Revisiting the Central Dogma in the 21st Century

James A. Shapiro

Department of Biochemistry and Molecular Biology, University of Chicago, Gordon Center for Integrative Science, Chicago, IL, USA

Since the elaboration of the central dogma of molecular biology, our understanding of cell function and genome action has benefited from many radical discoveries. The discoveries relate to interactive multimolecular execution of cell processes, the modular organization of macromolecules and genomes, the hierarchical operation of cellular control regimes, and the realization that genetic change fundamentally results from DNA biochemistry. These discoveries contradict atomistic pre-DNA ideas of genome organization and violate the central dogma at multiple points. In place of the earlier mechanistic understanding of genomics, molecular biology has led us to an informatic perspective on the role of the genome. The informatic viewpoint points towards the development of novel concepts about cellular cognition, molecular representations of physiological states, genome system architecture, and the algorithmic nature of genome expression and genome restructuring in evolution.

Key words: **biological theory; evolutionary theory; genome system architecture; cognition; informatics**

The Irony of Molecular Biology

When the structure of DNA was figured out in 1953, there was a strong belief among the pioneers of the new science of molecular biology that they had uncovered the physicochemical basis of heredity and fundamental life processes.[1] Following discoveries about the process of protein synthesis, the consensus view was most cogently summarized a half-century ago in 1958[2] (and then again in 1970[3]) by Crick's declaration of "the central dogma of molecular biology." The concept was that information basically flows from DNA to RNA to protein, which determines the cellular and organismal phenotype. While it was considered a theoretical possibility that RNA could transfer information to DNA, information transfer from proteins to DNA, RNA, or other proteins was considered outside the dogma and "would shake the whole intellectual basis of molecular biology."[3] This DNA/nucleic acid-centered view is still dominant in virtually all public discussions of biological questions, ranging from the role of heredity in disease to arguments about the process of evolutionary change. Even in the technical literature, there is a widespread assumption that DNA, as the genetic material, determines cell action and that observed deviations from strict genetic determinism must be the result of stochastic processes.

The idea of a "dogma" in science has always struck me as inherently self-contradictory. The scientific method is based upon continual challenges to accepted ideas and the recognition that new information inevitably leads to new conceptual formulations. So it seems appropriate to revisit Crick's dictum and ask how it stands up in the light of ongoing discoveries in molecular biology and genomics. The answer is "not well." The last four decades of biomolecular investigation have brought a wealth of discoveries about the informatics of living systems

Address for correspondence: James A. Shapiro, Department of Biochemistry and Molecular Biology, University of Chicago, Gordon Center for Integrative Science, 929 E. 57th Street, Chicago, IL 60637, USA. Voice: 773-702-1625; fax: 773-947-9345. jsha@uchicago.edu

Natural Genetic Engineering and Natural Genome Editing: Ann. N.Y. Acad. Sci. 1178: 6–28 (2009).
doi: 10.1111/j.1749-6632.2009.04990.x

and made the elegant simplifications of the central dogma untenable. Let us review what some of these discoveries have been and see how they revolutionize our concepts of information processing in living cells. The great irony of molecular biology is that it has led us inexorably from the mechanistic view of life it was believed to confirm to an informatic view that was completely unanticipated by Crick and his fellow scientific pioneers.[1]

Basic Molecular Functions

The molecular analysis of fundamental biochemical processes in living cells has repeatedly produced surprises about unexpected (or even "forbidden") activities. A short (and partial) list of these activities provides many illustrative complications or contradictions of the central dogma.

- **Reverse transcription.** The copying of RNA into DNA was predicted by Temin from his studies of RNA tumor viruses that pass through a latent DNA stage.[4] Crick published his 1970 formulation of the central dogma in response to the announcement by Temin and Mitzutani of the discovery of an RNA-dependent DNA polymerase, now called reverse transcriptase.[5] Thus, information can flow from RNA to DNA. We now know that reverse transcriptase activity is present in both prokaryotic and eukaryotic organisms and fulfills a number of different functions related to the modification or addition of genomic DNA sequences. Genome sequencing has revealed abundant evidence of the importance of reverse transcription in genome evolution.[6–8] Indeed, over one-third of our own genomes comes from DNA copies of RNA.[9]
- **Posttranscriptional RNA processing.** Early in the studies of RNA biogenesis, it became apparent that RNA was modified after it was copied from DNA. In some cases, such as tRNA, the modifications altered the individual nucleotides and also involved its cleavage from precursor transcripts.[10,11] With the advent of recombinant DNA technology, it was discovered that many messenger RNAs encoding proteins are processed from initial transcripts by internal cleavage and splicing of intervening sequences.[12,13] We now recognize that differential splicing is an important aspect of biological regulation and differential expression of genomic information.[14,15] In addition, processes of trans-splicing were found to join pieces of two different transcripts[16,17] and RNA editing could alter the base sequence of transcripts.[18,19] Thus, the information content of RNA molecules has many potential inputs besides the sequence of the DNA template for transcription.
- **Catalytic RNA.** Studies of RNA processing by Altman and Cech revealed that some RNA molecules could undergo structural changes in the absence of proteins.[10,20] These discoveries opened the floodgates on the recognition that RNA molecules can have catalytic processes in many ways analogous to those of proteins. This means that RNA plays a more direct role in determining cellular characteristics than the limited protein-coding role assigned by Crick.
- **Genome-wide (pervasive) transcription.** In a widely cited 1980 article published with Leslie Orgel, Crick applied the central dogma view to discriminate genomic DNA into classes that do and do not encode proteins, labeling the latter as "junk DNA" unable to make a meaningful contribution to cell function.[21] One criterion propounded to distinguish informational DNA is whether it is transcribed into RNA. Employing this criterion, the evidence for functionality of all regions of the genome has recently been extended by a detailed investigation of 1% of the human genome.[22] This

study has indicated that virtually all DNA in the genome, most of which does not encode protein, is transcribed from one or both strands.[23] So the central dogma-based notion that the genome can be functionally discriminated into transcribed (informational, coding) and nontranscribed (junk) regions appears to be invalid. There are other reasons for discounting the notion that only protein-coding DNA contains biologically meaningful information.[24]

- **Posttranslation protein modification.** In the early days of molecular biology, it was expected that the rich structural information in protein sequences was sufficient to determine their functional properties. However, biochemical analysis quickly revealed that proteins were subject to functional modulation via an enormous range of covalent alterations after translation on the ribosomes. These modifications included proteolytic cleavage,[25–27] adenylylation,[28] phosphorylation,[29–32] methylation,[33] acetylation,[34,35] attachment of peptides,[36] addition of sugars and polysaccharides,[37–40] decoration with lipids,[41,42] and cis- and trans-splicing.[43] Thus, like RNA, the information content of protein has many potential inputs other than the sequence code maintained in the DNA. It is significant to note that these protein-catalyzed modifications are critical to cellular signal transduction and regulatory circuits. They clearly fall into one of Crick's excluded catgories.[3]
- **DNA proofreading and repair.** In the early days of molecular biology and the central dogma, the stability of genomic information was assumed to be an inherent property of the DNA molecule and the replication machinery. Studies of mutagenesis have revealed that cells possess several levels of protein-based proofreading and error correction systems that maintain the stability of the genome, which is subject to chemical and physical damage, replication errors, and collapse of the replication complex leading to broken DNA molecules.[44–46] In some cases, these protein systems are also responsible for making specific localized changes in the DNA sequence.[47] Thus, the maintenance of genomic information during the replication loop in the central dogma has protein inputs as well.

Cellular Sensing and Intercellular Communication

A major achievement of molecular biology has been the identification of molecules that cells use to acquire information about their chemical, physical, and biological environment and to keep track of internal processes. Many of the biological indicators include molecules produced by the cells themselves. Recognizing the chemical basis for sensing and communication constitutes a major advance in understanding how cells are able to carry out the appropriate actions needed for survival, reproduction, and multicellular development.

