected and emergent properties. During evolution, genetic networks must perform a similar search of possible subnetwork dynamics, which can also be greatly accelerated when multitasking is employed. It is far easier to modify and expand the numbers of small control sequences than to duplicate and mutate entire subnetworks of genes, Additionally, simply turning off controls may reset the program, perhaps important in reproduction and survival. Most importantly, a control architecture makes it possible to coordinate activity across interacting sets of genes, while variation of this architecture can generate a large spectrum of different protein expression profiles.

However, multitasking controls are only useful to the extent that they convey information about the dynamical state of the network and its surrounding environment. To do this, nodes within the network must not only receive multiple inputs, but also generate multiple outputs (endogenous controls). In cells, molecular switches which act as input controls to relay metabolic, physiological, and environmental information by modifying protein structure and protein-protein and protein-nucleic acid binding affinities have been known for many years. However, endogenous controls need to be correlated with the internal cellular state, the central component of which is gene expression status. Importantly, in a fully integrated network, endogenously sourced controls are likely to be more numerous than externally sourced controls, just as computers must internally regulate millions of internal subnetwork controls to communicate with a few peripherals in the environment.

Ideally then, in order for a molecular genetic network to be capable of complex programming and multitasking, each of the gene subnetworks within a cell must produce numerous control molecules in parallel with their primary gene products, which dynamically communicate with other subnetworks (via transcriptional, splicing, and translational controls, among others). Such a system would be expected to display an exponential increase in its ability to manage and integrate larger genetic data sets and in its functionality and phenotypic range. In addition, because modulation of system dynamics can be readily achieved by mutation of control molecules, such a system should be able to explore new expression space at fast evolutionary rates over short evolutionary timescales.

A controlled multitasked molecular network is schematically shown in figure 1 in contrast to an uncontrolled regulated network. This network architecture can be equally applied to computer networks, neural networks, and cellular networks.

The Evolution of Controlled Multitasked Gene Networks

The nodes of a controlled multitasked network must be capable of generating and integrating multiple
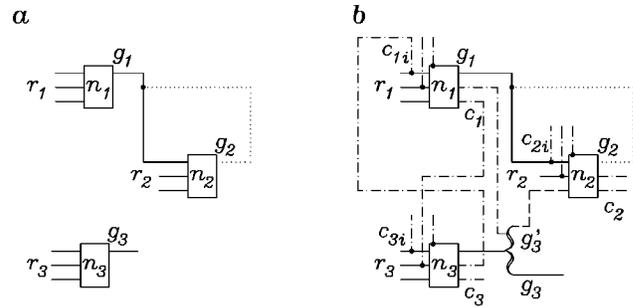

Fig. 1.—Schematic representation of subnetworks of an uncontrolled regulated network and a controlled multitasked network. a, An uncontrolled subnetwork wherein nodes take limited numbers of regulatory inputs r_k and generate limited numbers of protein outputs g_k. Here, g_1 regulates n_2 while being subject to feedback interactions from g_2 (dotted line). b, The same subnetwork with each node expressing a multiplex output of protein product g_k and many control molecules c_k, each capable of targeted interactions to multitask the subnetwork. A sample of possible interactions (shown as dot-dash lines) includes control c_1 determining the alternative splicing of the node n_3 output giving g_3 or g_3', the latter of which regulates node n_2 when expressed, while nodes n_1 and n_3 each feedback controls onto the other. It is evident that controls increase interconnectivity, which increases network dynamical output complexity.

inputs and outputs. Such networks are generally stable and scale-free, with some nodes having high connectivity and others having low connectivity, similar to most communication and social networks, including the Internet (Albert, Jeong, and Barabasi 2000). Multiply connected networks are widely employed in other complex information processing systems, including neurobiology, where secondary networking signals, termed "efference" signals, underlie sensory awareness and motor coordination (Bridgeman 1995; Andersen et al. 1997). The concept of multiple inputs and outputs is also a well-established feature of neural networks in cognition, language, and memory (Plunkett et al. 1997; Elman 1998). These networks involve densely connected webs of processing units that propagate and transform complex patterns of activity and are capable of self-organization. They operate by a form of parallel distributed processing, whereby information is distributed across the system such that patterns of activation across sets of "hidden units" (i.e., controls), which define the state of the network, then determine the pattern of activation across output nodes (McClelland and Rumelhart 1985; Rumelhart and McClelland 1986; McClelland and Plaut 1993; Plunkett et al. 1997; Elman 1998).

In cells, genetic information is transduced into RNAs and proteins, the latter of which are considered to be the major functional outputs of the genome and to comprise the structural, metabolic, and regulatory systems by which cells and organisms function. Theoretically, it is possible for proteins to provide multiple input controls, and combinatorial regulation does occur in the case of, e.g., transcription factors, but for each genetic node to be multiply connected, a multiplex output is also required from each node, at least on average. At present, however, there is no evidence that proteins are used to provide an output connection function (i.e., in parallel with a primary gene product), and no output (network-

ing) molecules acting as controls influencing the activity of other genes (or RNAs or proteins) have been identified, although intronic RNA could fulfill this function.

Prokaryote genomes consist almost entirely of protein-coding sequences that are separated by short intergenic regions containing promoters and transcription termination signals, and are flanked by 5′ and 3′ untranslated signals that are involved in translational control, mRNA localization, and mRNA stabilization. Prokaryotic genes are frequently arranged in operons allowing cotranscription of genes with related functions, such as the *lac* operon, although rarely if ever are broader regulatory (output control) proteins expressed from the same node (operon). Most regulatory proteins are expressed from separate nodes. For the *lac* operon, input control comes from the *lac* repressor (polling cellular lactose status) and the CAP protein (polling cellular cAMP/energy status), both of which are expressed separately (Reznikoff 1992). Transcription of the *lac* operon (and most operons) is therefore blind—no secondary communication signals are coexpressed and other cellular nodes remain unaware of the event, except indirectly through delayed feedback loops which relay metabolic state information. The number of regulatory proteins in bacteria is a relatively low proportion of the total, and the system appears to function as a set of sparsely connected local area networks, with each regulator contacting a limited number of nodes in the genome, and with controls usually composed of metabolic or environmental chemicals that intersect with these regulators.

Prokaryotes have limited genome sizes (upper limit ~10 Mb) and low phenotypic complexity, suggesting that advanced integrated control technologies are not widely employed in these organisms. The absence of a prokaryotic multiplex control system also implies that a system built primarily on proteins has inherent limitations. It is not as if prokaryotes have had insufficient time to evolve such a system—they have had four billion years and countless generations in which to explore all possible protein and phenotypic space, aided by lateral transfer to spread innovation. However, while multiplex input at complex promoters is possible (see below), a multiplex output (synchronous control signals based on proteins) is far more difficult. Prokaryotic gene transcripts are not processed to produce subspecies, and the only parallel outputs that are possible are separate proteins translated from polycistronic mRNAs. To average just one (additional) protein output per node requires doubling genome size, and the multiplex output necessary for true dynamical systems integration requires huge increases in both genome size and energy cost to the cell, making such integration unmanageable and ultimately impossible by this means. The lack of a sophisticated systems control technology in prokaryotes may be the primary reason why genomic and developmental complexity has not arisen in these lineages. This also reciprocally suggests that this constraint had to be solved before more complex organisms could evolve and that the network control mechanisms operating in

the higher eukaryotes may not be principally protein-based.

The complexity and phenotypic versatility of the higher eukaryotes is thought to result primarily from a larger set of proteins and combinatorial (input) control of gene expression by such proteins. This includes multiple "transcription factors" and intersecting signal transduction pathways influencing gene expression, along with alternative splicing producing different proteins from the same gene (Lopez 1998; Croft et al. 2000; Smith and Valcarcel 2000), generating subtly or substantially different functions in different tissues. While gene number is higher in complex eukaryotes, and alternative splicing greatly increases protein isoform numbers, combinatorial control of gene expression only allows multiplex input control, and alternative splicing mainly provides flexibility in endpoint specialization. Neither of these systems allows multiplex output of control molecules at the point of gene expression, a principal requirement for a multitasked network.

One possible population of cellular molecules with the attributes required to act as controls in genetic multitasking are functional introns and other noncoding RNAs. These have previously been suggested to potentiate a parallel processing system with vastly expanded regulatory options, leading to more complex genetic data sets, programs, and phenotypes which was perhaps critical to the evolution of multicellular organisms (Mattick 1994). These RNAs were initially christened iRNA (intronic/informational RNA) (Mattick 1994), but because of the ambiguity in that term (mRNA is also informational) and potential confusion with the recently discovered phenomenon termed RNAi (RNA interference), we have chosen to denote non-protein-coding RNAs which are involved in network integration and control as eRNA ("efference" RNA).

A Role for Introns and Other Noncoding RNAs in Dynamical Gene-Gene Communication, Genetic Multitasking, and Systems Integration

Potential cellular control molecules enabling multitasking and system integration must be capable of specifically targeted interactions with other molecules, must be plentiful (as limited numbers impair connectivity and adaptation in real and evolutionary time), and must carry information about the dynamical state of cellular gene expression. These goals are most simply achieved by spatially and temporally synchronizing control molecule production with gene expression. Most protein-coding genes of higher eukaryotes are mosaics containing one or more intervening sequences (introns) of generally high sequence complexity, which are spliced out during pre-mRNA processing to generate a nuclear population of intronic RNA with concentration profiles linked to that of the exons, which are reassembled during this process to form mRNA, and subsequently translated into protein. The numbers of protein-coding genes do not increase exponentially in complex organisms and hence cannot provide large-scale cellular connectivity (which does increase exponentially). The genomes of higher or-

ganisms are nevertheless much larger than those of single-celled organisms, with the vast majority of this size increase (after accounting for variable amounts of repetitive DNA) occurring within intron sequences and other non-protein-coding RNAs. Introns therefore fulfill the essential conditions for system connectivity and multitasking—(1) multiple output in parallel with gene expression; (2) large numbers, especially if, as is likely (see below), they are further processed to smaller molecules after excision from the primary transcript; and (3) the potential for specifically targeted interactions as a function of their sequence complexity. Sequences of just 20–30 nt should generally have sufficient specificity for homology-dependent or structure-specific interactions. Introns are therefore excellent candidates for, and perhaps the only source of, possible control molecules for multitasking eukaryotic molecular networks, which relieve the problems associated with protein-based systems, as genetic output can be multiplexed and target specificity can be efficiently encoded, assuming a receptive infrastructure.

Before considering the evidence that introns might fulfill such a function, it is necessary to address some preconceptions. The widely held idea that introns are nonfunctional is an assumption which dates back to the initial discovery of these sequences—a great surprise at the time (Williamson 1977)—which were interpreted in the light of the prevailing dogma that all cellular functions were directed by proteins and that genes were simply repositories of protein-coding sequences, which in turn was based on bacterial molecular genetics, in the absence of any understanding of the evolutionary history and origin of nuclear introns. There is no evidence to support the assumption that nuclear introns generally (as a class of sequences in the higher organisms) are nonfunctional, although the issue is confused by the fact that most introns are less conserved in sequence than accompanying protein-coding exons and that some or many will not have evolved function, as each intron will be evolving largely independently (see below).

Introns Populated the Eukaryotic Lineage Late in Evolution

It is now clear that modern nuclear introns are not ancient remnants of the prebiotic assembly of genes, but the evolutionary descendants of self-catalytic group II introns, which have similar splicing mechanisms (Lambowitz and Belfort 1993; Eickbush 2000). These elements appear to have penetrated the eukaryotic lineage late in evolution (Cavalier-Smith 1991; Palmer and Logsdon 1991; Mattick 1994; Stoltzfus et al. 1994; Cho and Doolittle 1997; Logsdon 1998; Wolf et al. 2000) and to have expanded initially by retrotransposition (Cousineau et al. 2000; Eickbush 2000) and later (after their sequence constraints were reduced by the evolution of the spliceosome) by other mutational, recombinational, and insertional processes (Tarrio, Rodriguez-Trelles, and Ayala 1998). Self-catalytic group II introns do occur in bacteria, usually in tRNA genes (Ferat and Michel 1993; Martinez-Abarca and Toro 2000), and the likely

reason that introns are generally absent from prokaryotic protein-coding sequences is the intimate coupling of transcription and translation in these cells, which does not allow time for intron excision (Mattick 1994).

The evolution of the nucleus and the separation of transcription and translation in the eukaryotes provided the opportunity for these introns to invade protein-coding genes, as long as their removal by self-splicing was efficient enough not to interfere with mRNA and protein production. The subsequent evolution of the spliceosome (involving the devolution of internal *cis*-acting catalytic RNAs into *trans*-acting spliceosomal RNAs and recruitment of accessory proteins) (Lambowitz and Belfort 1993; Mattick 1994; Newman 1994; Stoltzfus 1999; Yean et al. 2000) made intron processing easier, which reduced the negative selection against introns and allowed them more latitude. It also relaxed their internal sequence requirements, leaving them free to evolve and to explore new evolutionary space, based on RNA molecules produced in parallel with protein-coding sequences (Mattick 1994). This would have been accelerated by the co-evolution of receptor systems for these molecules, involving RNA-protein, RNA-RNA, and RNA-DNA/chromatin interactions, in the same way as other complex systems such as the ribosome and the spliceosome have evolved (Stoltzfus 1999). It does not follow that all introns in a given lineage will have evolved function (see below), but, rather, there will have been increasing opportunity to do so. This also applies to other types of insertion elements (International Human Genome Sequencing Consortium 2001). Any useful functions that may have been acquired would have provided a positive selection pressure, which is the basis of Darwinian evolution. The general hypothesis that intron-derived RNAs may have evolved *trans*-acting functions is therefore eminently feasible and should be entertained.

