

Getting Over the Code Delusion

Steve Talbott

***Editor's Note:** This is the first in a planned set of essays by Mr. Talbott explaining the significance of a revolution in genetics and molecular biology that has only just begun to receive public attention. Although this essay is at times necessarily technical, we trust that our readers will not find it prohibitively so—and we have appended a modest and informal glossary to help smooth the path, beginning on page 26.*

When it emerged a few years ago that humans and chimpanzees shared, by some measures, 98 or 99 percent of their DNA, a good deal of verbal hand-wringing and chest-beating ensued. How could we hold our heads up with high-browed, post-simian dignity when, as the *New Scientist* reported in 2003, “chimps are human”? If the DNA of the two species is nearly the same, and if, as most everyone seemed to believe, DNA is destiny, what remained to make us special?

Such was the fretting on the human side, anyway. To be truthful, the chimps didn’t seem much interested. And their disinterest, it turns out, was far more fitting than our angst.

In 1992, Nobel prize-winning geneticist Walter Gilbert wrote that you and I will one day hold up a CD containing our DNA sequence and say, “Here is a human being; it’s me!” His essay was entitled “A Vision of the Grail.” Today one can only wonder how we became so invested in the almost sacred importance of an abstract and one-dimensional genetic code—a code so thinly connected to the full-fleshed reality of our selves that its entire import could be captured in a skeletal string of four repeating letters, like so:

ATGCGATCTGTGAGCCGAGTCTTAAGTTCAATTGCAATG

It’s true that the code, as it was understood at the height of the genomic era, had some grounding in material reality. Each of the four different letters stands for one of the four nucleotide bases constituting the DNA sequence. And each group of three successive letters (referred to as

Steve Talbott, a New Atlantis contributing editor, is a senior researcher at the Nature Institute and the editor of NetFuture, an online newsletter about science, technology, and human responsibility (netfuture.org). He is the coauthor, most recently, of Beyond Biotechnology: The Barren Promise of Genetic Engineering (University Press of Kentucky, 2008).

a “codon”) potentially represents an amino acid, a constituent of protein. The idea was that the bases in a protein-coding DNA sequence, or gene, led to the synthesis of the corresponding sequence of amino acids in a protein. And proteins, folded into innumerable shapes, play a decisive role in virtually all living processes. By specifying the production of proteins, genes were presumed to be bearers of the blueprint, or master program, or molecular instruction book of our lives. As Richard Dawkins summed up in his 1986 book *The Blind Watchmaker*:

There is a sense, therefore, in which the three-dimensional coiled shape of a protein is determined by the one-dimensional sequence of code symbols in the DNA....The whole translation, from strictly sequential DNA ROM [read-only memory] to precisely invariant three-dimensional protein shape, is a remarkable feat of digital information technology.

Certainly the idea of a master program seemed powerful to those who were enamored of it. In their enthusiasm they heralded one revolutionary gene discovery after another—a gene for cystic fibrosis (from which the string of letters above is excerpted), a gene for cancer, a gene for obesity, a gene for depression, a gene for alcoholism, a gene for sexual preference. Building block by building block, genetics was going to show how a living organism could be constructed from mindless, indifferent matter.

And yet the most striking thing about the genomic revolution is that the revolution never happened. Yes, it’s been an era of the most amazing technical achievement, marked by an overwhelming flood of new data. It’s true that we are gaining, even if largely by trial and error, certain manipulative powers. But our understanding of the integrity and unified functioning of the living cell has, if anything, been more obscured than illumined by the torrent of data. “Many of us in the genetics community,” write Linda and Edward McCabe in *DNA: Promise and Peril* (2008), “sincerely believed that DNA analysis would provide us with a molecular crystal ball that would allow us to know quite accurately the clinical futures of our individual patients.” Unfortunately, as they and many others now acknowledge, the reality did not prove so straightforward.