- **Allosteric binding proteins.** One of the key triumphs of early molecular biologists was deciphering how small molecules regulate protein synthesis through interactions with DNA-binding transcription factors.[48] This accomplishment was expanded by the more general theory of allosteric transitions in proteins that bind two or more ligands.[49] Binding of one ligand alters the protein shape and alters the interaction with the second ligand. Through these structural and functional alterations, allosteric proteins serve as microprocessors that can transmit information from one cellular component to another.
- **Riboswitches and ribosensors.** The discovery of catalytic RNA led to a dynamic view of RNA structure and function.[50] Information is contained in three-dimensional structure as well as

one-dimensional nucleotide sequence. One aspect of this dynamic view is the realization that RNA can also bind ligands and behave allosterically. Riboswitches, the RNA molecules that bind small molecule ligands and then interact with nucleic acids or proteins, can intervene at all steps in information transfer between the genome and the rest of the cell.[51]

- **Surface and transmembrane receptors.** The first allosteric proteins and RNAs to be studied operated as soluble molecules in the cytoplasm or (in eukaryotic cells) nucleoplasm. Embedded in cell membranes and attached to the cell surface, molecular biologists have identified a wide variety of receptor proteins for detecting extracellular signals, including those indicating the presence of other cells.[52,53] Either the receptors themselves or associated proteins span the cell membrane(s) and transmit external information to the cytoplasm and other cell compartments, including the genome.[54,55]
- **Surface signals.** Complementary to receptors are molecular signals attached to the cell surface that indicate the presence and status of the cell.[56,57] These signals include proteins, polysaccharides, and lipids, and their presence or precise structure can change depending upon cellular physiology, stress, or differentiation. They interact with cognate receptors on other cells.[58] Thus, a great deal of metabolic, developmental, and historical information can be conveyed from one cell to another.[59] Without this kind of information transfer between cell surfaces, successful multicellular development would not be possible.[60]
- **Intercellular protein transfer.** In some cases, multiprotein surface structures serve as conduits for the transmission of proteins from the cytoplasm of one cell to another[61] (see also papers by Baluska, Heinlein, and Rustom from this symposium). Such molecular injections are basic to interkingdom communication in microbial pathogenesis and symbiosis with multicellular hosts.[62–64]
- **Exported signals.** In addition to cell-attached signaling, there is intercellular communication that occurs by molecular diffusion through the atmosphere or aqueous environments. Molecular classes as diverse as gases,[65,66] amino acids or their derivatives,[67] vitamins,[68] oligopeptides,[69] and larger proteins (often decorated with polysaccharide or lipid attachments) serve as alarm signals, hormones, pheromones, and cytokines to carry information between cells that are not in direct contact. Both prokaryotes and eukaryotes use these signals to regulate genetic exchange, homeostasis, metabolism, differentiation, multicellular defense, and morphogenesis.
- **Internal monitors.** The sensory capabilities of cells are not exclusively dedicated to the external chemical or biological environments. Monitoring internal processes and detecting actual or potential malfunctions are critical for reliable cellular reproduction. Molecular studies have revealed a wide range of functions that provide information about the accuracy of DNA replication,[44–46] protein synthesis,[70] membrane composition,[71] and progress through the cell cycle.[72] Current ideas about aberrations in the control of cellular proliferation in cancer attribute a major role to breakdowns in these internal monitoring processes, which often lead to uncontrolled proliferation and genomic instability.

Cellular Control Regimes

As genetic and molecular analysis of cell and organismal phenotypes progressed in the 1970s and 1980s, it quickly became evident that each character depends as much on the cellular functions that regulate expression of genomic information as on the functions that execute the underlying biochemical processes. It is now

taken for granted that every cell process is subject to a control regime that operates algorithmically to adjust to the changing contingencies of both the external and internal environments. Many features of these control regimes have been identified over the past few decades, but it is important to note that we still lack a comprehensive theory of cellular regulation.

- **Feedback regulation circuits.** The molecular analysis of metabolism and protein synthesis at the cellular and multicellular levels has revealed repeated patterns of positive and negative feedback circuitry that is used to achieve and maintain distinct states necessary for reproduction and development.[73] These patterns occur in the control of all cell processes (e.g., replication, transcription, posttranscriptional processing, translation, posttranslational processing, enzyme activity, RNA and protein turnover, etc.), but it is remarkable that the diversity of the molecular components is compatible with a relatively limited set of formal logical descriptions.
- **Signal transduction networks.** Molecular studies of cell growth and differentiation have shown that information about the response to external or internal signals can be transmitted along multimolecular pathways by processes such as sequential protein modifications.[30] These informational transmission chains are often interconnected, so it is more appropriate to describe and analyze them as signal transduction networks than as separate pathways.
- **Second messengers.** In many signal transduction networks, information is transmitted in the form of a small, freely diffusible molecule in the cytoplasm, such as cAMP (used both in pro- and eukaryotes). These cytoplasmic molecules are called second messengers,[74,75] and they constitute chemical symbols of various conditions. In *Escherichia coli*, for example, elevated levels of cAMP represent an absence of glucose in the external environment.[76]
- **Checkpoints.** An important conceptual advance in understanding emergency responses and regulation of the cell cycle was the concept of a checkpoint, a monitoring system that halts progress through the cell cycle until essential preliminary steps have been completed.[77] Concerning the genome, checkpoints have been identified that monitor DNA integrity, completion of DNA replication, and alignment of chromosomes at metaphase.[72] The same concept can be applied to other complex biological processes, such as cellular differentiation and morphogenesis.
- **Epigenetic regulation.** A major focus of current studies on genomic regulation is the control of chromosome regions by alternative chromatin structures. Since chromatin states do not alter DNA sequence but are heritable over many cell generations, and also because chromatin restructuring plays a critical role in cellular differentiation, this control mode is now included under the rubric "epigenetic."[78,79] Epigenetic processes encompass many phenomena, including parental imprinting and erasure of expression states,[80] higher order regulation of multiple linked genetic loci,[81] restriction of genome expression in differentiation,[82] silencing of mobile genetic elements and nearby genetic loci,[83] chromosome position effects,[84] and X chromosome inactivation in mammals.[85] Biochemical analysis has revealed a large number of protein- and DNA-modifying activities that can reformat chromatin from one state to another, often in response to particular stimuli[86,87] or after nuclear transfer.[88]
- **Regulatory RNAs.** Although regulatory RNA molecules had been known for several decades in bacteria, the realization in the 1990s that certain animal "genes" had RNA rather than protein products stimulated extensive research into the role

that small RNA molecules play in cellular regulation.[89] Frequently, the various regulatory effects are gathered under the label of RNAi (for RNA inhibition), but we beginning to learn about positive as well as negative effects of regulatory RNA molecules.[90] We now know about various classes of micro- (mi-), small inhibitory or silencing (si-), repeat-associated silencing (rasi-), and piwi-associated (pi-) RNA classes that control chromatin structure, transcription and translation through a variety of molecular mechanisms.[91] These regulatory RNAs are produced from larger primary transcripts by multiprotein complexes, and they target DNA or RNA molecules on the basis of nucleotide sequence complementarity. This means that any region of the genome can be targeted for control by regulatory RNAs without the need for sequence-specific DNA binding proteins.

- **Subnuclear localization.** An emerging field in cell regulation studies has developed because advances in light microscopy now make it possible to visualize where specific proteins and nucleic acid sequences localize in the nucleus. The new molecular cytology has revealed intricate spatial and functional organization in the prokaryotic cell and the eukaryotic nucleus.[92,93] Processes, such as replication, transcription, splicing, and DNA repair are seen to occur in distinct specialized subnuclear domains (sometimes called "factories"). This subdivision of the nucleus into different compartments indicates that cells have a previously unknown capacity to position DNA and RNA molecules together with distinct functional complexes.

Composite Organization of Macromolecules

In the early days of molecular biology, the prevailing view was that protein molecules and their corresponding DNA sequences (or "genes") functioned as unique intact entities. Today, this unitary perspective has broken down, and we realize that biological macromolecules are generally composites of separable functional components. The same components may be found in molecules that play very different roles in the life of the organism. This combinatorial modularity leads us to think of biomolecules as being the products of a Lego-like assembly process. Modularity is evident at many levels.

- **Multidomain structure of proteins.** Protein sequence databases and genetic engineering experiments have made it clear that proteins contain discrete functional domains.[94] These domains are characterized by the presence of critical amino acids in key positions that are found repeatedly in many proteins. The domains correspond to different functions, such as DNA binding, ATP hydrolysis, membrane localization, protein dimerization, protein phosphorylation, nuclease activity, etc. A domain may be taken from one protein and added to another without losing its functional specificity. Nowadays, a protein's cellular role is generally assessed by determining its domain structure and then trying to figure out how the individual functions work in combination. In other words, proteins are generally considered systems of separate repeatedly utilized domains. Comparative genomics has led to the view that a major force in protein evolution consists of the accretion and shuffling of domains as organisms diverge.[9]
- **Introns, exons, and splicing.** At about the same time that the domain structure of proteins was becoming evident, the separation of many eukaryotic (and some prokaryotic) coding regions into exons and introns was discovered.[95,96] As noted previously, this discovery meant that primary transcripts were composed of discrete coding elements that had to be spliced together

to form a functional mRNA to direct translation. The splicing process provides opportunities for producing more than one product from a particular genetic locus (alternative splicing) and even for producing products encoded by more than one genetic locus (trans-splicing).