Intron Density Correlates with Developmental Complexity

Intron size and sequence complexity correlates well with developmental complexity, and introns comprise the majority of pre-mRNA sequences in the higher organisms. In developmentally simple eukaryotes like *Schizosaccharomyces pombe,* Aspergillus, and Dictyostelium, introns compose only 10%–20% of the primary transcript and are generally small, with an average length of less than 100 bases and a density of about one to three introns per kilobase of protein-coding sequence. These data are consistent with hybridization kinetic analyses of the relative sequence complexity of ''heterogeneous nuclear RNA'' (hnRNA) versus mRNA in lower eukaryotes (Davidson 1976). In the higher plants, there are two to four introns per gene of an average length of about 250 bases, comprising about 50% of the primary transcript. In animals, the average intron size increases to about 500 bases in Drosophila and *C. elegans* and to about 3,400 bases in humans (six to seven introns per gene, average over 95% of the primary transcript) (Palmer and Logsdon 1991; Deutsch and Long

1999; International Human Genome Sequencing Consortium 2001; Venter et al. 2001).

Organisms with streamlined genomes provide a good test of the stringency of intron expansion. The pufferfish *Fugu rubripides* has, for unknown reasons, almost no repetitive (presumably superfluous) sequences in its genome: three quarters of pufferfish introns are very small, whereas the remainder are much larger and still account for the majority of total unique sequence (Brenner et al. 1993; Elgar 1996). A similar skewed distribution is observed in the compact genome of *Arabidopsis thaliana* (Carels and Bernardi 2000). (A comprehensive analysis of eukaryotic intron size can be found at http://isis.bit.uq.edu.au; Croft et al. 2000). Most of the small introns are probably vestigial, whereas, in these and probably in most organisms, larger introns with high sequence complexity may be considered to indicate functionality. This is the case in at least one instance (Cecconi et al. 1996). Interestingly, the complex alga *Volvox carteri* appears to possess large introns (Fabry et al. 1993). Since the order Volvocales contains a number of closely related members ranging from unicellular (Chlamydomonas) through a series of colonial forms to fully differentiated forms, this may represent a useful test case for the appearance of larger introns through an evolutionary developmental series.

Introns Have the Signatures of Information

Introns (and other nonprotein-coding RNAs; see below) of higher organisms exhibit all the signatures of information. They generally have high sequence complexity (Tautz, Trick, and Dover 1986), although one must distinguish between introns that may have evolved function and those that have not (which will be more degenerate) and take account of the differing proportions of functional and nonfunctional introns in lineages of different developmental complexity. While introns generally show less conservation than adjacent protein-coding sequences, which are subject to strong constraints, so also do adjacent promoters and 5' and 3' untranslated regions of mRNA, all of which are known to be important in gene regulation. The plasticity and more rapid evolution of these regulatory sequences does not mean they are nonfunctional, and we suggest the same holds in general for introns.

Nonetheless, some introns are highly conserved over substantial evolutionary distances (Garbe and Pardue 1986; Rieger and Franke 1988; Tournier-Lasserve et al. 1989; Lloyd and Gunning 1993; Starke and Gogarten 1993; Koop and Hood 1994; Bagavathi and Malathi 1996; John, Smith, and Kaiser 1996; Rosby, Alestrom, and Berg 1997; Kazmierczak et al. 1998; Aruscavage and Bass 2000; Sun et al. 2000; Yatsuki et al. 2000), often in large blocks (Jareborg, Birney, and Durbin 1999), indicating that they are under functional constraint. While such conservation might, in some cases, be ascribed to the presence of important *cis*-acting elements such as transcription enhancers, this cannot account for the extensive homology between, for example, the 94 kb of introns in the mouse and human T-cell receptor genes showing a high level of nucleotide sequence conservation (over 70%) similar to that of the accompanying exons (Koop and Hood 1994). Intron sequences can also evolve faster than silent positions in accompanying exons (Kloek et al. 1996) (sites that are presumably relatively neutral), indicating positive selection, further evidence of intron functionality. Moreover, if introns are acting as networking controls, the important issue is not the conservation of the sequence per se (i.e., to produce functional domains in the protein sense), but the conservation of interactions.

Noncoding RNAs Comprise the Majority of Genomic Output

Many (if not most; see below) transcripts from the genomes of higher organisms do not encode proteins at all (Eddy 1999; Erdmann et al. 1999). Where they have been examined, these nonprotein-coding transcripts are conserved and clearly functional. Well-documented examples include XIST (involved in female X-chromosome inactivation) (Brockdorff 1998; Lee, Davidow, and Warshawsky 1999; Hong, Ontiveros, and Strauss 2000) and H19 (mutants of which promote tumor development) (Wrana 1994; Hurst and Smith 1999), both of which are imprinted and differentially spliced without encoding any protein (Hurst and Smith 1999; Hong, Ontiveros, and Strauss 2000; F. Clark, personal communication). Others include *roX1* and *roX2* RNAs involved in dosage response (male X-chromosome activation) in Drosophila, heat shock response RNA in Drosophila, oxidative stress response RNAs in mammals, His-1 RNA involved in viral response/carcinogenesis in humans and mice, SCA8 RNA involved in spinocerebellar ataxia type 8 which is antisense to an actin-binding protein, and ENOD40 RNA in legumes and other plants (Eddy 1999; Erdmann et al. 1999; Nemes, Benzow, and Koob 2000). The 200-kb bithorax-abdominalA/B locus of Drosophila produces seven major transcripts (there may be minor ones as well), only three of which encode proteins, but all of which have phenotypic signatures and are developmentally regulated (Akam et al. 1985; Hogness et al. 1985; Lipshitz, Peattie, and Hogness 1987; Sanchez-Herrero and Akam 1989). These are not isolated examples. Many loci, including imprinted loci, express noncoding antisense and intergenic transcripts, some of which are alternatively spliced and developmentally regulated (Ashe et al. 1997; Lipman 1997; Potter and Branford 1998; Lee, Davidow, and Warshawsky 1999; Filipowicz 2000; Hastings et al. 2000; Nemes, Benzow, and Koob 2000), in addition to being stably detectable in the nucleus (Ashe et al. 1997).

There is a general point to be made here. Gene regulation often involves "enhancers" located either downstream of the transcription start site (in introns) or in the upstream promoter region spanning many kilobases of DNA, as well as more distant regions sometimes referred to as "locus control regions." In some, and perhaps many, cases, these intergenic regions are themselves transcribed (into noncoding RNAs), suggesting that their effects might be related to *trans*-acting,

not *cis*-acting sequences, which can confound interpretation of mutational analysis of ''promoter regions.'' Such transcripts have been discovered by careful analysis of transcriptional activity around a locus of interest, such as β-globin (Ashe et al. 1997), but this has not often been done.

Also, as noted by Eddy (1999), most systematic genomic screens are biased against discovering noncoding RNAs. PolyA+RNA preparations used in cDNA library construction are depleted of noncoding RNAs, and bioinformatic searches are limited by a lack of knowledge about the signatures and variety of these molecules, although comparative genomics to identify regions of sequence homology outside of protein-coding regions may provide clues. Many such homology regions are evident from comparison of the human and mouse genomes (V. R. Bonazzi, personal communication), and many noncoding regions in *C. elegans* encode sequences predicted to form thermodynamically stable complex secondary structures (F. Clark, personal communication). Genetic screens are probably also compromised by the likelihood that noncoding RNAs are less likely to be badly affected by point mutations. In Drosophila, most known mutants in ''regulatory'' regions that have strong phenotypic signatures are either large insertions or deletions. Furthermore, while there are very few known cases of point mutations in introns (or promoter regions) giving observable phenotypes in mammals, there is an unexpectedly high frequency of insertional mutants which give observable phenotypes in transgenic mice, most of which occur in introns or other noncoding regions (Meisler 1992). These observations not only strengthen the case that introns may have functions, but also suggest that these functions may only be readily revealed via extensive sequence disruption or deletion. This may also explain some of the unexpected results of gene knockouts in transgenic mice and confound interpretation of such experiments, which have not traditionally been designed to take account of introns and other non-protein-coding RNAs produced from the locus under study.

Additional evidence for large numbers of noncoding RNA transcripts in animal nuclei comes from earlier studies (preceding the discovery of introns) on the sequence complexity of heterogeneous nuclear RNA (hnRNA) (Davidson 1976), from which it was speculated that this RNA may represent regulatory transcripts (Britten and Davidson 1969; Davidson, Klein, and Britten 1977). Hybridization renaturation kinetics shows that hnRNA complexity in echinoderms is approximately 10–30 times that of mRNA (Davidson 1976), whereas we now know that protein-coding primary transcripts in vertebrates are about 5–20 times as complex as the resulting mRNAs (Deutsch and Long 1999). While these comparisons are crude, they suggest that a significant proportion of nuclear transcripts, perhaps more than half, do not contain protein-coding sequences. The nucleus of the higher organisms appears to be a very complex ball of RNA-DNA-protein interactions. On reflection, it may not be surprising that if an RNA communication network based on introns expressed in parallel

with protein-coding sequences has evolved, a higher-order control network involving eRNA alone may also have evolved. In addition, even though a substantial proportion of the human genome is composed of repeated elements, many of these are transcribed, and it is well within the bounds of possibility that they have also evolved to form part of the regulatory architecture (International Human Genome Sequencing Consortium 2001).

Examples of Gene Regulation and Communication by Introns and Noncoding RNAs

Clear-cut instances of RNA-mediated gene regulation are beginning to appear. The activities of the heterochronic genes *lin-14* and *lin-41,* which regulate developmental timing in *C. elegans,* are controlled by *lin-4* and *let-7* gene products encoding small RNAs that are antisense to repeated elements in the 3′ untranslated region of target mRNAs and which appear to inhibit translation by RNA-RNA interactions (Lee, Feinbaum, and Ambros 1993; Wightman, Ha, and Ruvkun 1993; Feinbaum and Ambros 1999; Reinhart et al. 2000), possibly by targeting the mRNA for endoribonuclease attack (Nashimoto 2000). *Lin-4* and *let-7* do not contain obvious protein-coding sequences, and the surrounding genomic sequences suggest that both are derived from functional introns surrounded by vestigial exons (Lee, Feinbaum, and Ambros 1993; Reinhart et al. 2000; L. Croft, personal communication). Moreover, *let-7* is functionally conserved in other bilaterian animals, from mollusks to mammals (Pasquinelli et al. 2000). Interestingly, the size of these RNAs (21–22 nt) is similar to that produced by the RNA interference (RNAi) pathway (Bass 2000; Parrish et al. 2000; Yang, Lu, and Erickson 2000; Zamore et al. 2000; Sharp 2001) (see below).

It has also been discovered that most small nucleolar RNAs (a group of more than 100 stable RNA molecules concentrated in the nucleolus) derive from processed introns of other genes, which encode various ribosomal proteins (e.g., L1, L5, L7, L13, S1, S3, S7, S8, S13, and others), ribosome-associated proteins (e.g., eIF-4A), nucleolar proteins (e.g., nucleolin, laminin, and fibrillarin), the heat shock protein hsc70, and the cell-cycle regulated protein RCC1, among others (Prislei et al. 1993; Sollner-Webb 1993; Bachellerie et al. 1995; Maxwell and Fournier 1995; Nicoloso et al. 1996; Rebane et al. 1998; Filipowicz et al. 1999; Filipowicz 2000). These provide both clear examples of dual gene outputs and potential instances of coordinate regulation (efference control) involving intronic sequences, in this case of ribosomal biogenesis and cell growth (Pelczar and Filipowicz 1998; Smith and Steitz 1998; Tanaka et al. 2000). More tellingly, some genes have so evolved that their protein-coding capacity no longer exists, and their primary product is intron-derived small nucleolar RNAs (Tycowski, Shu, and Steitz 1996; Bortolin and Kiss 1998; Pelczar and Filipowicz 1998; Smith and Steitz 1998; Tanaka et al. 2000), leading to the statement that ''genes generating functionally important RNAs exclusively from their intron regions are probably more

frequent than has been anticipated'' (Bortolin and Kiss 1998).