As minor tokens of the changing consciousness among biologists, one could cite recent articles in the world’s two premier scientific journals, each reflecting upon the 1989 discovery of the “gene for cystic fibrosis.” “The Promise of a Cure: 20 Years and Counting”—so ran the headline in *Science*, followed by this slightly sarcastic gloss: “The discovery of the cystic fibrosis gene brought big hopes for gene-based medicine; although

a lot has been achieved over two decades, the payoff remains just around the corner.” An echo quickly came from *Nature*, without the sarcasm: “One Gene, Twenty Years: When the cystic fibrosis gene was found in 1989, therapy seemed around the corner. Two decades on, biologists still have a long way to go.”

The story has been repeated for one gene after another, which may be why molecular biologist Tom Misteli offered such a startling postscript to the unbounded optimism of the Human Genome Project. “Comparative genome analysis and large-scale mapping of genome features,” he wrote in the journal *Cell*, “shed little light onto the Holy Grail of genome biology, namely the question of how genomes actually work *in vivo*” (that is, in living organisms).

But is this surprising? The human body is not a mere implication of clean logical code in abstract conceptual space, but rather a play of complexly shaped and intricately interacting physical substances and forces. Yet the four genetic letters, in the researcher’s mind, became curiously detached from their material matrix. In many scientific discussions it hardly would have mattered whether the letters of the “Book of Life” represented nucleotide bases or completely different molecular combinations. All that counted were certain logical correspondences between code and protein together with a few bits of regulatory logic, all buttressed by the massive weight of an unsupported assumption: somehow, by neatly executing an immaculate, computer-like DNA logic, the organism would fulfill its destiny as a living creature. The details could be worked out later.

The misdirection in all this badly needs elaborating—a task I hope to advance here. As for the differences between humans and chimpanzees, the only wonder is that so many were so exercised by it. If we had wanted to compare ourselves to chimps, we could have done the obvious and direct and scientifically respectable thing: we could have observed ourselves and chimps, noting the similarities and differences. Not such a strange notion, really—unless one is so transfixed by a code abstracted from human and chimp that one comes to prefer it to the organisms themselves.

I’m not aware of any pundit who, brought back to reality from the realm of code-fixated cerebration, would have been so confused about the genetic comparison as to invite a chimp home for dinner to discuss world politics. If we had been looking to ground our levitated theory in scientific observation, we would have known that the proper response to the code similarity in humans and chimps was: “Well, so much for the central, determining role we’ve been assigning to our genes.”

The central truth arising from genetic research today is that the hope of finding an adequate explanation of life in terms of inanimate, molecular-level machinery was misconceived. Just as we witness the distinctive character of life when we observe the organism as a whole, so, too, we encounter that same living character when we analyze the organism down to the level of molecules and genes. One by one every seemingly reliable and predictable “molecular mechanism” has been caught deviating from its “program” and submitting instead to the fluid life of its larger context. And chief among the deviants is that supposed First Cause, the gene itself. We are progressing into a post-genomic era—the new era of *epigenetics*.

Genomic Perplexities

The term “epigenetics” most commonly refers to heritable changes in gene activity not accounted for by alterations or mutations in the DNA sequence. But in order to understand the important developments now underway in biology, it’s more useful to take “epigenetics” in its broadest sense as “putting the gene in its living context.”

The genetic code was supposed to reassure us that something like a computational machine lay beneath the life of the organism. The fixity, precision, and unambiguous logical relations of the code seemed to guarantee its strictly mechanistic performance in the cell. Yet it is this fixity, this notion of a precisely characterizable march from cause to effect—and, more broadly, from gene to trait—that has lately been dissolving more and more into the fluid, dynamic exchange of living processes. Organisms, it appears, must be understood and explained at least in part from above downward, from context to subcontext, from the general laws or character of their being to the never-fully-independent details. To realize the full significance of the truth so often remarked in the technical literature today—namely, that context matters—is indeed to embark upon a revolutionary adventure. It means reversing one of the most deeply engrained habits within science—the habit of explaining the whole as the result of its parts. If an organic context really does rule its parts in the way molecular biologists are beginning to recognize, then we have to learn to speak about that peculiar form of governance, turning our usual causal explanations upside down.