- **Complex nature of genomic coding elements.** The genetic dissection of how the genome encodes proteins revealed an unexpected and still-growing array of separate signals in the DNA that are needed for accurate expression. These signals include promoters and transcription factor binding sites for correctly initiating transcription,[97,98] splice donor and splice acceptor signals for proper splicing,[99,100] ribosome binding sites for initiation of translation,[101] and transcriptional termination signals.[102,103] At each level of expression, these signals provide targets for cellular regulatory regimes to intervene in the reading of genomic coding sequences.
- **Repetitive and other "noncoding" DNA.** In most genomes, there are significant amounts of repetitive and other DNA sequences that do not appear to be involved in coding protein or specific RNA products.[104] This is the part of the genome that Crick and Orgel characterized as "junk DNA."[21] In many eukaryotic genomes, such as our own, the abundance of this "noncoding" DNA exceeds the known coding regions by more than an order of magnitude. A wide range of genetic and biochemical studies show that this "noncoding" DNA contains many types of information essential for proper genome expression, replication, and transmission to progeny cells.[24] Through its abundance and taxonomic specificity, it appears that "noncoding" DNA plays a key role in establishing the functional spatial architecture of the genome. The role of repetitive DNA in the organization of chromatin domains is becoming increasingly apparent.[83,105] The recent discovery of pervasive transcription indicates that cells interpret much of this "noncoding" information through RNA transcripts.[23]

Natural Genetic Engineering

Underlying the central dogma and conventional views of genome evolution was the idea that the genome is a stable structure that changes rarely and accidentally by chemical fluctuations[106] or replication errors. This view has had to change with the realization that maintenance of genome stability is an active cellular function and the discovery of numerous dedicated biochemical systems for restructuring DNA molecules.[107–110] Genetic change is almost always the result of cellular action on the genome. These natural processes are analogous to human genetic engineering, and their activity in genome evolution has been extensively documented.[6–8,111,112]

- **Intercellular DNA transfer.** Molecular genetics began with the study of intercellular DNA transfer in bacteria.[113,114] We now know that all prokaryotes have elaborate transmembrane systems for transferring DNA to other cells (even to higher plants) and many also possess them for taking up DNA from the environment.[115–117] This exogenous genetic information can be incorporated into the genome in the form of "islands" encoding specialized adaptive functions.[118] Eukaryotic cells are also capable of taking up and integrating exogenous DNA, but there has been little study of the molecular mechanisms involved.
- **Homology-dependent and -independent recombination.** For many years, geneticists spoke of legitimate and "illegitimate" recombination. The former was used in genetic mapping studies and exchanged segments in DNA molecules that had extensive homologous sequences. The latter produced rearrangements involving

exchanges between DNA molecules with little or no sequence homology. We now know that living cells contain multiple biochemical systems for joining together DNA molecules in ways that are either homology-dependent or -independent.[110] These systems play a critical role in protecting the cell against DNA breakage.[44] Where there is extensive DNA breakage, nonhomologous recombination generates chromosome rearrangements.[119,120] In addition, homology-dependent recombination plays a key role in sexual reproduction by aligning homologous chromosomes in meiosis.

- **DNA rearrangement modules.** In addition to the general systems that work more or less indiscriminately throughout the genome for repairing broken DNA molecules, cells contain defined DNA segments, or modules, and corresponding proteins that mediate homology-independent recombination between the module and a target site elsewhere in the genome. These modules are called mobile genetic elements or transposons, and they also include site-specific recombination systems.[108,110,121] These modular systems can move a defined DNA segment to a new location or make larger DNA rearrangements that bring outside DNA sequences into new relationships along the genome.[112]
- **Retrotransposition, retrotransduction, and reverse splicing.** In addition to mobile DNA modules, there are at least three classes of genetic elements that move via RNA intermediates, which are reverse transcribed and inserted into the genome.[108,110] These retro-elements include retroviruses and related retrotransposons characterized by long terminal repeats (LTRs), non-LTR retrotransposons, and retrohoming introns. In many higher organisms, retrotransposons are the most common form of repetitive DNA; for example, they account for over 30% of the sequenced human genome.[9] The sequence and mechanism of reverse transcription into DNA and insertion into target sequences are different for each class. These elements not only move through the genome and multiply in numbers as they do so, they can also incorporate other cellular sequences and mobilize them to new locations (retrotransduction[111]). Thus, while DNA modules carry out large-scale DNA rearrangements, retrotransposons carry out smaller-scale changes, such as the mobilization of exons to new locations.[122]
- **Protein engineering by DNA rearrangements and targeted mutagenesis.** In cells ranging from bacteria to trypanosomes to mammalian lymphocytes, there are advantages in being able to generate multiple protein structures from a limited DNA coding repertoire.[123] Depending on the particular cell, altering protein coding can involve targeted mutagenesis,[124] reverse transcription,[125] homologous and site-specific recombination,[126–129] rearrangement of exon segments and insertion of untemplated DNA sequences.[130] In some cases, the control of these DNA alterations is tightly controlled, while other examples have the appearance of occurring stochastically.
- **Genome reorganization in normal life cycles.** In organisms from bacteria and yeast to ciliated protozoa and invertebrates, genome restructuring is a programmed part of the normal life cycle. In many of these examples, DNA restructuring removes parts of the genome and occurs only in cells or nuclei that do not contribute to later generations.[131] In other cases involving vegetative cells, the changes do not result in loss of unique information.[132,133] As in protein engineering, these regularly programmed DNA restructurings involve a variety of biochemical mechanisms, from targeted homologous recombination[132] and

site-specific recombination[134,135] to RNA-guided chromosome breakage and reassembly.[136,137] In all these examples, normal genome restructuring is tightly regulated.

- **Response to stress and other stimuli.** While it is conventional wisdom to assert that genetics changes arise sporadically, there is a growing literature documenting the activation of natural genetic engineering systems in response to various stimuli, many of which represent stress or challenges to reproduction.[107] A few dozen different kinds of stimuli that have been documented to activate natural genetic engineering systems in a wide variety of organisms are listed in Table 1. Because natural genetic engineering represents cellular biochemistry acting on the genome, it should not be surprising to find it responsive to outside signals and cell signal transduction networks, like all other aspects of cell biochemistry. Note that at least two retrotransposons (Ty3 and MMTV) have evolved to respond to host organism pheromone/hormone molecules (Table 1).
- **Targeting.** Another inaccurate assertion of conventional wisdom is the idea that DNA changes must occur randomly throughout the genome. Once again, there is a large and growing literature documenting examples (and sometimes clarifying mechanisms) where particular natural genetic engineering systems show decidedly nonrandom specificities of action[112] (Box 5 of Ref. 138). It is of considerable importance to note that many distinct kinds of intermolecular recognition have evolved to target natural genetic engineering functions: sequence-specific DNA binding by proteins, DNA structure recognition by proteins, protein-protein binding, and complementary RNA-DNA base-pairing (Table 1 of Ref. 112).

What can These Molecular Biology Discoveries Teach Us?

If we recognize that the application of new technologies inevitably leads to conceptual changes in science, then we can ask about the basic lessons to be learned from the kind of molecular discoveries outlined above. The lessons are likely to lead us to a significant reformulation of our basic assumptions about the organization and role of the genome in phenotypic expression, heredity, and evolution. I can identify at least six broad lessons. Doubtless, other important lessons remain to be spelled out.

Lesson 1. There is no unidirectional flow of information from one class of biological molecule to another. If we were to attempt a contemporary figure depicting cellular information transfers analogous to Crick's 1970 scheme, it would have to contain at least a dozen Boolean propositions as illustrated in Figure 1. In this far more complex scheme, it is obvious that many types of molecules participate in information transfer from one molecule to any other. In particular, genomic functions are inherently interactive because isolated DNA is virtually inert (and probably never exists in that state at all in a cellular context). DNA cannot replicate or segregate properly to daughter cells or template synthesis of RNA by itself. That is the reason for proposition #1. This fundamental biochemical reality alone would invalidate the central dogma, even if we did not know about the many specific mechanisms that cells possess to complex, modify, and change the structure and function of DNA.

Lesson 2. Classical atomistic concepts of genome organization are no longer tenable. We cannot any more define a "gene" as a unitary component of the genome or specify a "gene product" as the unique

result of expressing a particular region of the genome. Every element of the genome has multiple components and interacts either directly or indirectly with many other genomic elements as it functions in coding, expression, replication, and inheritance. The importance of chromatin configuration, RNA processing, and protein modification are clear examples of how separate genomic elements influence expression of any individual coding sequence. Similarly, the idea of any cellular or organismal character as being "determined" by a single region of the genome has no logical connection with our knowledge of biogenesis. An electronic circuit provides a useful analogy. We can identify individual circuit components by removing or modifying them, but the output is always from the entire circuit, not an individual component. The most that we can conclude from genetic studies is that a particular segment of the genome contains information important for the correct operation of a corresponding cellular (or multicellular) process. Each process involves multiple molecular components, and one region of the genome may be important for more than one process. Our basic concepts of heredity thus have to reflect the inherently systemic and distributed nature of genome Organization.