These nucleolar RNAs are processed from introns by specific mechanisms involving endonucleolytic cleavage by double-stranded RNase III–related enzymes (Caffarelli et al. 1997; Chanfreau et al. 1998; Qu et al. 1999) (also implicated in RNAi, transgene silencing, and methylation [Mette et al. 2000]; see below), exonucleolytic trimming (Cecconi, Mariottini, and Amaldi 1995; Kiss and Filipowicz 1995; Mitchell et al. 1997; Allmang et al. 1999*a*, 1999*b*; van Hoof and Parker 1999; van Hoof, Lennertz, and Parker 2000), and possibly even adjacent RNA sequences that have self-cleaving activity (Prislei et al. 1995). This processing occurs in large RNA processing complexes called exosomes, which are also involved in processing rRNA and small nuclear RNAs, contain at least 10 3′–5′ exonucleases, helicases, and RNA-binding proteins, and are found in both the nucleus and the cytoplasm (Mitchell et al. 1997; Allmang et al. 1999*a,* 1999*b*; van Hoof and Parker 1999; Mitchell and Tollervey 2000).

Intron Processing, Stability, Decay, and Memory

Intronic RNAs are more stable than is generally thought. The widespread view that excised introns are simply discarded and degraded derives from the unjustified a priori assumption that introns are nonfunctional. For example, it has been stated that ''the half-life of excised introns is of the order of a few seconds'' (Sharp et al. 1987), but closer examination of the primary literature indicates that this estimate is the time taken to splice introns from primary transcripts (Padgett et al. 1986), not the half-life of the introns themselves. Free introns are rarely observed in Northern blots, as these are mostly performed with polyA+RNA preparations and/or cDNA probes and with different questions in mind. However, when examined, free introns in both lariat and linear form have been found to be present in ''abundance'' (Zeitlin and Efstratiadis 1984), and some are relatively stable (Qian et al. 1992). In situ hybridization studies suggest that while excised intronic sequences diffuse away from the spliceosome, they remain detectable (by this relatively insensitive technique) in the nucleus, exhibiting a broad signal with a ''punctate'' (spotted) pattern (Xing et al. 1993), consistent with the possibility of a life for intron-derived RNAs within the nuclear domain, and perhaps beyond.

After splicing, introns (initially in lariat form) are debranched (Ruskin and Green 1985), a process that is itself subject to regulation (Ruskin and Green 1985; Qian et al. 1992), but subsequent events are unknown. We suggest that it is likely that excised introns are processed by specific pathways similar to those used to produce small nucleolar RNAs and which generate multiple smaller species which can function independently as *trans*-acting signals in the network (Mattick 1994), affecting the metabolism of other RNAs and the modulation of chromatin structure, among other things (see below). The intronic origins of small nucleolar RNAs became known only because of their relative stability

and abundance, and they may be just one tip of a large iceberg of a much more complex milieu (tens of thousands) of other intron-derived and other non-protein-coding RNAs, which may be more transient and in much lower individual abundance and which have not yet been detected except by their genetic signatures, as in the case of *lin-4* and *let-7.*

There are other documented examples of small *trans*-acting functional RNAs processed from longer transcripts (Sit, Vaewhongs, and Lommel 1998; Cavaille et al. 2000). There are also large numbers of ribonucleases and other RNA-related proteins in plants and animals (see below), most of whose functions and substrates are not well defined. Such processing may also involve other splicing pathways (Santoro et al. 1994; Kreivi and Lamond 1996) and guide RNAs, possibly derived from introns or other nonprotein-coding RNAs. These have been described as ''riboregulators'' (in relation to antisense RNAs) (Delihas 1995) and the ''ribotype'' (in relation to alternatively spliced mRNAs) (Herbert and Rich 1999*a*) and may be considered part of the ''soft wiring'' of the cell (Mattick 1994; Herbert and Rich 1999*b*).

The decay characteristics of eRNAs are likely to be important to their function. Both short- and long-lived eRNAs would provide a molecular memory of prior gene activation status, a significant efficiency gain over the use of bistable regulated gene networks as memories (Gardner, Cantor, and Collins 2000). Differential eRNA decay (Qian et al. 1992) and diffusion rates would create spatially and temporally complex signal pulses that enable specific communication speeds, half lives, and maximal communication radii for eRNA information transfer, allowing fine control of cellular activities. Evidence suggests that nuclear chromosomes and transcription factors are spatially organized and functionally compartmentalized and that this is dynamically affected during cellular differentiation and by transcriptional activity, as is chromatin architecture (Stenoien et al. 1998; Croft et al. 1999; Bridger et al. 2000; Vassetzky, Hair, and Mechali 2000). There is also evidence that the positions of genes are nonrandom and that the regulation of genes by antisense RNAs and ribozymes is strongly affected by their relative location (Arndt and Rank 1997), indicating that spatial relativity is important in relation to both regulatory proteins and RNAs.

Unexplained Genetic Phenomena Involving RNA

There are many mysterious molecular genetic phenomena in which the involvement of RNA has been implicated, all of which are consistent with the general thesis that *trans*-acting RNAs play important roles as regulators in cell and developmental biology. These include imprinting, transvection, suppression of transposition, position effect variegation, chromosomal methylation, co-suppression, transcriptional and posttranscriptional gene silencing, and RNA interference (RNAi). The last three of these appear to be related (Dernburg et al. 2000; Fagard et al. 2000; Hammond et

eRNAs in Eukaryotic Evolution and Development 1619

al. 2000; Ketting and Plasterk 2000; Sijen and Kooter 2000; Sharp 2001), and they may all share features in common through intersecting pathways (Judd 1995; Brenton et al. 1998; Broday, Lee, and Costa 1999; Fire 1999; Jones et al. 1999; Wu and Morris 1999; Bosher and Labouesse 2000; Mette et al. 2000; Morel et al. 2000; Wassenegger 2000; Sharp 2001).

RNAi is thought to be a mechanism for defense against double-stranded RNA (dsRNA) viruses and possibly for prevention of transposon mobilization (Tabara et al. 1999; Birchler, Bhadra, and Bhadra 2000; Bosher and Labouesse 2000; C. P. Hunter 2000; Sijen and Kooter 2000; Baulcombe 2001). RNAi was discovered by chance, when it was found that injecting dsRNA into adult *C. elegans* caused potent and specific interference of genes containing these sequences (Fire et al. 1998), and has been subsequently demonstrated in other organisms, including mice, Drosophila, zebrafish, Arabidopsis, trypanosomes, and others (Ngo et al. 1998; Bass 2000; Bosher and Labouesse 2000; Chuang and Meyerowitz 2000; Clemens et al. 2000; Sijen and Kooter 2000). RNAi occurs posttranscriptionally and appears to target principally mRNA sequences (Fire et al. 1998; Ngo et al. 1998), although there are reports that it can also target pre-mRNA (Montgomery and Fire 1998; Bosher et al. 1999). The mechanism of RNAi action appears to involve cleavage of dsRNA into 21–23-bp fragments which act as catalytic cofactors for targeted degradation of homologous mRNA sequence (Bass 2000; Hammond et al. 2000; Parrish et al. 2000; Yang, Lu, and Erickson 2000; Zamore et al. 2000; Bernstein et al. 2001), apparently involving dsRNaseIIIs of the Dicer family (Bernstein et al. 2001), RNA helicases, RNaseD-type 3′–5′ exonucleases (Mut-7), RNA-dependent RNA polymerases, some (but not all) proteins involved in nonsense-mediated mRNA decay, the protein RDE-1 (of unknown function), possibly adenosine deaminases that act on dsRNAs (ADARs), and others identified in genetic screens but yet to be defined biochemically (Bass 2000; Bosher and Labouesse 2000; Clissold and Ponting 2000; Fagard et al. 2000; Sijen and Kooter 2000). Similar mechanisms have been implicated in transgene silencing and DNA methylation (Hamilton and Baulcombe 1999; Mette et al. 2000).

RNAi is remarkably active and can cross cell and generational boundaries. It can also be made stably heritable by transgene constructs which express the dsRNA as a hairpin-loop structure from an inverted repeat (Chuang and Meyerowitz 2000; Kennerdell and Carthew 2000; Shi et al. 2000; Tavernarakis et al. 2000), raising the possibility that such sequences might also occur naturally. Intriguingly, sequences that fulfill the conditions for RNAi (inverted repeats in introns that could fold into an RNA hairpin loop which are homologous to sequences in the exons of other genes) are common in the human genome (F. Clark, personal communication).

Mutants in some of the genes associated with RNAi do not show obvious defects in growth or development (Tabara et al. 1999), but others do (Fagard et al. 2000; Smardon et al. 2000; Bernstein et al. 2001). The partial overlap between RNAi and other processes suggests that this system is very complex and probably involves multiple pathways (Fire 1999; Bosher and Labouesse 2000; Dernburg et al. 2000; Fagard et al. 2000; Mette et al. 2000; Sharp 2001), some of which almost certainly have other roles in normal cell and developmental biology. Of course, those that are crucial will be lethal. RNAi-mediated degradation of mRNA may involve cytoplasmic exosomes which are functionally distinct from nuclear exosomes involved in RNA processing and which involve different components (van Hoof and Parker 1999). Many dsRNaseIII homologs occur in metazoan genomes, as do genes encoding the Dicer family of proteins that contain similar domains (N-terminal of the ribonuclease domain and double-stranded RNA binding motif), together with an RNA helicase domain and a PAZ domain (Jacobsen, Running, and Meyerowitz 1999; Bass 2000; Cerutti, Mian, and Bateman 2000; Bernstein et al. 2001). There are also various RNaseD homologs. The RDE-1 protein is a member of a growing family (the Argonaute/piwi/zwille family) of proteins found in plants, fungi, invertebrates, and mammals which also contains a PAZ domain (Cerutti, Mian, and Bateman 2000; Baulcombe 2001; Bernstein et al. 2001), with at least 20 homologs in *C. elegans* (Bosher and Labouesse 2000), suggesting a large set of proteins with related but as yet unknown functions in RNA metabolism. There are many other types of RNases, RNA-binding proteins, and proteins that bind other forms of nucleic acids in animal and plant genomes. The presence of RNA-dependent RNA polymerases (Smardon et al. 2000) also indicates that RNA metabolism is far from well understood in the higher eukaryotes, and all of these observations (and many others of which space limitations preclude discussion) hint at a very large and very complex system of RNA-mediated gene regulation of which only some parts are yet visible. These effects may not simply be cell-autonomous, as there is evidence that RNAi and transgene silencing can act systemically in animals and plants (Fire et al. 1998; Fagard and Vaucheret 2000; Voinnet, Lederer, and Baulcombe 2000), which suggests that RNA-mediated regulation may also be involved in long-range developmental processes.

Antisense nonprotein-coding RNA transcripts have also been implicated in X-inactivation and genomic imprinting (Wutz et al. 1997; Lee, Feinbaum, and Ambrose 1999; Sleutels, Barlow, and Lyle 2000; Wroe et al. 2000), processes which also involve DNA methylation (Wutz et al. 1997; Peters et al. 1999). In plants, methylation of transgenes, and probably endogenous DNA, is RNA-directed and can involve target sequences of only 23–30 bp (Wassenegger and Pelissier 1998; Fire 1999; Jones et al. 1999; Pelissier et al. 1999; Mette et al. 2000; Pelissier and Wassenegger 2000; Wassenegger 2000). The link between DNA methylation, specific antisense RNAs, co-suppression, transcriptional and posttranscriptional gene silencing, and RNAi suggests that RNA-directed DNA methylation is involved in epigenetic gene regulation throughout the eukaryotes (Wassenegger 2000). Co-suppression has also been reported in animals (Cameron and Jennings 1991; Bingham 1997; Pal-Bhadra, Bhadra, and Birchler 1997; Bahra-

mian and Zarbl 1999; Ketting and Plasterk 2000; Plasterk and Ketting 2000) and, at least in *Drosophila* and *C. elegans,* is dependent on polycomb group proteins (Pal-Bhadra, Bhadra, and Birchler 1997, 1999; Sharp 2001), as is transgene silencing (Birchler, Bhadra, and Bhadra 2000), which implicates not only RNA but also the structure of chromatin complexes in co-suppression and gene silencing (Jones, Cowell, and Singh 2000; Morel et al. 2000; Jedrusik and Schulze 2001; Sharp 2001). This suggests that *trans*-acting RNA signals can influence chromatin structure (and hence gene activity) via Polycomb-group proteins and provides a link to another apparently unrelated and poorly understood genetic phenomenon, transvection.

Transvection and Chromatin Structure

One would predict that if eRNAs do have an important function in regulating gene expression, there should be genetic clues from intensively studied systems. A good candidate for such a system is the Drosophila bithorax complex, which is the archetypal developmental control locus and has been subjected to a considerable amount of genetic and molecular scrutiny. The bithorax region of this complex locus covers over 100 kb and contains three transcription units, one of which (*Ubx*) contains large introns and is differentially spliced to produce several variants of the morphogenetic homeobox protein UBX (Hogness et al. 1985; Duncan 1987). The others, referred to as the early and late *bxd* units, are located upstream and do not appear to encode proteins. Mutants of this locus can be classified into *Ubx* alleles, which disrupt the protein-coding sequence, and the *abx, bx, pbx,* and *bxd* alleles, which are located either within the introns of the Ubx unit (*abx, bx*) or in the 40-kb upstream region (*pbx, bxd*) and affect the spatial pattern of UBX expression. The latter alleles are thought to represent *cis*-acting regulatory sequences controlling *Ubx* transcription and are usually interpreted in terms of conventional enhancer elements, despite the fact that they are themselves transcribed. The *bxd* transcription unit produces a 27-kb transcript early in embryogenesis which has a number of large introns and is subject to differential splicing to give various small (~1.2 kb) polyA+RNAs which do not contain any significant open reading frame (Akam et al. 1985; Hogness et al. 1985; Lipshitz, Peattie, and Hogness 1987). The expression of this transcript is highly regulated during embryogenesis, in a pattern that is partially reflexive of *Ubx* transcript (Akam et al. 1985; Irish, Martinez-Arias, and Akam 1989). A number of *bxd* insertional mutations have no effect on the amount or size of the *bxd* polyA+RNA, suggesting that this species is irrelevant to the observed phenotypes and that the real import of the transcription and processing of this gene is to produce intronic RNAs (Hogness et al. 1985). The "*cis*-regulatory" elements in this region also appear to be able to regulate the expression of *Ubx* in *trans,* since defective elements can be complemented by wild-type sequences on the other chromosome.