A number of conundrums have helped to nudge molecular biology toward a more contextualized understanding of the gene. To begin with, the Human Genome Project revised the human gene count downward from 100,000 to 20,000–25,000. What made the figure startling was

the fact that much simpler creatures—for example, a tiny, transparent roundworm—were found to have roughly the same number of genes. More recently, researchers have turned up a pea aphid with 34,600 genes and a water flea with 39,000 genes. Not even the “chimps are human” boosters were ready to set themselves on the same scale with a water flea. The difference in gene counts required some sort of shift in understanding.

A second oddity centered on the fact that, upon “deciphering” the genetic Book of Life, we found that our coding scheme made the vast bulk of it read like nonsense. That is, some 95 or 98 percent of human DNA was useless for making proteins. Most of this “noncoding DNA” was at first dismissed as “junk”—meaningless evolutionary detritus accumulated over the ages. At best, it was viewed as a kind of bag of spare parts, borne by cells from one generation to another for possible employment in future genomic innovations. But that’s an awful lot of junk for a cell to have to lug around, duplicate at every cell division, and otherwise manage on a continuing basis.

Another conundrum—perhaps the most decisive one—has been recognized and wrestled with (or more often just ignored) since the early twentieth century. With few exceptions, every different type of cell in the human body contains the same chromosomes and the same DNA sequence as the original, single-celled zygote. Yet somehow this zygote manages to differentiate into every manner of tissue—liver, skin, muscle, brain, blood, bone, retina, and so on. If genes determine the form and substance of the organism, how is it that such radically different cellular architectures result from the same genes? What directs genes to produce the intricately sculpted and differentiated form of a complex organism, and how can this directing agency be *governed* by the very genes that it directs?

The developmental biologist F. R. Lillie, remarking in 1927 on the contrast between “genes which remain the same throughout the life history” and a developmental process that “never stands still from germ to old age,” asserted that “Those who desire to make genetics the basis of physiology of development will have to explain how an unchanging complex can direct the course of an ordered developmental stream.”

Think for a moment about this ordered developmental stream. When a cell of the body divides, the daughter cells can be thought of as “inheriting” traits from the parent cell. The puzzle about this cellular-level inheritance is that, especially during the main period of an organism’s development, it leads to a dramatic, highly directed differentiation of tissues. For example, embryonic cells on a path leading to heart muscle tissue become progressively more specialized. The changes each step of

the way are “remembered” (that is, inherited)—but what is remembered is caught up within a process of continuous change. During development you cannot say that every cell reproduces “after its own likeness.”

Over successive generations, cells destined to become a particular type lose their ability to be transformed into any other tissue type. And so the path of differentiation leads from *totipotency* (the single-celled zygote is capable of developing into every cell of the body), to *pluripotency* (embryonic stem cells can transform themselves into many, but not all, tissue types during fetal development), to *multipotency* (blood stem cells can yield red cells, white cells, and platelets), to the final, fully differentiated cell of a particular tissue. In tissues where cell division continues further, the inheritance thereafter may take on a much greater constancy, with like giving rise (at least approximately) to like.

Cells of the mature heart and brain, then, have inherited entirely different destinies, but the difference in those destinies was not written in their DNA sequences, which remain identical in both organs. If we were stuck in the “chimp equals human” mindset, we would have to say that the brain is the same as the heart.

From Junk to Living Organism

So what’s going on? These puzzles turn out to be intimately related. As organisms rise on the evolutionary scale, they tend to have more “junk DNA.” Noncoding DNA accounts for some 10 percent of the genome in many one-celled organisms, 75 percent in roundworms, and 98 percent in humans. The ironic suspicion became too obvious to ignore: maybe it’s precisely our “junk” that differentiates us from water fleas. Maybe what counts most is not so much the genes themselves as the way they are regulated and expressed. Noncoding DNA could provide the complex regulatory functions that direct genes toward service of the organism’s needs, including its developmental needs.

That suspicion has now become standard doctrine—though a still much-too-simplistic doctrine if one stops there. For noncoding as well as coding DNA sequences continue unchanged throughout the organism’s entire trajectory of differentiation, from single cell to maturity. Lillie’s point therefore remains: it is hardly possible for an unchanging complex to explain an ordered developmental stream. Constant *things* cannot by themselves explain dynamic *processes*.