Lesson 3. This lesson applies to the molecular basis for specificity and precision. The traditional view, inherited from the period around the end of the 19th century, is of a hardwired "lock and key" kind of interaction.[139] While complementary surfaces are still critical to understanding molecular binding, the postcentral dogma discoveries have taught us about the importance of multivalent and combinatorial determination of specificity.[140–142] Increasingly, we appreciate the mobility and interaction of different submolecular domains and the stepwise recruitment of factors in building up multimolecular cellular machinery for high-precision operations.[143,144] In this regard, biological specificity has a "fuzzy logic" rather than rigidly deterministic character.[145,146] It is of great biological significance that multivalent operations provide the potential for feedback, regulation, and robustness that simple mechanical devices lack.

Lesson 4. Genome change arises as a consequence of natural genetic engineering, not from accidents. Replication errors and DNA damage are subject to cell surveillance and correction. When DNA damage correction does produce novel genetic structures, natural genetic engineering functions, such as mutator polymerases and nonhomologous end-joining complexes, are involved. Realizing that DNA change is a biochemical process means that it is subject to regulation like other cellular activities. Thus, we expect to see genome change occurring in response to different stimuli (Table 1) and operating nonrandomly throughout the genome, guided by various types of intermolecular contacts (Table 1 of Ref. 112). These expectations open up new ways of thinking about the role of natural genetic engineering in normal life cycles and the potential for nonrandom processes in evolution.

Lesson 5. Informatic rather than mechanical processes control cell functions. The prevailing 20th-century conception of living cells arose out of the mechanism-vitalism debates of the 1890s-1920s.[147,148] The cell was often viewed as a complex mechanical device that operated on a large set of independent linear responses to conditions. This dominant mechanistic view began to break down at the end of the 20th century with the discovery of increasingly dense and interconnected

TABLE 1. Responses of Natural Genetic Engineering Functions to Various Stimuli

Signal or condition	Natural genetic engineering function	Organism(s)	Reference
Quorum pheromones	DNA release and competence for DNA uptake	Multiple bacteria	Spoering, A.L. & M.S. Gilmore. 2006. Quorum sensing and DNA release in bacterial biofilms. *Curr. Opin. Microbiol.* **9:** 133–137. Sturme, M.H. *et al.* 2002. Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie Van Leeuwenhoek.* **81:** 233–243. Miller, M.B. & B.L. Bassler. 2001. Quorum sensing in bacteria. *Ann. Rev. Microbiol.* **55:** 165–199.
Chitin	Competence for DNA uptake	*Vibrio cholerae*	Meibom, K.L. *et al.* 2005. Chitin induces natural competence in *Vibrio cholerae*. *Science* **310:** 1824–1827.
Various stress conditions	Competence for DNA uptake	Gram-positive bacteria	Claverys, J.P., M. Prudhomme & B. Martin. 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Ann. Rev. Microbiol.* **60:** 451–475.
DNA damage	Recombination and mutator polymerases (SOS response)	*Escherichia coli*, *Bacillus subtilis*, and other bacteria	Sutton M.D. *et al.* 2000. The SOS response: Recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Ann. Rev. Genet.* **34:** 479–497. Au, N. *et al.* 2005. Genetic composition of the *Bacillus subtilis* SOS system. *J. Bacteriol.* **187:** 7655–7666.
DNA damage	Prophage excision	*E. coli*, *B. subtilis*, and other bacteria	Rokney, A. *et al.* 2008. Host responses influence on the induction of lambda prophage. *Mol. Microbiol.* **68:** 29–36. Goranov, A. *et al.* 2006. Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in *Bacillus subtilis*. *J. Bacteriol.* **188:** 5595–5605.
DNA damage	Horizontal transfer of intgrated conjugative (ICE) elements	Multiple bacteria	Beaber, J.W., B. Hochhut & M.K. Waldor. 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427:** 72–74. Auchtung, J.M. *et al.* 2005. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. USA* **102:** 12554–12559.

Continued

TABLE 1. *Continued*

Signal or condition	Natural genetic engineering function	Organism(s)	Reference
Oxidative stress	SOS responses	Multiple bacteria	Giuliodori, A.M. *et al.* 2007. Review on bacterial stress topics. *Ann. N. Y. Acad. Sci.* **1113:** 95–104.
Antibiotic	Prophage excision	*Staphylococcus aureus*	Goerke, C., J. Koller & C. Wolz. 2006. Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in *Staphylococcus aureus*. *Antimicrob. Agents Chemoth.* **50:** 171–177.
Antibiotic	Mutator polymerase	*E. coli*	Pérez-Capilla, T. *et al.* SOS-independent induction of dinB transcription by beta-lactam-mediated inhibition of cell wall synthesis in *Escherichia coli* 2005 *J. Bacteriol.* **187:** 1515–1518.
Tetracycline	CTnDOT excision and conjugal trransfer	*Bacteroides* sp.	Moon, K. *et al.* Regulation of excision genes of the *Bacteroides* conjugative transposon CTnDOT. *J. Bacteriol.* **187:** 5732–5741.
Quorum pheromones, plant metabolites (opines)	Conjugal transfer	*Agrobacterium tumefaciens*	Fuqua, W.C. & S.C. Winans. 1994. A LuxR-LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. *J. Bacteriol.* **176:** 2796–2806.
Plant phenolics	T-DNA transfer to plant cell	*A. tumefaciens*	Gelvin, S.B. 2006. *Agrobacterium* virulence gene induction. *Methods Mol. Biol.* **343:** 77–84.
Extracyto-plasmic stress	F plasmid transfer	*E. coli*	Lau-Wong, I.C. *et al.* 2007. Activation of the Cpx regulon destabilizes the F plasmid transfer activator, TraJ, via the HslVU protease in *Escherichia coli*. *Mol. Microbiol.* **67:** 516–527.
Heat shock	F plasmid transfer	*E. coli*	Zahrl, D. *et al.* 2007. GroEL plays a central role in stress-induced negative regulation of bacterial conjugation by promoting proteolytic degradation of the activator protein TraJ. *J. Bacteriol.* **189:** 5885–5894.
Growth phase	F plasmid	*E. coli*	Will, W.R., J. Lu & L.S. Frost. 2004. The role of H-NS in silencing F transfer gene expression during entry into stationary phase. *Mol. Microbiol.* **54:** 769–782.
Genome reduction	Stress-induced IS elements	*E. coli*	Pósfai, G. *et al.* 2006. Emergent properties of reduced-genome *Escherichia coli*. *Science* **312:** 1044–1046.

Continued

TABLE 1. *Continued*

Signal or condition	Natural genetic engineering function	Organism(s)	Reference
Sex pheromones	Conjugation agglutinins	*Enterobacter fecaelis*	Clewell, D.B. 2007. Properties of *Enterococcus faecalis* plasmid pAD1, a member of a widely disseminated family of pheromone-responding, conjugative, virulence elements encoding cytolysin. *Plasmid* **58:** 205–227.
Aerobic starvation	Mu prophage activation	*E. coli*	Maenhaut-Michel, G. & J.A. Shapiro. 1994. The roles of starvation and selective substrates in the emergence of araB-lacZ fusion clones. *EMBO J.* **13:** 5229–5239. Lamrani. S. *et al.* 1999. Starvation-induced Mucts62-mediated coding sequence fusion: Roles for ClpXP, Lon, RpoS and Crp. *Molec. Microbiol.* **32:** 327–343.
Aerobic starvation	Tn4652 activation	*Pseudomonas putida*	Hõrak, R. *et al.* The ColR-ColS two-component signal transduction system is involved in regulation of Tn4652 transposition in *Pseudomonas putida* under starvation conditions. *Molec. Microbiol.* **54:** 795–807.
Aerobic starvation	Plasmid DNA amplification and mutagenesis	*E. coli*	Slack, A. *et al.* 2006. On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genetics* **2:** 385–398.
Aerobic starvation	Base substitutions	*E. coli*	Bjedov, I. *et al.* 2003. Stress-induced mutagenesis in bacteria. *Science* 300: 1404–1409.
Aerobic starvation	Tandem duplications and amplifications	*Salmonella enterica*	Kugelberg, E. *et al.* 2006. Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl. Acad. Sci. USA* **103:** 17319–17324.
Heat shock	IS element activation	*Burkholderia* sp.	Taghavi, S., M. Mergeay & D. van der Lelie. 1997. Genetic and physical maps of the *Alcaligenes eutrophus* CH34 megaplasmid pMOL28 and its derivative pMOL50 obtained after temperature-induced mutagenesis and mortality. *Plasmid* **37:** 22–34.
Heat shock, high culture density	IS4Bsu1 element	*B. subtilis*	Takahashi, K. *et al.* 2007. Development of an intermolecular transposition assay system in *Bacillus subtilis* 168 using IS4Bsu1 from *Bacillus subtilis* (natto). *Microbiology* **153:** 2553–2559.