This phenomenon (partial complementation, or "allelic cross-talk," between a mutation in a "*cis*-regulator" on one chromosome and one in the coding region of the adjacent gene on the other chromosome) has been known for many years and is termed "transvection" (Judd 1988; Pirrotta 1990). Transvection has been observed in a number of different loci and appears to be synapsis-dependent, since translocation of the "regulatory" sequences to other chromosomal sites normally diminishes or eliminates this *trans*-complementation of gene expression patterns (Judd 1988; Pirrotta 1990; Wu and Morris 1999). Mechanistically, this has been interpreted in terms of enhancer elements from one copy of the gene being able to interact directly with its homolog on the other chromosome (i.e., to influence both promoters) because of their close alignment (Geyer, Green, and Corces 1990), although there are other propositions, mostly based on the same theme of chromosome pairing (Wu and Morris 1999). However, translocation of these regulatory sequences can in fact lead to a spectrum of transvection effects, ranging from weak to strong, suggesting that remote action is possible (Micol, Castelli-Gair, and Garcia-Bellido 1990) and that a simple model of chromosome pairing and transcriptional crossover is incorrect (Goldsborough and Kornberg 1996). Moreover, these effects may be simply interpreted by regarding the "*cis*-acting regulatory regions" as encoding separate (noncoding RNA) genes.

Transvection at distance is accentuated in the presence of mutant alleles of the *Polycomb* gene (which normally acts to maintain repression of transcription of *Ubx* and other genes in cells where it was not initially activated) and at many loci is dependent on the *zeste* gene product, which acts in opposition to polycomb-group proteins to enhance transcription (Wu and Goldberg 1989; Laney and Biggin 1992; Pirrotta 1999), indicating that factors other than chromosome pairing are involved in this process (Castelli-Gair and Garcia-Bellido 1990; Castelli-Gair, Micol, and Garcia-Belido 1990). *Zeste* null mutants do not affect chromosome pairing, even though transvection at some loci is entirely dependent on *zeste* (Gemkow, Verveer, and Arndt-Jovin 1998; Pirrotta 1999). Moreover, it has been shown that a region in the vicinity of the late *bxd* transcript which can attenuate *Ubx* expression can exert its action independent of its position (Castelli-Gair et al. 1992; Castelli-Gair, Muller, and Bienz 1992). To explain such observations, one has to invoke either DNA looping over enormous (interchromosomal) distances to bring regulatory proteins into contact with the *Ubx* promoter or a (diffusible) substance expressed from these sequences, i.e., RNA. It is worth recalling that, as mentioned above, the nucleus is highly ordered and at least some RNA-regulated interactions in the nucleus are known to be distance-, or at least location-, dependent. Transvection-mediated expression of *Ubx* can also be affected by mutant *Cbx* alleles, which are located within the second intron of *Ubx* (Castelli-Gair, Micol, and Garcia-Belido 1990; Goldsborough and Kornberg 1996). These alleles (which cause ectopic expression of *Ubx* in imaginal discs) are suppressed by mutations in *zeste* and by chro-

mosome rearrangements, which also reduce transcription from both homologs. For these types of reasons, it is thought that *trans*-activation by transvection involves the same type of interactions that normally control gene expression in *cis* (Goldsborough and Kornberg 1996; Muller et al. 1999).

Similar observations have been made for the downstream *abdA–AbdB* region of the bithorax complex, which also encodes homeotic proteins controlling segment identity. As in the case of bithorax itself, the sequences upstream of *abdA* and *AbdB,* which are referred to as the infra-abdominal (*iab*) region, are thought to function as *cis*-acting regulatory elements, despite the fact that this region, like *bxd,* is also itself transcribed. Transvection (involving *iab* and *abdA*/*AbdB* alleles) at this locus is synapsis (pairing) independent and relatively insensitive to location, again suggesting that a *trans*-acting RNA may be involved (Hendrickson and Sakonju 1995; Hopmann, Duncan, and Duncan 1995; Sipos et al. 1998). The efficiency of this transvection is also different in different tissues, indicating that the state of differentiation has an effect on this process (Sipos et al. 1998). Another (small, 800 bp) "element" in this region (*Mcp*) has also been shown to be capable of "*trans*-silencing," independent of homology or homology pairing in the immediate vicinity of *Mcp* transgene inserts, leading Muller et al. (1999) to pose the question: "how does such a short DNA sequence interact specifically with an *Mcp* partner over large distances, and why does it differ from other DNAs?" Leaving aside the problem posed by the second half of this question, it was speculated that the answer to the first may be that this element searches and finds its homolog by way of a "constrained random walk" within nuclear compartments, which must be repeated after each mitosis (Muller et al. 1999). A more parsimonious explanation for both questions is that *Mcp* encodes a *trans*-acting RNA whose ability to communicate with its target loci is affected by spatial separation and by polycomb/zeste-mediated effects on chromatin architecture.

These genetic events are complex, show locus- and allele-specific idiosyncrasies, and are extremely difficult to unravel. There are bewildering combinations of genotypes and phenotypes to sift through, many of the mutations studied have not been characterized at the sequence level, and experimental design and data analysis has often been tacitly couched within the current models of gene regulation, which makes conclusions difficult to reach if the correct explanations lie outside the current paradigm. However, it is likely that transvection is a general phenomenon in gene control in the higher eukaryotes (Bollmann, Carpenter, and Coen 1991; Tsai and Silver 1991; Aramayo and Metzenberg 1996), albeit most obvious in Drosophila because of the powerful and intense genetic analysis of this species.

Transvection has also been implicated in genomic imprinting and X-chromosome inactivation in mammals (Tsai and Silver 1991; Marahrens 1999). *Polycomb* and *zeste* are clearly involved in mediating transvection, although the relative effects of these proteins are locus-dependent, as are their effects on developmental phe-

notypes (Campbell et al. 1995). The trithorax group (TrxG) of activators (which includes *zeste*) and the polycomb group (PcG) of repressors (Campbell et al. 1995; Gould 1997) are multigene families that are believed to control the expression of several key developmental regulators by changing the structure of chromatin (Judd 1995; Gebuhr, Bultman, and Magnuson 2000), although, with one exception (Brown et al. 1998), they do not appear to bind DNA per se (Zink et al. 1991), and their target specificity is not known: TrxG- and PcG-response elements are definable only in vivo (Tillib et al. 1999; Farkas, Leibovitch, and Elgin 2000). These genes not only influence transvection but also the correct spatial expression of genes. They are thought to be responsible for the maintenance of transcriptional regulation by providing a "cellular memory mechanism throughout development" (Kennison 1995; Hanson et al. 1999; Jacobs and van Lohuizen 1999; Gebuhr, Bultman, and Magnuson 2000) by altering chromatin structure, but what determines their activity in different lineages is unknown. They are required only during active phases of development, as once the chromatin conformation is fixed, it remains stable, possibly through deacetylation of histones (van der Vlag and Otte 1999; van Lohuizen 1999). Homologs of these genes occur in plants and mammals and are probably a general feature of the biology of the higher eukaryotes (Goodrich et al. 1997; Gould 1997; Schumacher and Magnuson 1997; Hashimoto et al. 1998). They also appear to act in large heterogeneous multimeric complexes, containing more than one member of the family, which in many cases are gene- and allele-specific (Campbell et al. 1995; Gould 1997; Strutt and Paro 1997; Hashimoto et al. 1998; Kyba and Brock 1998; van der Vlag and Otte 1999; Farkas, Leibovitch, and Elgin 2000). As noted already, polycomb group proteins have also been shown to influence co-suppression and gene silencing, which is RNA-dependent and involves methylation (Jones, Thomas, and Maule 1998; Jones et al. 1999; Morel et al. 2000), leading to the suggestion that *trans*-acting RNAs may direct the gene-specific binding of polycomb complexes (Sharp 2001). Polycomb group proteins are also involved in transgene silencing (Birchler, Bhadra, and Bhadra 2000), which also involves homology-dependent RNA mechanisms and methylation (Baulcombe 1996; Broday, Lee, and Costa 1999; Jones et al. 1999), as well as histone H1.1 (Jedrusik and Schulze 2001).

Significantly, it has recently been shown that a conserved domain called a chromodomain, which occurs in polycomb-group proteins, as well as in other proteins involved in chromatin remodeling, such as the HP1 and CHD families (Jones, Cowell, and Singh 2000), is an RNA-binding module (Akhtar, Zink, and Becker 2000). Furthermore, association of the histone acetyltransferase MOF with the male X chromosome in Drosophila depends on its binding to the nonprotein-coding RNA *roX2* via its chromodomain (Akhtar, Zink, and Becker 2000). In this context, it is also interesting that a nonprotein-coding RNA has been shown to act as a transcriptional coactivator for steroid receptors (Lanz et al. 1999), whose action also requires chromatin remodeling

and the recruitment of histone acetyltransferases (Zhang and Lazar 2000).

Thus, all of these genetic phenomena are connected, with the common features being nonprotein-coding RNAs and dynamic interactions and remodeling of chromatin involving DNA methylation and trithorax- and polycomb-group proteins occurring in large complexes with a variety of other proteins, including histone-modifying factors and transcription factors. The influence on transvection and other phenomena of complexes containing trithorax- and polycomb-group proteins may therefore be interpreted more easily in terms of maintaining, enhancing, or inhibiting accessibility of these sites to *trans*-acting RNAs and/or executing signals from such RNAs. The fact that *zeste* mutants often die during development but adult survivors are relatively healthy (Goldberg, Colvin, and Mellin 1989) suggests that such communication is most critical during development, as one might predict would be the case.

In this context, it is relevant to note that the target specificity of "transcription factors" may not be duplex DNA but higher-order structures. For example, it has been shown that some zinc finger proteins (such as Sp1) that are considered conventional transcription factors have a comparable or greater affinity for RNA-DNA hybrids than for double-stranded DNA, which is strand-specific (Shi and Berg 1995). A number of other "transcription factors" including Y-box (cold shock) proteins are also able to bind RNA (Ladomery 1997; Matsumoto and Wolffe 1998; Shnyreva et al. 2000). Other domains found in regulatory proteins in the higher eukaryotes, such as HOX, PAX, LIM, brahma/SWI/SNF complexes, forkhead/winged helix-loop-helix proteins, etc., may also in fact bind not (just) to duplex DNA but to other nucleic acid structures involving RNA, including triplexes (which may also be involved in the catalytic mechanism of RNAi). The adenosine deaminases that act on dsRNAs (ADARS) and play a role in RNAi (Bass 2000) (see above) have been shown to contain domains (related to winged helix-turn-helix domains and the globular domain of histone H5; Herbert and Rich 1999*c*) which bind Z-DNA (Herbert et al. 1995, 1997, 1998; Herbert and Rich 1999*c*) and/or catalyze its formation (Kim et al. 2000).

Genetic Programming and the Evolution of Complex Organisms

The evolution of complex phenotypes is usually understood to proceed by a sequence from cells that were entirely unregulated and whose dynamics were governed by rate processes and input constraints. The existence of these cells provided the preconditions for the appearance of regulatory mechanisms which fine-tuned rate processes. We propose that these regulated networks, following a change in gene structure and output in the eukaryotic lineage, provided the necessary precondition for the appearance of controlled multitasked networks, which in turn led to the appearance of programmed response networks capable of implementing stored sequences of dynamical activities in response

to internal and external stimuli. Furthermore, we suggest that there is only one plausible mechanism for the evolution and control of multitasking in cell and developmental biology and that, far from being evolutionary junk, nuclear introns and other nonprotein-coding RNAs have evolved this function.

The majority of information in a multitasked network is held in control sequences. Nonprotein-coding RNAs compose the majority of the genomic output and unique sequence information in the higher eukaryotes, and the evidence is growing that these RNAs are functional, as is the realization that RNA metabolism in these organisms is much more complex than previously realized.