We need a more living understanding. It is not only that noncoding DNA is by itself inadequate to regulate genes. What we are finding is

that at the molecular level the organism is so dynamic, so densely woven and multidirectional in its causes and effects, that it cannot be explicated as living process through strictly local investigations. When it begins to appear that, as one European research team puts it, “everything does everything to everything,”¹ the search for “regulatory control” necessarily leads to the unified and irreducible functioning of the cell and organism as a whole—a living, metamorphosing form within which each more or less distinct partial activity finds its proper place.

The Dynamic Chromosome

The usual formula has it that DNA makes RNA and RNA makes protein. The DNA double helix forms a kind of spiraling ladder, with pairs of nucleotide bases constituting the rungs of the ladder: a nucleotide base attached to one siderail (strand) of the ladder bonds with a base attached to the other strand. These two bases are normally complementary, so that an A on one strand pairs only with a T on the other (and vice versa), just as C and G are paired. Because the chemical subunits making up the double helix are asymmetrical and oriented oppositely on the two strands, the strands can be said to “point” in opposite directions.

The enzyme that transcribes DNA into RNA must move along the length of a gene in the proper direction, separating the two strands and using one of them as a template for synthesizing an RNA transcript—a transcript that complements the DNA template in much the same way that one DNA strand complements the other. It is by virtue of this complementation that the code for a protein is passed from DNA to RNA. RNA, however, is commonly single-stranded, unlike DNA. Once formed, much of it passes through the nuclear envelope to the cytoplasm, where it is translated into protein.

Or so the usual story runs—which is more or less correct as far as it goes. But let’s look at some of what else must go on in order to make the story happen.

If you arranged the DNA in a human cell linearly, it would extend for nearly two meters. How do you pack all that DNA into a cell nucleus just five or ten millionths of a meter in diameter? According to the usual comparison, it’s as if you had to pack 24 miles of extremely thin thread into a tennis ball. Moreover, this thread is divided into 46 pieces (individual chromosomes) averaging, in our tennis-ball analogy, over half a mile long. Can it be at all possible not only to pack the chromosomes into the nucleus, but also to keep them from becoming hopelessly entangled?

Obviously it must be possible, however difficult to conceive—and in fact an endlessly varied packing and unpacking is going on all the time. The first thing to realize is that chromosomes do not consist of DNA only. Their actual substance, an intricately woven structure of DNA, RNA, and protein, is referred to as “chromatin.” “Histone proteins,” several of which can bind together in the form of an extremely complex “spool,” are the single most prominent constituent of this chromatin. Every cell contains numerous such spools—there are some 30 million in a typical human cell—and the DNA double helix, after wrapping a couple of times around one of them, extends for a very short stretch and then wraps around another one. The spool with its DNA is referred to as a “nucleosome,” and between 75 and 90 percent of our DNA is wrapped up in nucleosomes.

But that’s just the first level of packing; it accounts for relatively little of the overall condensation of the chromosomes. If you twist a long, double-stranded rope, you will find the rope beginning to coil upon itself, and if you continue to twist, the coils will coil upon themselves, and so on without particular limit, depending on the fineness and length of the rope. Something like this “supercoiling” happens with the chromosome, mediated in part by the nucleosome spools. As a result, the spools, and the DNA along with them, become tightly packed almost beyond comprehension, in a dense, three-dimensional geometry that researchers have yet to visualize in any detail. This highly condensed state, characterizing great stretches of every chromosome, contrasts with other, relatively unconensed stretches known as “open chromatin.”

At any one time—and with the details depending on the tissue type and stage of the organism’s development, among other things—some parts of every chromosome are heavily condensed while others are open. Every overall configuration represents a unique balance between constrained and liberated expression of our total complement of 25,000 genes. This is because the transcription of genes generally requires an open state; genes in condensed chromatin are largely silenced.

The supercoiling has another direct, more localized role in gene expression. Think again of twisting a rope: depending on the direction of your twist, the two strands of the helix will either become more tightly wound around each other or will be loosened and unwound. (This tightening or loosening of the two strands is independent of the overall supercoiling of the rope, which occurs in either case.) And if, taking a double-stranded rope in hand, you insert a pencil between the strands and force it in one direction along the rope, you will find the strands winding ever more tightly ahead of the pencil’s motion and unwinding behind.