Continued

TABLE 1. *Continued*

Signal or condition	Natural genetic engineering function	Organism(s)	Reference
Adenine starvation	Ty1 retrotransposon activation	*Saccharomyces cerevisiaea*	Todeschini, A.L. *et al.* 2005. Severe adenine starvation activates Ty1 transcription and retrotransposition in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **25:** 7459–7472.
DNA damage (radiation or carcinogen)	Ty1 retrotransposon activation	*S. cerevisiaea*	Bradshaw, V.A. & K. McEntee. 1989. DNA damage activates transcription and transposition of yeast Ty retrotransposons. *MGG Molec. Gen. Genet.* **218:** 465–474.
Telomere erosion	Ty1 retrotransposon activation	*S. cerevisiaea*	Scholes, DT. *et al.* 2003. Activation of a LTR-retrotransposon by telomere erosion. *Proc. Natl. Acad. Sci. USA* **100:** 15736–15741.
Oxidative conditions (H_2O_2)	Tf2 retrotransposon activation	*Schizosaccharomyces pombe*	Cam, H.P. *et al.* 2007. Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature* **451:** 431–436. Sehgal, A., C.Y. Lee & P.J. Espenshade. 2007. SREBP controls oxygen-dependent mobilization of retrotransposons in fission yeast. *PLoS Genet.* **3:** e131.
Mating pheromone	Ty3 retrotransposon activation	*S. cerevisiaea*	Kinsey, P.T. & S.B. Sandmeyer. 1995. Ty3 transposes in mating populations of yeast: A novel transposition assay for Ty3. *Genetics* **139:** 81–94.
DNA damage (Mitomycin C)	Transposon and retrotransposon activation	*Drosophila melanogaster*	Georgiev, P.G. *et al.* 1990. Mitomycin C induces genomic rearrangements involving transposable elements in *Drosophila melanogaster*. *Molec. Gen. Genet.* **220:** 229–233.
DNA damage	Alu retransposition	*Homo sapiens*	Hagan, C.R., R.F. Sheffield & C.M. Rudin. 2003. Human Alu element retrotransposition induced by genotoxic stress. *Nat. Genet.* **35:** 219–220.
Steroid hormones	Mouse mammary tumor virus (MMTV) activation	*Mus musculus*	Truss, M., G. Chalepakis, M. Beato. 1992. Interplay of steroid hormone receptors and transcription factors on the mouse mammary tumor virus promoter. *J. Steroid Biochem. Mol. Biol.* **43:** 365–378.
Plant alarm chemicals	Retrotransposon activation	*Nicotiana tabacum*	Beguiristain, T. *et al.* 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol.* **127:** 212–221.

Continued

TABLE 1. *Continued*

Signal or condition	Natural genetic engineering function	Organism(s)	Reference
Hydrostatic pressure	MITE DNA transposons	rice	Lin, X. *et al.* 2006. In planta mobilization of mPing and its putative autonomous element Pong in rice by hydrostatic pressurization. *J. Exp. Bot.* **57:** 2313–2323.
Cutting or wounding	Retrotransposon activation	*N. tabacum*	Sugimoto, K., S. Takeda & H. Hirochika. 2000. MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. *Plant Cell* **12:** 2511–2527.
Protoplasting & growth in tissue culture	Transposon and retrotransposon activation	various plants	Grandbastien, M.-A. 1998. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3:** 181–187. Hirochika, H. 1993. Activation of tobacco retrotransposons during tissue culture. *EMBO J.* **12:** 2521–2528. Kikuchi, K. *et al.* 2003. The plant MITE mPing is mobilized in anther culture. *Nature* **421:** 167–170.
Protoplasting & growth in tissue culture	Tos17 retrotransposon activation	rice	Hirochika, H. *et al.* 1996. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA* **93:** 7783–7788.
Cell culture growth	1731 LTR retrotransposon	*D. melanogaster*	Maisonhaute C. *et al.* 2007. Amplification of the 1731 LTR retrotransposon in *Drosophila melanogaster* cultured cells: origin of neocopies and impact on the genome. *Gene* **393:** 116–126.
Fungal metabolites	TnT1 retrotransposon	*Nicotiana tabacum*	Melayah, D. *et al.* 2001. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *Plant J.* **28:** 159–168.
Fungal infection	(CT)n microsatellite contraction	wheat	Schmidt, A.L. & V. Mitter. 2004. Microsatellite mutation directed by an external stimulus. *Mut. Res.* **568:** 233–243. Boyko, A. *et al.* Transgenerational changes in the genome stability and methylation in pathogen-infected plants (Virus-induced plant genome instability). *Nucl. Ac. Res.* **35:** 1714–1725.

Continued

TABLE 1. *Continued*

Signal or condition	Natural genetic engineering function	Organism(s)	Reference
Temperature	Amplification & reduction in DNA repeats	*Festuca arundinacea* (Tall Fescue)	Ceccarelli, M. *et al.* 2002. Genome plasticity in *Festuca arundinacea*: Direct response to temperature changes by redundancy modulation of interspersed DNA repeats. *Theoret. Appl. Genet.* **104:** 901–907.
Elevation and moisture	BARE-1 retrotransposition	*Hordeum spontaneum* (wild barley)	Kalendar, R. *et al.* 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* **97:** 6603–6607.
Heat shock, toxic chemicals	SINE transcription	*Bombyx morii* (silkworm)	Kimura, R.H. *et al.* 2001. Stress induction of Bm1 RNA in silkworm larvae: SINEs, an unusual class of stress genes. *Cell Stress Chaper.* **6:** 263–272.
Various stress conditions	SINE transcription	*H. sapiens*	Li, T.-H. & C.W. Schmid. 2001. Differential stress induction of individual Alu loci: Implications for transcription and retrotransposition. *Gene* **276:** 135–141.
Heat shock	B1 SINE transcription	*M. musculus*	Li, T.-H. *et al.* 1999. Physiological stresses increase mouse short interspersed element (SINE) RNA expression *in vivo*. *Gene* **239:** 367–372.
Industrial air pollution	Microsatellite expansion	*M. musculus*	Somers, C.M. *et al.* 2002. Air pollution induces heritable DNA mutations. *Proc. Natl. Acad. Sci. USA* **99:** 15904–15907.
Industrial air pollution	Microsatellite expansion	Herring gulls	Yauk, C.L. *et al.* 2000. Induced minisatellite germline mutations in herring gulls (*Larus argentatus*) living near steel mills. *Mut. Res.* **452:** 211–218.
Chemical mutagens, etoposide	Microsatellite expansion	*M. musculus*	Vilariño-Güell, C., A.G. Smith & Y.E. Dubrova. 2003. Germline mutation induction at mouse repeat DNA loci by chemical mutagens. *Mut. Res.* **526:** 63–73.
Diet (extra folic acid, vitamin B12, choline, and betaine)	IAP retrotransposon at Agouti locus (Avy allele)	*M. musculus*	Waterland, R.A. & R.L. Jirtle. 2003. Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Molec. Cell. Biol.* **23:** 5293–5300.

Typical expression of the Central Dogma of Molecular Biology:

(DNA ==>2X DNA) ==> RNA ==> Protein ==> Phenotype

Contemporary statements of molecular information transfer in cell:

1. DNA + 0 ==> 0
2. DNA + Protein + ncRNA ==> chromatin
3. Chromatin + Protein + ncRNA ==> DNA replication, chromatin maintenance/reconstitution
4. Protein + RNA + lipids + small molecules ==> signal transduction
5. Chromatin + Protein + signals ==> RNA (primary transcript)
6. RNA + Protein + ncRNA ==> RNA (processed transcript)
7. RNA + Protein + ncRNA ==> Protein (primary translation product)
8. Protein + nucleotides + Ac-CoA + SAM + sugars + lipids ==> Processed and decorated protein
9. DNA + Protein ==> New DNA sequence (mutator polymerases)
10. Signals + Chromatin + Protein ==> New DNA structure (DNA rearrangements subject to stimuli; Table 1)
11. RNA + Protein + chromatin ==> New DNA structure (retrotransposition, retroduction, retrohoming)
12. Signals + chromatin + proteins + ncRNA + lipids ==> nuclear/nucleoid localization

SUMMARY: DNA + Protein + ncRNA + signals + other molecules <==> Genome Structure & Phenotype

Figure 1. Some of the molecular types involved in cellular information transfer events written as Boolean propositions. Note that the involvement of numerous signals and protein or RNA processing steps mentioned in the text have been omitted from many of the propositions for clarity.

regulatory circuits controlling the basic operations of metabolism, biogenesis, the cell cycle, damage responses, and multicellular development.[149,150] Genetic studies of virtually any biological process reliably identify regulatory molecules as well as the expected functions needed to carry out the particular process under investigation. A variety of nonlinear modeling approaches are routinely applied to biological circuits (386 hits from querying PubMed with *nonlinear modeling*). These modeling attempts reflect a growing awareness that information processing is a central aspect of all vital functions.