The three critical steps in the evolution of this system were (1) the entry of introns into protein-coding genes in the eukaryotic lineage, (2) the subsequent relaxation of internal sequence constraints through the evolution of the spliceosome and the exploration of new sequence space, and (3) the co-evolution of processing and receiver mechanisms for *trans*-acting RNAs, which are not yet well characterized but are likely to involve the dynamic modeling and remodeling of chromatin and DNA, as well as RNA-RNA and RNA-protein interactions in other parts of the cell. Steps 2 and 3 probably occurred, at least initially, through constructive neutral evolution (Stoltzfus 1999), involving biased variation, epistatic interactions, and excess capacities underlying a complex series of steps giving rise to novel structures and operations, and later through molecular co-evolution (Dover and Tautz 1986). Once this system of RNA communication began to be established, the rate of evolution of functional introns would have accelerated (by positive selection) and led also to the evolution of other nonprotein-coding RNAs, which are also usually spliced and are probably derived from genes that had lost their protein-coding capacity, as appears to have occurred in the case of transcripts producing small nucleolar RNAs.

In practical terms then, we propose that functional introns provide a cellular memory of recent transcriptional events and underpin a multiple output parallel processing system in which gene activity at one locus can connect to other genes and gene products in real time, allowing integration and multitasking of a sophisticated network of cellular activity. In this scheme, nonprotein-coding RNAs are control molecules in the network that do not require concomitant production of protein. Thus, there are two levels of information produced by gene expression in the higher organisms—mRNA and eRNA—allowing the concomitant expression of both structural (i.e., protein-coding) and networking information, with the latter involving multiplex contacts between different genes and gene products via RNA signals that are implicit in primary transcripts. As some genes have evolved to express only eRNA and some genes lack introns, there are three types of genes in the higher organisms—those that encode only protein (which are rare), those that encode only eRNA, and those that encode both.

One prediction of this model is that many core proteins in the higher eukaryotes will be multitasked, i.e.,

have different roles in different subnetworks to produce different phenotypic outcomes. This appears to occur. For example, it has been shown that glycogen synthase kinase-3β participates in both the specification of the vertebrate embryonic dorsoventral axis (via the Wnt/ wingless signaling pathway) and the NF-κB-mediated cell survival response following TNF activation (Hoeflich et al. 2000). Both cytochrome *c* and a flavoprotein (apoptosis-inducing factor) have redox functions in mitochondria as well as specific apoptogenic functions (Chinnaiyan 1999; Daugas et al. 2000; Loeffler and Kroemer 2000). The XPD gene product functions in both transcription and excision repair of DNA (Lehmann 2001). There are many other documented examples of proteins that participate in more than one developmental and signaling pathway (subnetwork) (see, e.g., Boutros and Mlodzik 1999; Szebenyi and Fallon 1999; Coffey et al. 2000; O'Brien et al. 2000). There are also examples of proteins having different, even antagonistic, functions in different settings, often as a result of alternative splicing (Jiang and Wu 1999; Lopez 1998; Hastings et al. 2000), a process that we predict will turn out to be regulated and guided not simply by tissue-specific RNA binding proteins/splicing factors, but also by *trans*-acting RNAs produced by the activity of other genes (see, e.g., Hastings et al. 2000). Consequently, developmental and phylogenetic profiling efforts will need to assign a range of biological, in addition to biochemical, functions to individual proteins and their splice variants in the network.

A multitasked network allows the rapid exploration of exponentially many protein expression profiles without equivalent increase in the size of the controlled parent network. The model therefore also predicts that the core proteome will be relatively stable in the higher organisms, which appears to be the case (Duboule and Wilkins 1998; Rubin et al. 2000), and that phenotypic variation will result primarily and quite easily from variation in the control architecture, rather than duplication and mutation of gene subnetworks. Once in place, therefore, a controlled multitasked network enables not only the efficient programming of different cellular phenotypes in the differentiation and development of multicellular organisms, but also rapid evolutionary radiation during expansions into uncontested environments, such as that initially observed in the Cambrian explosion and those seen after major extinction events.

The corollary is that prokaryotes and simpler eukaryotes operating on simple protein control circuitry are limited in their phenotypic range, genome size, and complexity not by the available diversity of polypeptide structures and chemistry, but by a primitive genetic operating system incapable of supporting integrated multitasking of gene networks. This would also explain why the earth was restricted to simpler unicellular and colonial life forms for over 3 billion years, the rapid evolution of complex life forms after the conditions for feasible parallel outputs were satisfied by the entry of introns into the eukaryotic lineage around 1.2 billion years ago, and the subsequent evolution of the necessary infrastructure for sending and receiving intronic and other nonprotein-coding RNA signals.

Genomes are data sets with controls. Our hypothesis examines biology and genomes from the viewpoint of information and network theory and unifies a wide range of evolutionary and molecular genetic observations, including the long lag followed by the sudden appearance of developmentally sophisticated multicellular organisms, the plasticity of phenotypic diversity despite the relative conservation of the core proteome, and a wide range of unexplained molecular genetic phenomena that all intersect with RNA, the enabling molecule. If correct, this would force a fundamental reassessment of our understanding of genetic programming in the higher organisms, with significant scientific and practical consequences.

Acknowledgments

LITERATURE CITED

AKAM, M. E., A. MARTINEZ-ARIAS, R. WEINZIERL, and C. D. WILDE. 1985. Function and expression of ultrabithorax in the *Drosophila* embryo. Cold Spring Harb. Symp. Quant. Biol. **50**:195–200.

AKHTAR, A., D. ZINK, and P. B. BECKER. 2000. Chromodomains are protein-RNA interaction modules. Nature **407**: 405–409.

ALBERT, R., H. JEONG, and A. L. BARABASI. 2000. Error and attack tolerance of complex networks. Nature **406**:378–382.

ALLMANG, C., J. KUFEL, G. CHANFREAU, P. MITCHELL, E. PETFALSKI, and D. TOLLERVEY. 1999*a*. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J. **18**:5399–5410.

ALLMANG, C., E. PETFALSKI, A. PODTELEJNIKOV, M. MANN, D. TOLLERVEY, and P. MITCHELL. 1999*b*. The yeast exosome and human PM-Scl are related complexes of 3′→5′ exonucleases. Genes Dev. **13**:2148–2158.

ALMEIDA, A. C., V. M. FERNANDES DE LIMA, and A. F. INFANTOSI. 1998. Mathematical model of the CA1 region of the rat hippocampus. Phys. Med. Biol. **43**:2631–2646.

ANDERSEN, R. A., L. H. SNYDER, D. C. BRADLEY, and J. XING. 1997. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu. Rev. Neurosci. **20**:303–330.

ARAMAYO, R., and R. L. METZENBERG. 1996. Meiotic transvection in fungi. Cell **86**:103–113.

ARNDT, G. M., and G. H. RANK. 1997. Colocalization of antisense RNAs and ribozymes with their target mRNAs. Genome **40**:785–797.

ARUSCAVAGE, P. J., and B. L. BASS. 2000. A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. RNA **6**:257–269.

ASHE, H. L., J. MONKS, M. WIJGERDE, P. FRASER, and N. J. PROUDFOOT. 1997. Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev. **11**:2494–2509.

BACHELLERIE, J. P., M. NICOLOSO, L. H. QU, B. MICHOT, M. CAIZERGUES-FERRER, J. CAVAILLE, and M. H. RENALIER. 1995. Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. Biochem. Cell Biol. **73**: 835–843.

BAGAVATHI, S., and R. MALATHI. 1996. Introns and protein revolution—an analysis of the exon/intron organisation of actin genes. FEBS Lett. **392**:63–65.

BAHRAMIAN, M. B., and H. ZARBL. 1999. Transcriptional and posttranscriptional silencing of rodent alpha1(I) collagen by a homologous transcriptionally self-silenced transgene. Mol. Cell. Biol. **19**:274–283.

BASS, B. L. 2000. Double-stranded RNA as a template for gene silencing. Cell **101**:235–238.

BAULCOMBE, D. C. 1996. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol. Biol. **32**:79–88.

———. 2001. Diced defence. Nature **409**:295–296.

BECSKEI, A., and L. SERRANO. 2000. Engineering stability in gene networks by autoregulation. Nature **405**:590–593.

BERNSTEIN, E., A. A. CAUDY, S. M. HAMMOND, and G. J. HANNON. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature **409**:363–366.

BHALLA, U. S., and R. IYENGAR. 1999. Emergent properties of networks of biological signaling pathways. Science **283**: 381–387.

BINGHAM, P. M. 1997. Cosuppression comes to the animals. Cell **90**:385–387.

BIRCHLER, J. A., M. P. BHADRA, and U. BHADRA. 2000. Making noise about silence: repression of repeated genes in animals. Curr. Opin. Genet. Dev. **10**:211–216.

BOLLMANN, J., R. CARPENTER, and E. S. COEN. 1991. Allelic interactions at the nivea locus of *Antirrhinum.* Plant Cell **3**: 1327–1336.

BORTOLIN, M. L., and T. KISS. 1998. Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. RNA **4**: 445–454.

BOSHER, J. M., P. DUFOURCQ, S. SOOKHAREEA, and M. LABOUESSE. 1999. RNA interference can target pre-mRNA: consequences for gene expression in a *Caenorhabditis elegans* operon. Genetics **153**:1245–1256.

BOSHER, J. M., and M. LABOUESSE. 2000. RNA interference: genetic wand and genetic watchdog. Nat. Cell. Biol. **2**:E31–E36.

BOUTROS, M., and M. MLODZIK. 1999. Dishevelled: at the crossroads of divergent intracellular signaling pathways. Mech. Dev. **83**:27–37.

BRENNER, S., G. ELGAR, R. SANDFORD, A. MACRAE, B. VENKATESH, and S. APARICIO. 1993. Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. Nature **366**:265–268.

BRENTON, J. D., J. F. AINSCOUGH, F. LYKO, R. PARO, and M. A. SURANI. 1998. Imprinting and gene silencing in mice and *Drosophila.* Novartis Found. Symp. **214**:233–244.

BRIDGEMAN, B. 1995. A review of the role of efference copy in sensory and oculomotor control systems. Ann. Biomed. Eng. **23**:409–422.

BRIDGER, J. M., S. BOYLE, I. R. KILL, and W. A. BICKMORE. 2000. Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. Curr. Biol. **10**:149–152.

BRITTEN, R. J., and E. H. DAVIDSON. 1969. Gene regulation for higher cells: a theory. Science **165**:349–357.

BROCKDORFF, N. 1998. The role of Xist in X-inactivation. Curr. Opin. Genet. Dev. **8**:328–333.

BRODAY, L., Y. W. LEE, and M. COSTA. 1999. 5-azacytidine induces transgene silencing by DNA methylation in Chinese hamster cells. Mol. Cell. Biol. **19**:3198–3204.

BROWN, J. L., D. MUCCI, M. WHITELEY, M. L. DIRKSEN, and J. A. KASSIS. 1998. The *Drosophila* Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol. Cell **1**:1057–1064.

CAFFARELLI, E., L. MAGGI, A. FATICA, J. JIRICNY, and I. BOZZONI. 1997. A novel Mn^{++}-dependent ribonuclease that functions in U16 SnoRNA processing in *X. laevis.* Biochem. Biophys. Res. Commun. **233**:514–517.

CAMERON, F. H., and P. A. JENNINGS. 1991. Inhibition of gene expression by a short sense fragment. Nucleic Acids Res. **19**:469–475.

CAMPBELL, R. B., D. A. SINCLAIR, M. COULING, and H. W. BROCK. 1995. Genetic interactions and dosage effects of Polycomb group genes of *Drosophila.* Mol. Gen. Genet. **246**:291–300.

CARELS, N., and G. BERNARDI. 2000. Two classes of genes in plants. Genetics **154**:1819–1825.

CASTELLI-GAIR, J. E., M. P. CAPDEVILA, J. L. MICOL, and A. GARCIA-BELLIDO. 1992. Positive and negative cis-regulatory elements in the bithoraxoid region of the *Drosophila* Ultrabithorax gene. Mol. Gen. Genet. **234**:177–184.

CASTELLI-GAIR, J. E., and A. GARCIA-BELLIDO. 1990. Interactions of Polycomb and trithorax with cis regulatory regions of Ultrabithorax during the development of *Drosophila* melanogaster. EMBO J. **9**:4267–4275.

CASTELLI-GAIR, J. E., J. L. MICOL, and A. GARCIA-BELLIDO. 1990. Transvection in the *Drosophila* Ultrabithorax gene: a Cbx1 mutant allele induces ectopic expression of a normal allele in trans. Genetics **126**:177–184.

CASTELLI-GAIR, J., J. MULLER, and M. BIENZ. 1992. Function of an Ultrabithorax minigene in imaginal cells. Development **114**:877–886.

CAVAILLE, J., K. BUITING, M. KIEFMANN, M. LALANDE, C. I. BRANNAN, B. HORSTHEMKE, J. P. BACHELLERIE, J. BROSIUS, and A. HUTTENHOFER. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. USA **97**: 14311–14316.

CAVALIER-SMITH, T. 1991. Intron phylogeny: a new hypothesis. Trends Genet. **7**:145–148.

CECCONI, F., C. CROSIO, P. MARIOTTINI, G. CESARENI, M. GIORGI, S. BRENNER, and F. AMALDI. 1996. A functional role for some *Fugu* introns larger than the typical short ones: the example of the gene coding for ribosomal protein S7 and snoRNA U17. Nucleic Acids Res. **24**:3167–3172.

CECCONI, F., P. MARIOTTINI, and F. AMALDI. 1995. The Xenopus intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precursor in oocytes. Nucleic Acids Res. **23**:4670–4676.