In a similar way, RNA polymerase, the enzyme that transcribes DNA into RNA, must separate the strands of the double helix as it moves along a gene sequence. This is much easier if the supercoiling of the chromatin has already loosened the strands, and harder if the strands are tightened. In this way, the variations in supercoiling along the length of a chromosome either encourage or discourage the transcription of particular genes. Moreover, by virtue of its own activity in moving along the DNA and separating the two strands, RNA polymerase (like the pencil) tends to unwind the strands in the chromosomal region behind it, rendering that region, too, more susceptible to gene expression. There are proteins that detect such changes in the torsion (a sort of twisting tension) propagating along chromatin, and they read the changes as “suggestions” about helping to activate nearby genes.

Picture the situation concretely. Every bodily activity or condition presents its own requirements for gene expression. Whether you are running or sleeping, starving or feasting, getting aroused or calming down, suffering a flesh wound or recovering from pneumonia—in all cases the body and its different cells have specific, almost incomprehensibly complex and changing requirements for differentiated expression of thousands of genes. And one thing necessary for achieving this expression in all its fine detail is the properly choreographed performance of the chromosomes.

This performance cannot be captured with an abstract code. Interacting with its surroundings, the chromosome is as much a living actor as any other part of its living environment. Maybe instead of summoning the image of a rope, I should have invoked a snake, coiling, curling, and sliding over a landscape that is itself in continual movement.

The Dynamic Space of the Cell Nucleus

There are many levels at which we discover significant form and organization in chromatin, which one scientist has dubbed “a plastic polymorphic dynamic elastic resilient flexible nucleoprotein complex.”² Each chromosome, for example, is structured by various means and in ever-changing ways into functionally significant chromosome “domains.” We’ve already seen that chromosomes have both condensed and more open regions. The boundaries between these regions are not always well-defined or digitally precise. Simply by residing close to a more compact region, a gene that otherwise would be very actively transcribed might be only intermittently expressed, or even silenced altogether.

Chromosome domains are also established by the torsion communicated more or less freely along bounded segments of the chromosome.

A region characterized by a particular torsion may attract its own distinctive regulatory proteins. The torsion also tends to correlate with the level of compaction of the chromatin fiber, which in turn correlates with many other aspects of gene regulation. And even on an extremely small scale, the twisting or untwisting of the short stretches of DNA between nucleosomes by various proteins is presumed to help drive the folding or unfolding of the local chromatin.

Genes expressed in the same cell type or at the same time, genes sharing common regulatory factors, and genes actively expressed (or mostly inactive) tend to be grouped together. One way such domains could be established is through the binding of the same protein complexes along a region of the chromosome, thereby establishing a common molecular and regulatory environment for the encompassed genes. But such regions are more a matter of fluid tendency than of absolute rule. All this reminds us that gene regulation is defined less by static elements of logic than by the quality and force of various movements and transformations.

So far we've been looking only at the structure of the chromosome itself. But organization at one level of an organism does not make sense except insofar as it reflects organization at other levels. The structured chromosome can fulfill its tasks only by participating in—mirroring and being mirrored by—a structured nucleus sharing the same dynamic character.

Every chromosome occupies a characteristic region of the nucleus—a “chromosome territory” that varies with the tissue type, the stage of the organism's development, and the life cycle of the individual cell. Chromosomes or parts of chromosomes near the center of the nucleus are marked by more intense gene expression, while those near the outer periphery tend to be repressed.

For local regions of a chromosome, this effect of location can be finely tuned to a degree and in ways that currently baffle all attempts at understanding. Spurred by as yet unknown signals and forces, a particular segment of a chromosome will loop out as an open-chromatin “thread” from its primary territory and come together with other looping segments of the same chromosome. This well-aimed movement brings certain genes and regulatory elements together while keeping others apart, and in this way properly coordinated gene expression is brought about. Sometimes the fraternizing genes are separated on their chromosome by tens of millions of nucleotide bases.