Lesson 6. Signals play a central role in cell operations. Inspection of Figure 1 shows that "signals" are included as molecular components in several of the Boolean propositions. These signals include diverse chemical classes, such as growth factors bound to surface receptors, small molecule pheromones, cytoplasmic second messengers, and chemical modifications on histones bound to DNA. It would actually be possible to add "signals" to all of the statements in Figure 1 because every one of these information transfer processes can be influenced by various signaling events. The use of signals is critical for such basic vital functions as homeostatic regulation, adaptation to changing conditions, cellular differentiation, and multicellular morphogenesis. The presence of unpredictable signals in biological processes generates an inescapable indeterminacy that contradicts the central dogma and other reductionist statements of genetic determinism. Signal-dependent indeterminacy also produces phenotypic differences between genetically identical cells that is fundamentally distinct from the kind of stochastic noise assumed in most studies of individual cell phenotypes.[151,152]

What New Informatic Concepts do We Need to Elaborate in a 21st-Century View of the Genome and Evolution?

Here are suggestions for a few of the novel ideas I believe will prove helpful as we try to rethink the role of information processing in living cells and organisms.

Cellular Cognition and Action on the Genome

If we are to give up the outmoded atomistic vocabulary of 20th-century genetics, we need to develop a new lexicon of terms based on a view of the cell as an active and sentient entity, particularly as it deals with its genome. The emphasis has to be on what the cell does with and to its genome, not what the genome directs the cell to execute. In some ways, the change in thinking reverses the instructional relationship postulated by the central dogma. The two basic ideas here are:

1. Sensing, computation, and decision-making are central features of cellular functions; and
2. The cell is an active agent utilizing and modifying the information stored in its genome.

Internal Symbolic Representations

In its information processing, the cell makes use of transient information about ambient conditions and internal operations. This information is carried by environmental constituents and signals received from other cells and organisms. The cell's receptors and signal transduction networks transform this transient information into various chemical forms (second messengers, modified proteins, lipids, polysaccharides and nucleic acids) that feed into the operation of cell proliferation, checkpoints, and cellular or multicellular developmental programs. These chemical forms act as symbols that allow the cell to form a virtual representation of its functional status and its surroundings. My argument here is that any successful 21st-century description of biological functions will include control models that incorporate cellular decisions based on symbolic representations.

Genome System Architecture

By flexible analogy with electronic information-processing systems, we need to recognize that every genome has a system architecture which makes it possible for cells to access and utilize the information stored there. It has been argued elsewhere that each genome serves as a read-write (RW) memory system on multiple time scales:[24,138]

1. Within the cell cycle by adjustment of DNA binding protein complexes;
2. Over several cell cycles by chromatin reformatting;
3. Over evolutionary time by natural genetic engineering.

As with electronic systems, different system architectures may accomplish similar functions. Thus genomes may differ in their functional architectures from one taxonomic group to another. The idea of genome system architecture facilitating information utilization can be applied to thinking about existing genomes and also to the potential for generating novel genomes in the face of inevitable but unpredictable challenges. In both situations, there have to be algorithmic processes for searching genome space. If we list the tasks these algorithms must facilitate, there turn out to be a striking number of similarities between what is needed for orderly transcription and what is needed for natural genetic engineering.

Algorithms for searching genome space in normal life cycles (transcription):

1. Locate locus in nucleus/nucleoid;
2. Adjust chromatin configuration;
3. Assemble transcription factors;

4. Move locus to proper functional domain ("transcription factory");
5. Execute transcription;
6. Process transcription product.

Algorithms for searching genome space by natural genetic engineering:

1. Express natural genetic engineering function;
2. Choose and locate substrate sequences (donor, target);
3. Move substrates to proper nuclear functional domain for rearrangements (e.g., DS break repair foci[153]);
4. Adjust chromatin configuration;
5. Assemble natural genetic engineering complex;
6. Process DNA substrates (e.g., reverse transcription);
7. Strand joining, replication and sealing to reconstitute full duplex molecules.

The idea that there are algorithmic processes governing transcription is relatively uncontroversial, but there will be resistance to applying the same concept to natural genetic engineering. The problem comes from the pre-DNA philosophical concept of genetic change as a random process. However, from a biochemical perspective, there are no fundamental differences between transcription and DNA restructuring. In fact, we possess counter-examples to randomness in those cases where DNA change has evolved to be a part of the normal life cycle, as in yeast mating-type switching,[132] postzygotic macronuclear development in ciliated protozoa,[154] and immune system development in vertebrates.[130] In those cases, we have even identified some of the molecular mechanisms involved in making the algorithmic searches that ensure reliability in the DNA changes. They include sequence recognition by proteins, small RNA guidance, and coupling of point mutation and DNA breakage to transcription. There are no mechanistic mysteries involved, only the application of the same molecular processes we recognize in all cell operations on the genome.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Judson, H.F. 1979. *The Eighth Day of Creation: Makers of the Revolution in Biology*. Simon and Schuster. New York.
2. Crick, F.H. 1958. On protein synthesis. *Symp. Soc. Exp. Biol.* **12:** 138–163.
3. Crick, F. 1970. Central dogma of molecular biology. *Nature* **227:** 561–563.
4. Temin, H.M. 1964. The participation of DNA in Rous sarcoma virus production. *Virology* **23:** 486–494.
5. Temin, H.M. & S. Mizutani. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226:** 1211–1213.
6. Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238:** 115–134.
7. Brosius, J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* **118:** 99–116.
8. Betrán, E., K. Thornton & M. Long. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **12:** 1854–1859.
9. International Human Genome Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
10. Altman, S. 2000. The road to RNase P. *Nat. Struct. Biol.* **7:** 827–828.
11. Iwata-Reuyl, D. 2008. An embarrassment of riches: The enzymology of RNA modification. *Curr. Opin. Chem. Biol.* **12:** 126–133.
12. Chow, L.T. *et al*. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12:** 1–8.
13. Sharp, P.A. 1994. Split genes and RNA splicing. *Cell* **77:** 805–815.
14. House, A.E. & K.W. Lynch. 2008. Regulation of alternative splicing: More than just the ABCs. *J. Biol. Chem.* **283:** 1217–1221.
15. Tarn, W.Y. 2007. Cellular signals modulate alternative splicing. *J. Biomed. Sci.* **14:** 517–522.
16. Graber, J.H. *et al*. 2007. *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. *RNA* **13:** 1409–1426
17. Denoeud, F. *et al*. 2007. Prominent use of distal 5′ transcription start sites and discovery of a large