CERUTTI, L., N. MIAN, and A. BATEMAN. 2000. Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the piwi domain. Trends Biochem. Sci. **25**:481–482.

CHANFREAU, G., G. ROTONDO, P. LEGRAIN, and A. JACQUIER. 1998. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. EMBO J. **17**:3726–3737.

CHERVITZ, S. A., L. ARAVIND, G. SHERLOCK et al. (13 co-authors). 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science **282**:2022–2028.

CHINNAIYAN, A. M. 1999. The apoptosome: heart and soul of the cell death machine. Neoplasia **1**:5–15.

CHO, G., and R. F. DOOLITTLE. 1997. Intron distribution in ancient paralogs supports random insertion and not random loss. J. Mol. Evol. **44**:573–584.

CHUANG, C. F., and E. M. MEYEROWITZ. 2000. Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA **97**:4985–4990.

CLEMENS, J. C., C. A. WORBY, N. SIMONSON-LEFF, M. MUDA, T. MAEHAMA, B. A. HEMMINGS, and J. E. DIXON. 2000. Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. Proc. Natl. Acad. Sci. USA **97**:6499–6503.

CLISSOLD, P. M., and C. P. PONTING. 2000. PIN domains in nonsense-mediated mRNA decay and RNAi. Curr. Biol. **10**: R888–R890.

COFFEY, E. T., V. HONGISTO, M. DICKENS, R. J. DAVIS, and M. J. COURTNEY. 2000. Dual roles for c-Jun N-terminal kinase in developmental and stress responses in cerebellar granule neurons. J. Neurosci. **20**:7602–7613.

COUSINEAU, B., S. LAWRENCE, D. SMITH, and M. BELFORT. 2000. Retrotransposition of a bacterial group II intron. Nature **404**:1018–1021.

CROFT, J. A., J. M. BRIDGER, S. BOYLE, P. PERRY, P. TEAGUE, and W. A. BICKMORE. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. J. Cell Biol. **145**:1119–1131.

CROFT, L., S. SCHANDORFF, F. CLARK, K. BURRAGE, P. ARCTANDER, and J. S. MATTICK. 2000. ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. Nat. Genet. **24**:340–341.

DANO, S., P. G. SORENSEN, and F. HYNNE. 1999. Sustained oscillations in living cells. Nature **402**:320–322.

DAUGAS, E., D. NOCHY, L. RAVAGNAN, M. LOEFFLER, S. A. SUSIN, N. ZAMZAMI, and G. KROEMER. 2000. Apoptosis-inducing factor (AIF): a ubiquitous mitochondrial oxidoreductase involved in apoptosis. FEBS Lett. **476**:118–123.

DAVIDSON, E. H. 1976. Gene activity in early development. Academic Press, New York.

DAVIDSON, E. H., W. H. KLEIN, and R. J. BRITTEN. 1977. Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. Dev. Biol. **55**: 69–84.

DELIHAS, N. 1995. Regulation of gene expression by trans-encoded antisense RNAs. Mol. Microbiol. **15**:411–414.

DERNBURG, A. F., J. ZALEVSKY, M. P. COLAIACOVO, and A. M. VILLENEUVE. 2000. Transgene-mediated cosuppression in the *C. elegans* germ line. Genes Dev. **14**:1578–1583.

DEUTSCH, M., and M. LONG. 1999. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res. **27**:3219–3228.

DOVER, G. A., and D. TAUTZ. 1986. Conservation and divergence in multigene families: alternatives to selection and drift. Philos. Trans. R. Soc. Lond. B Biol. Sci. **312**:275–289.

DUBOULE, D., and A. S. WILKINS. 1998. The evolution of 'bricolage.' Trends Genet. **14**:54–59.

DUNCAN, I. 1987. The bithorax complex. Annu. Rev. Genet. **21**:285–319.

EDDY, S. R. 1999. Noncoding RNA genes. Curr. Opin. Genet. Dev. **9**:695–699.

EICKBUSH, T. H. 2000. Molecular biology: introns gain ground. Nature **404**:940–941.

ELGAR, G. 1996. Quality not quantity: the pufferfish genome. Hum. Mol. Genet. **5**:1437–1442.

ELMAN, J. L. 1998. Connectionism, artificial life, and dynamical systems: new approaches to old questions. Pp. 488–505 *in* W. BECHTEL and G. GRAHAM, eds. A companion to cognitive science. Blackwell Science.

ELOWITZ, M. B., and S. LEIBLER. 2000. A synthetic oscillatory network of transcriptional regulators. Nature **403**:335–338.

ERDMANN, V. A., M. SZYMANSKI, A. HOCHBERG, N. DE GROOT, and J. BARCISZEWSKI. 1999. Collection of mRNA-like noncoding RNAs. Nucleic Acids Res. **27**:192–195.

FABRY, S., A. JACOBSEN, H. HUBER, K. PALME, and R. SCHMITT. 1993. Structure, expression, and phylogenetic relationships of a family of ypt genes encoding small G-proteins in the green alga *Volvox carteri.* Curr. Genet. **24**:229–240.

FAGARD, M., S. BOUTET, J. B. MOREL, C. BELLINI, and H. VAUCHERET. 2000. AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc. Natl. Acad. Sci. USA **97**:11650–11654.

FAGARD, M., and H. VAUCHERET. 2000. Systemic silencing signal(s). Plant Mol. Biol. **43**:285–293.

FARKAS, G., B. A. LEIBOVITCH, and S. C. ELGIN. 2000. Chromatin organization and transcriptional control of gene expression in *Drosophila.* Gene **253**:117–136.

FEINBAUM, R., and V. AMBROS. 1999. The timing of lin-4 RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans.* Dev. Biol. **210**: 87–95.

FERAT, J. L., and F. MICHEL. 1993. Group II self-splicing introns in bacteria. Nature **364**:358–361.

FILIPOWICZ, W. 2000. Imprinted expression of small nucleolar RNAs in brain: time for RNomics. Proc. Natl. Acad. Sci. USA **97**:14035–14037.

FILIPOWICZ, W., P. PELCZAR, V. POGACIC, and F. DRAGON. 1999. Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. Acta Biochim. Pol. **46**:377–389.

FIRE, A. 1999. RNA-triggered gene silencing. Trends Genet. **15**:358–363.

FIRE, A., S. XU, M. K. MONTGOMERY, S. A. KOSTAS, S. E. DRIVER, and C. C. MELLO. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans.* Nature **391**:806–811.

GARBE, J. C., and M. L. PARDUE. 1986. Heat shock locus 93D of *Drosophila melanogaster*: a spliced RNA most strongly conserved in the intron sequence. Proc. Natl. Acad. Sci. USA **83**:1812–1816.

GARDNER, T. S., C. R. CANTOR, and J. J. COLLINS. 2000. Construction of a genetic toggle switch in *Escherichia coli.* Nature **403**:339–342.

GEBUHR, T. C., S. J. BULTMAN, and T. MAGNUSON. 2000. Pc-G/trx-G and the SWI/SNF connection: developmental gene regulation through chromatin remodeling. Genesis **26**:189–197.

GEMKOW, M. J., P. J. VERVEER, and D. J. ARNDT-JOVIN. 1998. Homologous association of the bithorax-complex during embryogenesis: consequences for transvection in *Drosophila melanogaster.* Development **125**:4541–4552.

GERHART, J., and M. KIRSCHNER. 1997. Cells, embryos and evolution: toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptibility. Blackwell Science, Malden, Mass.

GEYER, P. K., M. M. GREEN, and V. G. CORCES. 1990. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in *Drosophila.* EMBO J. **9**:2247–2256.

GOLDBERG, M. L., R. A. COLVIN, and A. F. MELLIN. 1989. The *Drosophila* zeste locus is nonessential. Genetics **123**:145–155.

GOLDSBOROUGH, A. S., and T. B. KORNBERG. 1996. Reduction of transcription by homologue asynapsis in *Drosophila* imaginal discs. Nature **381**:807–810.

GOODRICH, J., P. PUANGSOMLEE, M. MARTIN, D. LONG, E. M. MEYEROWITZ, and G. COUPLAND. 1997. A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis.* Nature **386**:44–51.

GOULD, A. 1997. Functions of mammalian Polycomb group and trithorax group related genes. Curr. Opin. Genet. Dev. **7**:488–494.

HAASE, S. B., and S. I. REED. 1999. Evidence that a free-running oscillator drives G1 events in the budding yeast cell cycle. Nature **401**:394–397.

HAMILTON, A. J., and D. C. BAULCOMBE. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286**:950–952.

HAMMOND, S. M., E. BERNSTEIN, D. BEACH, and G. J. HANNON. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature **404**:293–296.

HANSON, R. D., J. L. HESS, B. D. YU et al. (11 co-authors). 1999. Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. Proc. Natl. Acad. Sci. USA **96**:14372–14377.

HARTWELL, L. H., J. J. HOPFIELD, S. LEIBLER, and A. W. MURRAY. 1999. From molecular to modular cell biology. Nature **402**:C47–C52.

HASHIMOTO, N., H. W. BROCK, M. NOMURA, M. KYBA, J. HODGSON, Y. FUJITA, Y. TAKIHARA, K. SHIMADA, and T. HIGASHINAKAGAWA. 1998. RAE28, BMI1, and M33 are members of heterogeneous multimeric mammalian Polycomb group complexes. Biochem. Biophys. Res. Commun. **245**:356–365.

HASTINGS, M. L., H. A. INGLE, M. A. LAZAR, and S. H. MUNROE. 2000. Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. J. Biol. Chem. **275**:11507–11513.

HASTY, J., J. PRADINES, M. DOLNIK, and J. J. COLLINS. 2000. Noise-based switches and amplifiers for gene expression. Proc. Natl. Acad. Sci. USA **97**:2075–2080.

HAYASHI, T., K. MAKINO, M. OHNISHI et al. (22 co-authors). 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**:11–22.

HENDRICKSON, J. E., and S. SAKONJU. 1995. Cis and trans interactions between the iab regulatory regions and abdominal-A and abdominal-B in *Drosophila melanogaster.* Genetics **139**:835–848.

HERBERT, A., J. ALFKEN, Y. G. KIM, I. S. MIAN, K. NISHIKURA, and A. RICH. 1997. A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. Proc. Natl. Acad. Sci. USA **94**:8421–8426.

HERBERT, A., K. LOWENHAUPT, J. SPITZNER, and A. RICH. 1995. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. Proc. Natl. Acad. Sci. USA **92**:7550–7554.

HERBERT, A., and A. RICH. 1999a. RNA processing and the evolution of eukaryotes. Nat. Genet. **21**:265–269.

———. 1999b. RNA processing in evolution: the logic of soft-wired genomes. Ann. N.Y. Acad. Sci. **870**:119–132.

———. 1999c. Left-handed Z-DNA: structure and function. Genetica **106**:37–47.

HERBERT, A., M. SCHADE, K. LOWENHAUPT, J. ALFKEN, T. SCHWARTZ, L. S. SHLYAKHTENKO, Y. L. LYUBCHENKO, and A. RICH. 1998. The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences. Nucleic Acids Res. **26**:3486–3493.

HOEFLICH, K. P., J. LUO, E. A. RUBIE, M. S. TSAO, O. JIN, and J. R. WOODGETT. 2000. Requirement for glycogen synthase kinase-3β in cell survival and NF-kappaB activation. Nature **406**:86–90.

HOGNESS, D. S., H. D. LIPSHITZ, P. A. BEACHY, D. A. PEATTIE, R. B. SAINT, M. GOLDSCHMIDT-CLERMONT, P. J. HARTE, E. R. GAVIS, and S. L. HELFAND. 1985. Regulation and products of the Ubx domain of the bithorax complex. Cold Spring Harb. Symp. Quant. Biol. **50**:181–194.

HOLLAND, P. W. 1999. The future of evolutionary developmental biology. Nature **402**:C41–C44.

HONG, Y. K., S. D. ONTIVEROS, and W. M. STRAUSS. 2000. A revision of the human XIST gene organization and structural comparison with mouse Xist. Mamm. Genome **11**:220–224.

HOPMANN, R., D. DUNCAN, and I. DUNCAN. 1995. Transvection in the iab-5,6,7 region of the bithorax complex of *Drosophila*: homology independent interactions in trans. Genetics **139**:815–833.

HUANG, F. 1998. Syntagms in development and evolution. Int. J. Dev. Biol. **42**:487–494.

HUNTER, C. P. 2000. Gene silencing: shrinking the black box of RNAi. Curr. Biol. **10**:R137–140.

HUNTER, T. 2000. Signaling—2000 and beyond. Cell **100**:113–127.

HURST, L. D., and N. G. SMITH. 1999. Molecular evolutionary evidence that H19 mRNA is functional. Trends Genet. **15**:134–135.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

IRISH, V. F., A. MARTINEZ-ARIAS, and M. AKAM. 1989. Spatial regulation of the Antennapedia and Ultrabithorax homeotic genes during *Drosophila* early development. EMBO J. **8**:1527–1537.

JACOBS, J. J., and M. VAN LOHUIZEN. 1999. Cellular memory of transcriptional states by Polycomb-group proteins. Semin. Cell Dev. Biol. **10**:227–235.