Such chromosome movements are now known to bring together genes and regulatory sites on *different* chromosomes as well (“kissing

chromosomes,” as some researchers have called them). This is a considerable feat of precision targeting, considering not only the chromosome-packing problem discussed above, but also the fact that there are billions of nucleotide bases in human chromosomes. Yet such synchronization of position can be decisive for the expression of particular genes.

Looking at all the coordinated looping and dynamic reorganization of chromosomes, a Dutch research team concluded:

Not only active, but also inactive, genomic regions can transiently interact over large distances with many loci in the nuclear space. The data strongly suggest that each DNA segment has its own preferred set of interactions. This implies that it is impossible to predict the long-range interaction partners of a given DNA locus without knowing the characteristics of its neighboring segments and, by extrapolation, the whole chromosome.³

So context indeed matters. Moreover, the relevant organization of the cell nucleus involves much more than the chromosomes themselves. There are so-called “transcription factories” within the nucleus where looping chromosome segments, regulatory proteins, transcribing enzymes (RNA polymerases), and other substances gather together, presumably making for highly efficient and coordinated gene expression.

Other nuclear functions besides transcription also seem to be localized in this way. But all these specialized locales lack rigid or permanent structure, and are typically marked by rapid turnover of molecules. The lack of well-defined structure in these functional locations contrasts with the cell cytoplasm, which is elaborately subdivided by membranes and populated by numerous organelles. The extraordinary “lightness” and fluidity of the nucleus provide an interesting counterbalance to the relative fixity of the DNA sequence.

With so much concerted movement going on—not to mention the coiling and packing and unpacking of chromosomes mentioned earlier—how does the cell keep all those “miles of string in the tennis ball” from getting hopelessly tangled? All we can say currently is that we know some of the players addressing the problem. For example, there are enzymes called “topoisomerases” whose task is to help manage the spatial organization of chromosomes. Demonstrating a spatial insight and dexterity that might amaze those of us who have struggled to sort out tangled masses of thread, these enzymes manage to make just the right local cuts to the strands in order to relieve strain, allow necessary movement of genes or regions of the chromosome, and prevent a hopeless mass of knots.

Some topoisomerases cut just one strand of the double helix, allow it to wind or unwind around the other strand, and then reconnect the severed ends. This alters the supercoiling of the DNA. Other topoisomerases cut both strands, pass a loop of the chromosome through the gap thus created, and then seal the gap again. (Imagine trying this with miles of string crammed into a tennis ball!) I don't think anyone would claim to have the faintest idea how this is actually managed in a meaningful, overall, contextual sense, although great and fruitful efforts are being made to analyze isolated local forces and "mechanisms."

In sum: the chromosome is engaged in a highly effective spatial performance. It is a living, writhing, gesturing expression of its cellular environment, and the significance of its gesturing goes far beyond the negative requirement that it be condensed and kept free of tangles. If the organism is to survive, chromosome movements must be well-shaped responses to sensitively discerned needs; every gene must be expressed or not according to the needs of the larger context. The chromosome, like everything else in the cell, is itself a manifestation of life, not a logic or mechanism explaining life.

Metamorphosis of the Code

Not only is DNA "managed" by the spatial dynamism of the nucleus and the complex structural folding and unfolding of the chromatin matrix, but the DNA sequence itself is subject to continual transformation. It happens, for example, that certain nucleotide bases are subject to "DNA methylation"—the attachment of methyl groups. These small chemical entities are said to "tag" or "mark" the affected bases, a highly significant process that occurs selectively and dynamically throughout the entire genome. Words such as "attach," "tag," and "mark," however, are grossly inadequate, suggesting as they do little more than a kind of binary coding function whereby we can classify every nucleotide base simply according to the presence or absence of a methyl group. What this leaves out is the actual qualitative change resulting from the chemical transaction.

Part of the problem lies in the mechanistic mindset that looks for the mere aggregation of parts, as if the methyl group and nucleotide base were discrete Lego blocks added together. But wherever chemical bonds are formed or broken, there is a transformation of matter. The result is not just an aggregation or mixture of the substances that came together, but something new, with different qualities and a different constellation of forces.