number of additional exons in ENCODE regions. *Genome Res.* **17:** 746–759.
18. Maas, S. & A. Rich. 2000. Changing genetic information through RNA editing. *Bioessays* **22:** 790–802.
19. Amariglio, N. & G. Rechavi. 2007. A-to-I RNA editing: A new regulatory mechanism of global gene expression. *Blood Cells Mol. Dis.* **39:** 151–155.
20. Cech, T.R. 1989. RNA as an enzyme. *Biochem. Int.* **18:** 7–14.
21. Orgel, L.E. & Crick, F.H. 1980. Selfish DNA: The ultimate parasite. *Nature* **284:** 604–607.
22. ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.
23. Gingeras, T.R. 2007. Origin of phenotypes: Genes and transcripts. *Genome Res.* **17:** 682–690.
24. Shapiro, J.A. & R. von Sternberg. 2005. Why repetitive DNA is essential for genome function. *Biol. Rev. (Camb.)* **80:** 227–250.
25. Marks, N., A. Suhar & M. Benuck. 1981. Peptide processing in the central nervous system. *Adv. Biochem. Psychopharmacol.* **28:** 49–60.
26. Lazure, C. *et al*. 1983. Proteases and posttranslational processing of prohormones: A review. *Can. J. Biochem. Cell Biol.* **61:** 501–515.
27. Mizushima, S. 1984. Post-translational modification and processing of outer membrane prolipoproteins in Escherichia coli. *Mol. Cell. Biochem.* **60:** 5–15.
28. Stadtman, E.R. 2001. The story of glutamine synthetase regulation. *J. Biol. Chem.* **276:** 44357–44364.
29. Krebs, E.G. 1983. Historical perspectives on protein phosphorylation and a classification system for protein kinases. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **302:** 3–11.
30. Krebs, E.G. & J.D. Graves. 2000. Interactions between protein kinases and proteases in cellular signaling and regulation. *Adv. Enz. Regul.* **40:** 441–470.
31. Igo, M.M., J.M. Slauch & T.J. Silhavy. 1990. Signal transduction in bacteria: Kinases that control gene expression. *New Biol.* **2:** 5–9.
32. Cohen, P. 2002. The origins of protein phosphorylation. *Nat. Cell Biol.* **4:** E127–E130.
33. Lee, D.Y. *et al*. 2005. Role of protein methylation in regulation of transcription. *Endocr. Rev.* **26:** 147–170.
34. Bayle, J.H. & G.R. Crabtree. 1997. Protein acetylation: More than chromatin modification to regulate transcription. *Chem. Biol.* **4:** 885–888.
35. Grunstein, M. 1997. Histone acetylation in chromatin structure and transcription. *Nature* **389:** 349–352.
36. Herrmann, J., L.O. Lerman & A. Lerman. 2007. Ubiquitin and ubiquitin-like proteins in protein regulation. *Circ. Res.* **100:** 1276–1291.
37. Schäffer, C., M. Graninger & P. Messner. 2001. Prokaryotic glycosylation. *Proteomics* **1:** 248–261.
38. Haltiwanger, R.S. 2002. Regulation of signal transduction pathways in development by glycosylation. *Curr. Opin. Struct. Biol.* **12:** 593–598.
39. Pilobello, K.T. & L.K. Mahal. 2007. Deciphering the glycocode: The complexity and analytical challenge of glycomics. *Curr. Opin. Chem. Biol.* **11:** 300–305.
40. Yurist-Doutsch, S. *et al*. 2008. Sweet to the extreme: Protein glycosylation in archaea. *Mol. Microbiol.* **68:** 1079–1084.
41. Marshall, C.J. 1993. Protein prenylation: A mediator of protein-protein interactions. *Science* **259:** 1865–1866.
42. Sinensky, M. 2000. Recent advances in the study of prenylated proteins. *Biochim. Biophys. Acta* **1484:** 93–106.
43. Saleh, L. & F.B. Perler. 2006. Protein splicing in cis and in trans. *Chem. Rec.* **6:** 183–193.
44. Cox MM. 2002. The nonmutagenic repair of broken replication forks via recombination. *Mutat. Res.* **510:** 107–120.
45. Kunkel, T.A. & D.A. Erie. 2005. DNA mismatch repair. *Ann. Rev. Biochem.* **74:** 681–710.
46. Jiricny, J. 2006. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **7:** 335–346.
47. Goodman, M.F. 2002. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Ann. Rev. Biochem.* **71:** 17–50.
48. Jacob, F. & J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3:** 318– 356.
49. Monod, J., J. Wyman & J.P. Changeux. 1965. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12:** 88–118.
50. Strobel, S.A. & J.C. Cochrane. 2007. RNA catalysis: Ribozymes, ribosomes, and riboswitches. *Curr. Opin. Chem. Biol.* **11:** 636–643.
51. Serganov, A. & D.J. Patel. 2007. Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins. *Nat. Rev. Genet.* **8:** 776–790.
52. Pierce, K.L., R.T. Premont & R.J. Lefkowitz. 2002. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **3:** 639–650.
53. Luttrell, L.M. 2006. Transmembrane signaling by G protein-coupled receptors. *Methods Mol. Biol.* **332:** 3–49.
54. Revenu, C. *et al*. 2004. The co-workers of actin filaments: From cell structures to signals. *Nat. Rev. Mol. Cell Biol.* **5:** 635–646.
55. Krolewski J.J. 2005. Cytokine and growth factor receptors in the nucleus: What's up with that? *J. Cell Biochem.* **95:** 478–487.

56. Aricescu, A.R. & E.Y. Jones. 2007. Immunoglobulin superfamily cell adhesion molecules: Zippers and signals. *Curr. Opin. Cell Biol.* **19:** 543–550.
57. Takada, Y., X. Ye & S. Simon. 2007. The integrins. *Genome Biol.* **8:** 215.
58. Yamada, S. & W.J. Nelson. 2007. Synapses: Sites of cell recognition, adhesion, and functional specification. *Ann. Rev. Biochem.* **76:** 267–294.
59. Widelitz, R. 2005. Wnt signaling through canonical and non-canonical pathways: Recent progress. *Growth Factors* **23:** 111–116.
60. Gumbiner, B.M. 2005. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat. Rev. Mol. Cell Biol.* **6:** 622–634.
61. Neijssen, J., B. Pang & J. Neefjes. 2007. Gap junction-mediated intercellular communication in the immune system. *Prog. Biophys. Mol. Biol.* **94:** 207–218.
62. Backert, S. & T.F. Meyer. 2006. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr. Opin. Microbiol.* **9:** 207–217.
63. Cambronne, E.D. & C.R. Roy. 2006. Recognition and delivery of effector proteins into eukaryotic cells by bacterial secretion systems. *Traffic* **7:** 929–939.
64. Stavrinides, J., H.C. McCann & D.S. Guttman. 2008. Host-pathogen interplay and the evolution of bacterial effectors. *Cell Microbiol.* **10:** 285–292.
65. Bogdan, C. 2001. Nitric oxide and the regulation of gene expression. *Trends Cell Biol.* **11:** 66–75.
66. Chen, Y.F., N. Etheridge & G.E. Schaller. 2005. Ethylene signal transduction. *Ann. Bot. (Lond.)* **95:** 901–915.
67. Fuqua, C., M.R. Parsek & E.P. Greenberg. 2001. Regulation of gene expression by cell-to-cell communication: Acyl-homoserine lactone quorum sensing. *Ann. Rev. Genet.* **35:** 439–468.
68. Chambon, P. 1995. The molecular and genetic dissection of the retinoid signaling pathway. *Rec. Prog. Horm. Res.* **50:** 317–332.
69. Lazazzera, B.A. 2001. The intracellular function of extracellular signaling peptides. *Peptides* **22:** 1519–1527.
70. Mühlemann, O. 2008. Recognition of nonsense mRNA: Towards a unified model. *Biochem. Soc. Trans.* **36**(Pt 3): 497–501.
71. Goldstein, J.L., R.A. DeBose-Boyd & M.S. Brown. 2006. Protein sensors for membrane sterols. *Cell* **124:** 35–46.
72. Novák, B., J.C. Sible & J.J. Tyson. 2003. Checkpoints in the cell cycle. Online Encyclopedia of Life Sciences, John Wiley and Sons (http://www3.interscience.wiley.com/emrw/resolve/oid?OID=112356534).
73. Alon, U. 2007. Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8:** 450–461.
74. Ashcroft, S.J. 1997. Intracellular second messengers. *Adv. Exp. Med. Biol.* **426:** 73–80.
75. Krysko, D.V., L. Leybaert, P. Vandenabeele, *et al.* 2005. Gap junctions and the propagation of cell survival and cell death signals. *Apoptosis* **10:** 459–469.
76. Saier, M.H., JR. 1991. A multiplicity of potential carbon catabolite repression mechanisms in prokaryotic and eukaryotic microorganisms. *New Biol.* **3:** 1137–1147.
77. Hartwell, L.H. & T.A. Weinert. 1989. Checkpoints: Controls that ensure the order of cell cycle events. *Science* **246:** 629–634.
78. Jablonka, E. & M.J. Lamb. 2002. The changing concept of epigenetics. *Ann. N. Y. Acad. Sci.* **981:** 82–96.
79. Goldberg, A.D., C.D. Allis & E. Bernstein. 2007. Epigenetics: A landscape takes shape. *Cell* **128:** 635–638.
80. Ooi, S.L. & S. Henikoff. 2007. Germline histone dynamics and epigenetics. *Curr. Opin. Cell Biol.* **19:** 257–265.
81. Van Driel, R., P.F. Fransz & P.J. Verschure. 2003. The eukaryotic genome: A system regulated at different hierarchical levels. *J. Cell Sci.* **116:** 4067–4075.
82. Kiefer, J.C. 2007. Epigenetics in development. *Dev. Dyn.* **236:** 1144–1156.
83. Slotkin, R.K. & R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8:** 272–285.
84. Schotta, G. *et al.* 2003. Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Sem. Cell Dev. Biol.* **14:** 67–75.
85. Wutz, A. & J. Gribnau. 2007. X inactivation Xplained. *Curr. Opin. Genet. Dev.* **17:** 387–393.
86. Goodrich, J. & S. Tweedie. 2002. Remembrance of things past: Chromatin remodeling in plant development. *Ann. Rev. Cell Dev. Biol.* **18:** 707–746.
87. Morgan, H.D. *et al.* 2005. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14:** R47–R58.
88. Hochedlinger, K. & R. Jaenisch. 2006. Nuclear reprogramming and pluripotency. *Nature* **441:** 1061–1067.
89. Mattick, J.S. 2004. RNA regulation: A new genetics? *Nat. Rev. Genet.* **5:** 316–323.
90. Jopling, C.L., K.L. Norman & P. Sarnow. 2006. Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA miR-122. *Cold Spring Harb. Symp. Quant. Biol.* **71:** 369–376.
91. Farazi, T.A., S.A. Juranek & T. Tuschl. 2008. The growing catalog of small RNAs and their association with distinct argonaute/piwi family members. *Development* **135:** 1201–1214.