JACOBSEN, S. E., M. P. RUNNING, and E. M. MEYEROWITZ. 1999. Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. Development **126**:5231–5243.

JAN, Y. N., and L. Y. JAN. 1993. Functional gene cassettes in development. Proc. Natl. Acad. Sci. USA **90**:8305–8307.

JAREBORG, N., E. BIRNEY, and R. DURBIN. 1999. Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. Genome Res. **9**:815–824.

JEDRUSIK, M. A., and E. SCHULZE. 2001. A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in *Caenorhabditis elegans.* Development **128**:1069–1080.

JIANG, Z. H., and J. Y. WU. 1999. Alternative splicing and programmed cell death. Proc. Soc. Exp. Biol. Med. **220**:64–72.

JOHN, T. R., J. J. SMITH, and I. I. KAISER. 1996. A phospholipase A2-like pseudogene retaining the highly conserved introns of Mojave toxin and other snake venom group II PLA2s, but having different exons. DNA Cell Biol. **15**:661–668.

JONES, A. L., C. L. THOMAS, and A. J. MAULE. 1998. De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. EMBO J. **17**:6385–6393.

JONES, D. O., I. G. COWELL, and P. B. SINGH. 2000. Mammalian chromodomain proteins: their role in genome organisation and expression. Bioessays **22**:124–137.

JONES, L., A. J. HAMILTON, O. VOINNET, C. L. THOMAS, A. J. MAULE, and D. C. BAULCOMBE. 1999. RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. Plant Cell **11**:2291–2302.

JUDD, B. H. 1988. Transvection: allelic cross talk. Cell **53**:841–843.

———. 1995. Mutations of zeste that mediate transvection are recessive enhancers of position-effect variegation in *Drosophila melanogaster*. Genetics **141**:245–253.

KAZMIERCZAK, B., J. BULLERDIEK, K. H. PHAM, S. BARTNITZKE, and H. WIESNER. 1998. Intron 3 of HMGIC is the most frequent target of chromosomal aberrations in human tumors and has been conserved basically for at least 30 million years. Cancer Genet. Cytogenet. **103**:175–177.

KENNERDELL, J. R., and R. W. CARTHEW. 2000. Heritable gene silencing in *Drosophila* using double-stranded RNA. Nat. Biotechnol. **18**:896–898.

KENNISON, J. A. 1995. The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu. Rev. Genet. **29**:289–303.

KETTING, R. F., and R. H. PLASTERK. 2000. A genetic link between co-suppression and RNA interference in *C. elegans*. Nature **404**:296–298.

KIM, Y. G., K. LOWENHAUPT, S. MAAS, A. HERBERT, T. SCHWARTZ, and A. RICH. 2000. The Zab domain of the human RNA editing enzyme ADAR1 recognizes Z-DNA when surrounded by B-DNA. J. Biol. Chem. **275**:26828–26833.

KISS, T., and W. FILIPOWICZ. 1995. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. Genes Dev. **9**:1411–1424.

KLOEK, A. P., J. P. MCCARTER, R. A. SETTERQUIST, T. SCHEDL, and D. E. GOLDBERG. 1996. Caenorhabditis globin genes: rapid intronic divergence contrasts with conservation of silent exonic sites. J. Mol. Evol. **43**:101–108.

KOOP, B. F., and L. HOOD. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. Nat. Genet. **7**:48–53.

KREIVI, J. P., and A. I. LAMOND. 1996. RNA splicing: unexpected spliceosome diversity. Curr. Biol. **6**:802–805.

KYBA, M., and H. W. BROCK. 1998. The *Drosophila* polycomb group protein Psc contacts ph and Pc through specific conserved domains. Mol. Cell. Biol. **18**:2712–2720.

LADOMERY, M. 1997. Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. Bioessays **19**:903–909.

LAMBOWITZ, A. M., and M. BELFORT. 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. **62**:587–622.

LANEY, J. D., and M. D. BIGGIN. 1992. zeste, a nonessential gene, potently activates Ultrabithorax transcription in the *Drosophila* embryo. Genes Dev. **6**:1531–1541.

LANZ, R. B., N. J. MCKENNA, S. A. ONATE, U. ALBRECHT, J. WONG, S. Y. TSAI, M. J. TSAI, and B. W. O'MALLEY. 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell **97**:17–27.

LEE, J. T., L. S. DAVIDOW, and D. WARSHAWSKY. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. **21**:400–404.

LEE, R. C., R. L. FEINBAUM, and V. AMBROS. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75**:843–854.

LEHMANN, A. R. 2001. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. Genes Dev. **15**:15–23.

LIPMAN, D. J. 1997. Making (anti)sense of non-coding sequence conservation. Nucleic Acids Res. **25**:3580–3583.

LIPSHITZ, H. D., D. A. PEATTIE, and D. S. HOGNESS. 1987. Novel transcripts from the Ultrabithorax domain of the bithorax complex. Genes Dev. **1**:307–322.

LLOYD, C., and P. GUNNING. 1993. Noncoding regions of the gamma-actin gene influence the impact of the gene on myoblast morphology. J. Cell Biol. **121**:73–82.

LOEFFLER, M., and G. KROEMER. 2000. The mitochondrion in cell death control: certainties and incognita. Exp. Cell Res. **256**:19–26.

LOGSDON, J. 1998. The recent origins of spliceosomal introns revisited. Curr. Opin. Genet. Dev. **8**:637–648.

LOPEZ, A. J. 1998. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Annu. Rev. Genet. **32**:279–305.

MCADAMS, H. H., and A. ARKIN. 1997. Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA **94**:814–819.

MCADAMS, H. H., and L. SHAPIRO. 1995. Circuit simulation of genetic networks. Science **269**:650–656.

MCCLELLAND, J. L., and D. C. PLAUT. 1993. Computational approaches to cognition: top-down approaches. Curr. Opin. Neurobiol. **3**:209–216.

MCCLELLAND, J. L., and D. E. RUMELHART. 1985. Distributed memory and the representation of general and specific information. J. Exp. Psychol. Gen. **114**:159–197.

MARAHRENS, Y. 1999. X-inactivation by chromosomal pairing events. Genes Dev. **13**:2624–2632.

MARTINEZ-ABARCA, F., and N. TORO. 2000. Group II introns in the bacterial world. Mol. Microbiol. **38**:917–926.

MATSUMOTO, K., and A. P. WOLFFE. 1998. Gene regulation by Y-box proteins: coupling control of transcription and translation. Trends Cell Biol. **8**:318–323.

MATTICK, J. S. 1994. Introns: evolution and function. Curr. Opin. Genet. Dev. **4**:823–831.

MAXWELL, E. S., and M. J. FOURNIER. 1995. The small nucleolar RNAs. Annu. Rev. Biochem. **64**:897–934.

MEISLER, M. H. 1992. Insertional mutation of 'classical' and novel genes in transgenic mice. Trends Genet. **8**:341–344.

MENDOZA, L., and E. R. ALVAREZ-BUYLLA. 1998. Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis. J. Theor. Biol. **193**:307–319.

MESTL, T., E. PLAHTE, and S. W. OMHOLT. 1995. A mathematical framework for describing and analysing gene regulatory networks. J. Theor. Biol. **176**:291–300.

METTE, M. F., W. AUFSATZ, J. VAN DER WINDEN, M. A. MATZKE, and A. J. MATZKE. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. **19**:5194–5201.

MICOL, J. L., J. E. CASTELLI-GAIR, and A. GARCIA-BELLIDO. 1990. Genetic analysis of transvection effects involving cis-regulatory elements of the *Drosophila* Ultrabithorax gene. Genetics **126**:365–373.

MITCHELL, P., E. PETFALSKI, A. SHEVCHENKO, M. MANN, and D. TOLLERVEY. 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple $3' \rightarrow 5'$ exoribonucleases. Cell **91**:457–466.

1628 Mattick and Gagen

MITCHELL, P., and D. TOLLERVEY. 2000. Musing on the structural organization of the exosome complex. Nat. Struct. Biol. **7**:843–846.

MONTGOMERY, M. K., and A. FIRE. 1998. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. Trends Genet. **14**:255–258.

MOREL, J., P. MOURRAIN, C. BECLIN, and H. VAUCHERET. 2000. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis.* Curr. Biol. **10**:1591–1594.

MULLER, M., K. HAGSTROM, H. GYURKOVICS, V. PIRROTTA, and P. SCHEDL. 1999. The mcp element from the *Drosophila melanogaster* bithorax complex mediates long-distance regulatory interactions. Genetics **153**:1333–1356.

NASHIMOTO, M. 2000. Anomalous RNA substrates for mammalian tRNA 3′ processing endoribonuclease. FEBS Lett. **472**:179–186.

NEMES, J. P., K. A. BENZOW, and M. D. KOOB. 2000. The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). Hum. Mol. Genet. **9**:1543–1551.

NEWMAN, A. J. 1994. Pre-mRNA splicing. Curr. Opin. Genet. Dev. **4**:298–304.

NGO, H., C. TSCHUDI, K. GULL, and E. ULLU. 1998. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei.* Proc. Natl. Acad. Sci. USA **95**:14687–14692.

NICOLOSO, M., L. H. QU, B. MICHOT, and J. P. BACHELLERIE. 1996. Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2′-O- ribose methylation of rRNAs. J. Mol. Biol. **260**:178–195.

NIEHRS, C., and N. POLLET. 1999. Synexpression groups in eukaryotes. Nature **402**:483–487.

O'BRIEN, S. P., K. SEIPEL, Q. G. MEDLEY, R. BRONSON, R. SEGAL, and M. STREULI. 2000. Skeletal muscle deformity and neuronal disorder in trio exchange factor-deficient mouse embryos. Proc. Natl. Acad. Sci. USA **97**:12074–12078.

PADGETT, R. A., P. J. GRABOWSKI, M. M. KONARSKA, S. SEILER, and P. A. SHARP. 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.

PAL-BHADRA, M., U. BHADRA, and J. A. BIRCHLER. 1997. Cosuppression in *Drosophila*: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. Cell **90**:479–490.

———. 1999. Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. Cell **99**:35–46.

PALMER, J. D., and J. M. LOGSDON JR. 1991. The recent origins of introns. Curr. Opin. Genet. Dev. **1**:470–477.

PARRISH, S., J. FLEENOR, S. XU, C. MELLO, and A. FIRE. 2000. Functional anatomy of a dsRNA trigger. Differential requirement for the two trigger strands in RNA interference. Mol. Cell **6**:1077–1087.

PASQUINELLI, A. E., B. J. REINHART, F. SLACK et al. (11 coauthors). 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature **408**:86–89.

PAWSON, T. 1995. Protein modules and signalling networks. Nature **373**:573–580.

PELCZAR, P., and W. FILIPOWICZ. 1998. The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5′-terminal oligopyrimidine gene family. Mol. Cell. Biol. **18**:4509–4518.

PELISSIER, T., S. THALMEIR, D. KEMPE, H. L. SANGER, and M. WASSENEGGER. 1999. Heavy de novo methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation. Nucleic Acids Res. **27**:1625–1634.

PELISSIER, T., and M. WASSENEGGER. 2000. A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. RNA **6**:55–65.

PETERS, J., S. F. WROE, C. A. WELLS, H. J. MILLER, D. BODLE, C. V. BEECHEY, C. M. WILLIAMSON, and G. KELSEY. 1999. A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. Proc. Natl. Acad. Sci. USA **96**:3830–3835.

PIRROTTA, V. 1990. Transvection and long-distance gene regulation. Bioessays **12**:409–414.

———. 1999. Transvection and chromosomal trans-interaction effects. Biochim. Biophys. Acta **1424**:M1–M8.

PLASTERK, R. H., and R. F. KETTING. 2000. The silence of the genes. Curr. Opin. Genet. Dev. **10**:562–567.

PLUNKETT, K., A. KARMILOFF-SMITH, E. BATES, J. L. ELMAN, and M. H. JOHNSON. 1997. Connectionism and developmental psychology. J. Child Psychol. Psychiatry **38**:53–80.

POTTER, S. S., and W. W. BRANFORD. 1998. Evolutionary conservation and tissue-specific processing of Hoxa 11 antisense transcripts. Mamm. Genome **9**:799–806.

PRISLEI, S., A. FATICA, E. DE GREGORIO, M. ARESE, P. FRAGAPANE, E. CAFFARELLI, C. PRESUTTI, and I. BOZZONI. 1995. Self-cleaving motifs are found in close proximity to the sites utilized for U16 snoRNA processing. Gene **163**: 221–226.

PRISLEI, S., A. MICHIENZI, C. PRESUTTI, P. FRAGAPANE, and I. BOZZONI. 1993. Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes. Nucleic Acids Res. **21**:5824–5830.

QIAN, L., M. N. VU, M. CARTER, and M. F. WILKINSON. 1992. A spliced intron accumulates as a lariat in the nucleus of T cells. Nucleic Acids Res. **20**:5345–5350.

QU, L. H., A. HENRAS, Y. J. LU, H. ZHOU, W. X. ZHOU, Y. Q. ZHU, J. ZHAO, Y. HENRY, M. CAIZERGUES-FERRER, and J. P. BACHELLERIE. 1999. Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. Mol. Cell. Biol. **19**:1144–1158.