To think of a methylated cytosine (the nucleotide base most commonly affected) as still the same letter “C” that it was before its methylation, but merely tagged with a methyl group, is to miss the full reality of the situation. What we are really looking at is a metamorphosis of millions of letters of the genetic code under the influence of pervasive and poorly understood cellular processes. And the altered balance of forces represented by all those transformed letters plays with countless possible nuances into the surrounding chromatin, reshaping its sculptural qualities and therefore its expressive potentials.

We are now learning about the consequences of these metamorphoses. In the first place, the transformations of structure brought about by methylation can render DNA locations no longer accessible to the protein transcription factors that would otherwise bind to them and activate the associated genes. Secondly, and perhaps more fundamentally, there are many proteins that *do* recognize methylated sites and bind specifically to them, recruiting in turn other proteins that restructure the chromatin—typically condensing it and resulting in gene repression.

It would be difficult to overstate the profound role of DNA methylation in the organism. In humans, distinctive patterns of DNA methylation are associated with Rett syndrome (a form of autism) and various kinds of mental retardation. Stephen Baylin, a geneticist at Johns Hopkins School of Medicine, says that the silencing, via DNA methylation, of tumor suppressor genes is “probably playing a fundamental role in the onset and progression of cancer. Every cancer that’s been examined so far, that I’m aware of, has this (pattern of) methylation.”⁴ In an altogether different vein, researchers have reported that “DNA methylation is dynamically regulated in the adult nervous system” and is a “crucial step” in memory formation.⁵ It also seems to play a key role in tissue differentiation.

Some patterns of DNA methylation are heritable, leading (against all conventional expectation) to a kind of Lamarckian transmission of acquired characteristics. According to geneticist Joseph Nadeau at the Case Western Reserve University School of Medicine, “a remarkable variety of factors including environmental agents, parental behaviors, maternal physiology, xenobiotics, nutritional supplements and others lead to epigenetic changes that can be transmitted to subsequent generations without continued exposure.”⁶

But by no means are all methylation patterns inherited. For the most part they are not, and for good reason. It would hardly do if tissue-specific patterns of methylation—for example, those in the heart, kidney, or brain—were passed along to the zygote, whose undifferentiated condition

is so crucial to its future development. In general, the slate upon which the developmental processes of the adult have been written needs to be wiped clean in order to clear a space for the independent life of the next generation. As part of this slate-cleaning, a restructuring wave of demethylation passes along each chromosome shortly after fertilization of an egg, and is completed by the time of embryonic implantation in the uterus. Immediately following this, a new methylation occurs, shaped by the embryo itself and giving it a fresh epigenetic start. When, in mammals, the stage of embryonic methylation is blocked artificially, the organism quickly dies.

This structuring and restructuring of DNA by the surrounding life processes is fully as central to a developing organism as the code-conforming DNA sequence.

DNA's Many Languages

We have seen that chromosomes are more than the coded sequence of their DNA. But even when we restrict our gaze to the DNA sequence itself, what we find is much more than a presumed logic—much more than a one-dimensional array of codons that map to the amino acids of protein. As one group of researchers summarize the matter: There is a “growing body of evidence that the topology and the physical features of the DNA itself is an important factor in the regulation of transcription.”⁷ I will try to illustrate this briefly.

Each nucleosome spool is enwrapped by a couple of turns of the DNA double helix. This takes some doing, since the double helix has a certain “stiffness,” or resistance to bending. Some combinations of nucleotide bases lend themselves more easily to bending than others. These combinations influence where nucleosomes will be positioned along the double-stranded DNA. And the positioning of nucleosomes matters at a highly refined level: a shift in position of as little as two or three base pairs can make the difference between an expressed or a silenced gene.

Further, not only the exact position of a nucleosome on the double helix but also the precise *rotation* of the helix on the nucleosome is important. “Rotation” refers to which part of the DNA faces toward the surface of the spool and which part faces outward. Depending on this orientation, the nucleotide bases will be more or less accessible to the various activating and repressing factors that recognize and bind to specific sequences. The orientation in turn depends considerably on the configuration of the local sequence of bases. All of which is to say that among the important

meanings of the language spoken by the genetic sequence are those relating to the distribution of forces in the double helix, which must appropriately complement the forces in nucleosome spools (which latter, as we will see below, can also express themselves with endless variation).