92. Thanbichler, M. & L. Shapiro. 2008. Getting organized–how bacterial cells move proteins and DNA. *Nat. Rev. Microbiol.* **6:** 28–40.
93. Schneider, R. & R. Grosschedl. 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.* **21:** 3027–3043.
94. Doolittle, R.F. 1995. The multiplicity of domains in proteins. *Ann. Rev. Biochem.* **64:** 287–314.
95. Doel, M.T. 1980. Gene inserts. Introns and exons, a condensed review. *Cell Biol. Int. Rep.* **4:** 433–445.
96. Witkowski, J.A. 1988. The discovery of 'split' genes: A scientific revolution. *Trends Biochem. Sci.* **13:** 110–113.
97. Harbison, C.T. *et al*. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104.
98. Barrera, L.O. & B. Ren. 2006. The transcriptional regulatory code of eukaryotic cells–insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr. Opin. Cell. Biol.* **18:** 291–298.
99. Eperon, L.P., J.P. Estibeiro & I.C. Eperon. 1986. The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. *Nature* **324:** 280–282.
100. Hiller, M. & M. Platzer. 2008. Widespread and subtle: Alternative splicing at short-distance tandem sites. *Trends Genet.* **24:** 246–255.
101. Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234:** 187–208.
102. Henkin, T.M. 1996. Control of transcription termination in prokaryotes. *Ann. Rev. Genet.* **30:** 35–57.
103. Lykke-Andersen, S. & T.H. Jensen. 2007. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89:** 1177–1182.
104. Britten, R.J. & D.E. Kohne. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161:** 529–540.
105. Podgornaya O.I. *et al*. 2003. Structure-specific DNA-binding proteins as the foundation for three-dimensional chromatin organization. *Int. Rev. Cytol.* **224:** 227–296.
106. Watson, J.D. & F.H. Crick. 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171:** 964–967.
107. McClintock, B. 1984. Significance of responses of the genome to challenge. *Science* **226:** 792–801.
108. Shapiro, J.A. 1983. *Mobile Genetic Elements*. Academic Press. New York, NY.
109. Shapiro, J.A. 1992. Natural genetic engineering in evolution. *Genetica* **86:** 99–111.
110. Craig, N.L. *et al*. (Eds.) 2002. *Mobile DNA II*. American Society of Microbiology Press. Washington, DC.
111. Kazazian, H.H., Jr. 2004. Mobile elements: Drivers of genome evolution. *Science* **303:** 1626–1632.
112. Shapiro, J.A. 2005. A 21st century view of evolution: Genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* **345:** 91–100.
113. Avery, O.T., C.M. MacLeod & M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated prom pneumococcus type III. *J. Exp. Med.* **79:** 137–158.
114. Hayes, W. 1968. *The Genetics of Bacteria and Their Viruses*, 2nd ed. Blackwell. London.
115. Christie, P.J. 2004. Type IV secretion: The agrobacterium VirB/D4 and related conjugation systems. *Bioch. Biophys. Acta* **1694:** 219–234.
116. Averhoff, B. & A. Friedrich. 2003. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. *Arch. Microbiol.* **180:** 385–393.
117. Chen, I. & D. Dubnau. 2004. DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* **2:** 241–249.
118. Dobrindt, U. *et al*. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2:** 414–424.
119. McClintock, B. 1987. *Discovery and Characterization of Transposable Elements: The Collected Papers of Barbara McClintock*. Garland. New York.
120. Weterings, E. & D.J. Chen. 2008. The endless tale of non-homologous end-joining. *Cell Res.* **18:** 114–124.
121. Bukhari, A.I., J.A. Shapiro & S.L. Adhya (Eds.) 1977. *DNA Insertion Elements, Plasmids and Episomes*. Cold Spring Harbor Press. Cold Spring Harbor, New York.
122. Moran, J.V., R.J. DeBerardinis & H.H. Kazazian, JR. 1999. Exon shuffling by L1 retrotransposition. *Science* **283:** 1530–1534.
123. Barbour, A.G. & B.I. Restrepo. 2000. Antigenic variation in vector-borne pathogens. *Emerg. Inf. Dis.* **6:** 449–457.
124. Teng, G. & F.N. Papavasiliou. 2007. Immunoglobulin somatic hypermutation. *Ann. Rev. Genet.* **41:** 107–120.
125. Medhekar, B. & J.F. Miller. 2007. Diversity-generating retroelements. *Curr. Opin. Microbiol.* **10:** 388–395.
126. Gyohda, A. *et al*. 2004. Structure and function of the shufflon in plasmid R64. *Adv. Biophys.* **38:** 183–213.
127. Davidsen, T. & T. Tønjum. 2006. Meningococcal genome dynamics. *Nat. Rev. Microbiol.* **4:** 11–22.

128. Taylor, J.E. & G. Rudenko. 2006. Switching trypanosome coats: What's in the wardrobe? *Trends Genet.* **22:** 614–620.
129. Palmer, G.H. & K.A. Brayton. 2007. Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends Parasitol.* **23:** 408–413.
130. Gellert, M. 2002. V(D)J recombination: RAG proteins, repair factors, and regulation. *Ann. Rev. Biochem.* **71:** 101–132.
131. Müller, F. & H. Tobler. 2000. Chromatin diminution in the parasitic nematodes ascaris suum and parascaris univalens. *Int. J. Parasitol.* **30:** 391–399.
132. Haber, J.E. 1998. Mating-type gene switching in *Saccharomyces cerevisiae*. *Ann. Rev. Genet.* **32:** 561–599.
133. Yamada-Inagawa, T., A.J. Klar & J.Z. Dalgaard. 2007. *Schizosaccharomyces pombe* switches mating type by the synthesis-dependent strand-annealing mechanism. *Genetics* **177:** 255–265.
134. Kunkel, B., R. Losick & P. Stragier. 1990. The *Bacillus subtilis* gene for the development transcription factor sigma K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. *Genes Dev.* **4:** 525–535.
135. Golden, J.W. & H.S. Yoon. 2003. Heterocyst development in *Anabaena*. *Curr. Opin. Microbiol.* **6:** 557–563.
136. Garnier, O. *et al*. 2004. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell Biol.* **24:** 7370–7379.
137. Nowacki, M. *et al*. 2008. RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451:** 153–158.
138. Shapiro, J.A. 2006. Genome informatics: The role of DNA in cellular computations. *Biol. Theory* **1:** 288–301.
139. Fischer, E. 1894. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dt. Chem. Ges.* **27:** 2985–2993.
140. Greenspan, N.S. & L.J. Cooper 1995. Complementarity, specificity and the nature of epitopes and paratopes in multivalent interactions. *Immunol. Today* **16:** 226–230.
141. Schengrund, C.L. 2003. "Multivalent" saccharides: Development of new approaches for inhibiting the effects of glycosphingolipid-binding pathogens. *Biochem. Pharmacol.* **65:** 699–707.
142. Ruthenburg, A.J. *et al*. 2007. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8:** 983–994.
143. Echols, H. 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. *Science* **233:** 1050–1056.
144. Ptashne, M. 1986. *A Genetic Switch: Phage lambda and Higher Organisms*, 2nd ed. Cell Press and Blackwell Scientific. Cambridge, MA.
145. Sadegh-Zadeh, K. 1999. Advances in fuzzy theory. *Artif. Intell. Med.* **15:** 309–323.
146. Jamshidi, M. 2003. Tools for intelligent control: Fuzzy controllers, neural networks and genetic algorithms. *Phil. Trans. A: Math. Phys. Eng. Sci.* **361:** 1781–1808.
147. Melzer, S.J. 1904. Vitalism and mechanism in biology and medicine. *Science* **19:** 18–22.
148. Lillie, R.S. 1914 The philosophy of biology: Vitalism versus mechanism. *Science* **40:** 840–846.
149. Gerhart, J. & M. Kirschner. 1997. *Cells, Embryos, and Evolution*. Blackwell Science. Malden, MA.
150. Alberts, B. *et al*. 2002. *Molecular Biology of the Cell*, 4th ed. Garland. New York.
151. McAdams, H.H. & A. Arkin. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15:** 65–69.
152. Blake, W.J. *et al*. 2003. Noise in eukaryotic gene expression. *Nature* **422:** 633–637.
153. Lisby, M. & R. Rothstein. 2005. Localization of checkpoint and repair proteins in eukaryotes. *Biochimie* **87:** 579–589.
154. Prescott, D.M. 2000. Genome gymnastics: Unique modes of DNA evolution and processing in ciliates. *Nat. Rev. Genet.* **1:** 191–198.