REBANE, A., R. TAMME, M. LAAN, I. PATA, and A. METSPALU. 1998. A novel snoRNA (U73) is encoded within the introns of the human and mouse ribosomal protein S3a genes. Gene **210**:255–263.

REINHART, B. J., F. J. SLACK, M. BASSON, A. E. PASQUINELLI, J. C. BETTINGER, A. E. ROUGVIE, H. R. HORVITZ, and G. RUVKUN. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans.* Nature **403**: 901–906.

REZNIKOFF, W. S. 1992. The lactose operon-controlling elements: a complex paradigm. Mol. Microbiol. **6**:2419–2422.

RIEGER, M., and W. W. FRANKE. 1988. Identification of an orthologous mammalian cytokeratin gene. High degree of intron sequence conservation during evolution of human cytokeratin 10. J. Mol. Biol. **204**:841–856.

ROEST CROLLIUS, H., O. JAILLON, A. BERNOT et al. (12 coauthors). 2000. Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. Nat. Genet. **25**:235–238.

ROSBY, O., P. ALESTROM, and K. BERG. 1997. High-degree sequence conservation in LPA kringle IV-type 2 exons and introns. Clin. Genet. **52**:293–302.

RUBIN, G. M., M. D. YANDELL, J. R. WORTMAN et al. (55 coauthors). 2000. Comparative genomics of the eukaryotes. Science **287**:2204–2215.

RUMELHART, D. E., and J. L. MCCLELLAND. 1986. Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: Foundations. MIT Press, Cambridge, Mass.

RUSKIN, B., and M. R. GREEN. 1985. An RNA processing activity that debranches RNA lariats. Science **229**:135–140.

SANCHEZ-HERRERO, E., and M. AKAM. 1989. Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila.* Development **107**:321–329.

SANTORO, B., E. DE GREGORIO, E. CAFFARELLI, and I. BOZZONI. 1994. RNA-protein interactions in the nuclei of *Xenopus* oocytes: complex formation and processing activity on the regulatory intron of ribosomal protein gene L1. Mol. Cell. Biol. **14**:6975–6982.

SCHUMACHER, A., and T. MAGNUSON. 1997. Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond. Trends Genet. **13**:167–170.

SHARP, P. A. 2001. RNA interference—2001. Genes Dev. **15**: 485–490.

SHARP, P. A., M. M. KONARKSA, P. J. GRABOWSKI, A. I. LAMOND, R. MARCINIAK, and S. R. SEILER. 1987. Splicing of messenger RNA precursors. Cold Spring Harb. Symp. Quant. Biol. **52**:277–285.

SHEARMAN, L. P., S. SRIRAM, D. R. WEAVER et al. (11 coauthors). 2000. Interacting molecular loops in the mammalian circadian clock. Science **288**:1013–1019.

SHI, H., A. DJIKENG, T. MARK, E. WIRTZ, C. TSCHUDI, and E. ULLU. 2000. Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. RNA **6**: 1069–1076.

SHI, Y., and J. M. BERG. 1995. Specific DNA-RNA hybrid binding by zinc finger proteins. Science **268**:282–284.

SHNYREVA, M., D. S. SCHULLERY, H. SUZUKI, Y. HIGAKI, and K. BOMSZTYK. 2000. Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein. J. Biol. Chem. **275**:15498–15503.

SIJEN, T., and J. M. KOOTER. 2000. Post-transcriptional gene-silencing: RNAs on the attack or on the defense? Bioessays **22**:520–531.

SIPOS, L., J. MIHALY, F. KARCH, P. SCHEDL, J. GAUSZ, and H. GYURKOVICS. 1998. Transvection in the Drosophila Abd-B domain: extensive upstream sequences are involved in anchoring distant cis-regulatory regions to the promoter. Genetics **149**:1031–1050.

SIT, T. L., A. A. VAEWHONGS, and S. A. LOMMEL. 1998. RNA-mediated trans-activation of transcription from a viral RNA. Science **281**:829–832.

SLEUTELS, F., D. P. BARLOW, and R. LYLE. 2000. The uniqueness of the imprinting mechanism. Curr. Opin. Genet. Dev. **10**:229–233.

SMARDON, A., J. M. SPOERKE, S. C. STACEY, M. E. KLEIN, N. MACKIN, and E. M. MAINE. 2000. EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans.* Curr. Biol. **10**:169–178.

SMITH, C. M., and J. A. STEITZ. 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. Mol. Cell. Biol. **18**:6897–6909.

SMITH, C. W., and J. VALCARCEL. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci. **25**:381–388.

SMOLEN, P., D. A. BAXTER, and J. H. BYRNE. 1999. Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems. Am. J. Physiol. **277**: C777–790.

———. 2000. Modeling transcriptional control in gene networks—methods, recent results, and future directions. Bull. Math. Biol. **62**:247–292.

SOLLNER-WEBB, B. 1993. Novel intron-encoded small nucleolar RNAs. Cell **75**:403–405.

STARKE, T., and J. P. GOGARTEN. 1993. A conserved intron in the V-ATPase A subunit genes of plants and algae. FEBS Lett. **315**:252–258.

STENOIEN, D., Z. D. SHARP, C. L. SMITH, and M. A. MANCINI. 1998. Functional subnuclear partitioning of transcription factors. J. Cell. Biochem. **70**:213–221.

STOLTZFUS, A. 1999. On the possibility of constructive neutral evolution. J. Mol. Evol. **49**:169–181.

STOLTZFUS, A., D. F. SPENCER, M. ZUKER, J. M. LOGSDON JR., and W. F. DOOLITTLE. 1994. Testing the exon theory of genes: the evidence from protein structure. Science **265**: 202–207.

STRUTT, H., and R. PARO. 1997. The polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes. Mol. Cell. Biol. **17**:6773–6783.

SUN, L., Y. LI, A. K. MCCULLOUGH, T. G. WOOD, R. S. LLOYD, B. ADAMS, J. R. GURNON, and J. L. VAN ETTEN. 2000. Intron conservation in a UV-specific DNA repair gene encoded by chlorella viruses. J. Mol. Evol. **50**:82–92.

SZEBENYI, G., and J. F. FALLON. 1999. Fibroblast growth factors as multifunctional signaling factors. Int. Rev. Cytol. **185**:45–106.

TABARA, H., M. SARKISSIAN, W. G. KELLY, J. FLEENOR, A. GRISHOK, L. TIMMONS, A. FIRE, and C. C. MELLO. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans.* Cell **99**:123–132.

TANAKA, R., H. SATOH, M. MORIYAMA, K. SATOH, Y. MORISHITA, S. YOSHIDA, T. WATANABE, Y. NAKAMURA, and S. MORI. 2000. Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27¼45) of human B-cell lymphoma. Genes Cells **5**:277–287.

TARRIO, R., F. RODRIGUEZ-TRELLES, and F. J. AYALA. 1998. New Drosophila introns originate by duplication. Proc. Natl. Acad. Sci. USA **95**:1658–1662.

TAUTZ, D., M. TRICK, and G. A. DOVER. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature **322**:652–656.

TAVERNARAKIS, N., S. L. WANG, M. DOROVKOV, A. RYAZANOV, and M. DRISCOLL. 2000. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. Nat. Genet. **24**:180–183.

THIEFFRY, D., A. M. HUERTA, E. PEREZ-RUEDA, and J. COLLADO-VIDES. 1998. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli.* Bioessays **20**:433–440.

TILLIB, S., S. PETRUK, Y. SEDKOV, A. KUZIN, M. FUJIOKA, T. GOTO, and A. MAZO. 1999. Trithorax- and Polycomb-group response elements within an Ultrabithorax transcription maintenance unit consist of closely situated but separable sequences. Mol. Cell. Biol. **19**:5189–5202.

TOURNIER-LASSERVE, E., W. F. ODENWALD, J. GARBERN, J. TROJANOWSKI, and R. A. LAZZARINI. 1989. Remarkable intron and exon sequence conservation in human and mouse homeobox Hox 1.3 genes. Mol. Cell. Biol. **9**:2273–2278.

TSAI, J. Y., and L. M. SILVER. 1991. Escape from genomic imprinting at the mouse T-associated maternal effect (Tme) locus. Genetics **129**:1159–1166.

TYCOWSKI, K. T., M. D. SHU, and J. A. STEITZ. 1996. A mammalian gene with introns instead of exons generating stable RNA products. Nature **379**:464–466.

VAN DER GUGTEN, A. A., and H. V. WESTERHOFF. 1997. Internal regulation of a modular system: the different faces of internal control. Biosystems **44**:79–106.

VAN DER VLAG, J., and A. P. OTTE. 1999. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nat. Genet. **23**:474–478.

VAN HOOF, A., P. LENNERTZ, and R. PARKER. 2000. Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. EMBO J. **19**:1357–1365.

VAN HOOF, A., and R. PARKER. 1999. The exosome: a proteasome for RNA? Cell **99**:347–350.

VAN LOHUIZEN, M. 1999. The trithorax-group and polycomb-group chromatin modifiers: implications for disease. Curr. Opin. Genet. Dev. **9**:355–361.

VASSETZKY, Y., A. HAIR, and M. MECHALI. 2000. Rearrangement of chromatin domains during development in Xenopus. Genes Dev. **14**:1541–1552.

VENTER, J. C., M. D. ADAMS, E. W. MYERS et al. (274 coauthors). 2001. The sequence of the human genome. Science **291**:1304–1351.

VOINNET, O., C. LEDERER, and D. C. BAULCOMBE. 2000. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana.* Cell **103**:157–167.

VON NEUMANN, J. 1982. First draft of a report on the EDVAC. Pp. 383–392 *in* B. RANDALL, ed. The origins of digital computers: selected papers. Springer, Berlin.

WASSENEGGER, M. 2000. RNA-directed DNA methylation. Plant Mol. Biol. **43**:203–220.

WASSENEGGER, M., and T. PELISSIER. 1998. A model for RNA-mediated gene silencing in higher plants. Plant Mol. Biol. **37**:349–362.

WENG, G., U. S. BHALLA, and R. IYENGAR. 1999. Complexity in biological signaling systems. Science **284**:92–96.

WIGHTMAN, B., I. HA, and G. RUVKUN. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans.* Cell **75**: 855–862.

WILLIAMSON, B. 1977. DNA insertions and gene structure. Nature **270**:295–297.

WOLF, D. M., and F. H. EECKMAN. 1998. On the relationship between genomic regulatory element organization and gene regulatory dynamics. J. Theor. Biol. **195**:167–186.

WOLF, Y. I., F. A. KONDRASHOV, and E. V. KOONIN. 2000. No footprints of primordial introns in a eukaryotic genome. Trends Genet. **16**:333–334.

WRANA, J. L. 1994. H19, a tumour suppressing RNA? Bioessays **16**:89–90.

WROE, S. F., G. KELSEY, J. A. SKINNER, D. BODLE, S. T. BALL, C. V. BEECHEY, J. PETERS, and C. M. WILLIAMSON. 2000. An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. Proc. Natl. Acad. Sci. USA **97**:3342–3346.

WU, C. T., and M. L. GOLDBERG. 1989. The *Drosophila* zeste gene and transvection. Trends Genet. **5**:189–194.

WU, C. T., and J. R. MORRIS. 1999. Transvection and other homology effects. Curr. Opin. Genet. Dev. **9**:237–246.

WUTZ, A., O. W. SMRZKA, N. SCHWEIFER, K. SCHELLANDER, E. F. WAGNER, and D. P. BARLOW. 1997. Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature **389**:745–749.

XING, Y., C. V. JOHNSON, P. R. DOBNER, and J. B. LAWRENCE. 1993. Higher level organization of individual gene transcription and RNA splicing. Science **259**:1326–1330.

YANG, D., H. LU, and J. W. ERICKSON. 2000. Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in drosophila embryos. Curr. Biol. **10**:1191–1200.

YATSUKI, H., H. WATANABE, M. HATTORI et al. (14 co-authors). 2000. Sequence-based structural features between Kvlqt1 and Tapa1 on mouse chromosome 7F4/F5 corresponding to the Beckwith-Wiedemann syndrome region on human 11p15.5: long-stretches of unusually well conserved intronic sequences of kvlqt1 between mouse and human. DNA Res. **7**:195–206.

YEAN, S. L., G. WUENSCHELL, J. TERMINI, and R. J. LIN. 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature **408**:881–884.

YUH, C. H., H. BOLOURI, and E. H. DAVIDSON. 1998. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science **279**:1896–1902.

ZAMORE, P. D., T. TUSCHL, P. A. SHARP, and D. P. BARTEL. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell **101**:25–33.

ZEITLIN, S., and A. EFSTRATIADIS. 1984. In vivo splicing products of the rabbit beta-globin pre-mRNA. Cell **39**:589–602.

ZHANG, J., and M. A. LAZAR. 2000. The mechanism of action of thyroid hormones. Annu. Rev. Physiol. **62**:439–466.

ZINK, B., Y. ENGSTROM, W. J. GEHRING, and R. PARO. 1991. Direct interaction of the Polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of *Drosophila melanogaster.* EMBO J. **10**:153–162.