The shape of a stretch of DNA matters in a different way as well. There are two grooves (the major and minor grooves) running the length of a DNA strand, and proteins that recognize an exact sequence of nucleotide bases typically do so in the major groove. However, many proteins bind to DNA in highly selective ways that are not determined by an exact sequence. Recent work has shown that the minor groove may be compressed so as to enhance the local negative electrostatic potential. Regulatory proteins “read” the compression and the electrostatic potential as cues for binding to the DNA. The “complex minor-groove landscape,” as one research team explained in *Nature*,⁸ is indeed affected by the DNA sequence, as well as by associated proteins; however, regulatory factors “reading” the landscape can hardly do so according to a strict digital code. By musical analogy: it’s less a matter of identifying a precise series of notes than of recognizing a melodic motif.

This discovery of the role of the minor groove also helps to solve a puzzle. “The ability to sense the variation in electrostatic potential in DNA,” according to bioinformatics researcher Tom Tullius, “may reveal how a protein could home in on its binding site in the genome without touching every nucleotide”—of which there are billions in every set of human chromosomes. The lesson in all this, Tullius suggests, has to do with what we lose when we simplify DNA to “a one-dimensional string of letters.” After all, “DNA is a molecule with a three-dimensional shape that is not perfectly uniform.”⁹ It is remarkable how readily the historical shift from direct observation of organisms to instrumental readouts of molecular-level processes encouraged a forgetfulness of material form and substance in favor of abstract codes fit for computers.

Meaningful Form

Distinct combinations of nucleotide bases not only assume different conformations themselves; by virtue of their structure, or pattern of forces, they can also impart different conformations to the proteins that bind to them—and these differences can matter a great deal. A group of California molecular biologists recently investigated the glucocorticoid receptor, one of many transcription factors that respond to hormones. Noting the general fact that “genes are not simply turned on or off, but

instead their expression is fine-tuned to meet the needs of a cell,” the researchers went on to report that the various DNA binding sequences for the glucocorticoid receptor may differ by as little as a single base pair. The receptor alters its conformation in response to such differences, and in this way its regulatory activity is modulated.¹⁰

Meanwhile, a Berlin research team looked at several different hormone-responding transcription factors. They concluded that not only did the DNA sequences to which these proteins were bound impart conformational changes to the proteins, but also that these changes led to selective recruitment of different co-regulators and perhaps even to distinct restructurings of the local chromatin architecture. The researchers refer to the “subtle information” conveyed by “unique differences” in the DNA sequence, and the consequent “fine-tuning” of the interplay among regulatory factors. “Small variations in DNA sites,” they write, “can thereby provide for high regulatory diversity, thus adding another level of complexity to gene-specific control.”¹¹

The influence of form works in the other direction as well: the bound protein can transform the shape of DNA in a decisive way, making it easier for a second protein to bind nearby, even without any direct protein-protein interactions. In the case of one gene relating to the production of interferon (an important constituent of the immune system), “eight proteins modulate [DNA] binding site conformation and thereby stabilize cooperative assembly without significant contribution from interprotein interactions.”¹² As a result of this intricate cooperation of proteins and DNA, mediated by the shifting structure of the double helix, the cell achieves proper expression of the interferon gene according to its needs.

On yet another front: the genetic code consists of sixty-four distinct codons, representing all the unique ways four different letters can be arranged in three-letter sequences. Because there are only twenty amino acid constituents of human protein, the code is redundant: several different codons can signify the same amino acid. Such codons have been considered “synonymous,” since the meaning of the code was thought to be exhausted in the specification of amino acids. However, biologists are now in the process of discovering how non-equivalent these synonymous codons really are.

“Synonymous mutations [that is, changes of codons into different, yet synonymous, forms] do not alter the encoded protein, but they can influence gene expression,” Joshua Plotkin and his colleagues write in *Science* magazine. To demonstrate the situation, these scientists engineered 154 versions of a gene—versions that differed randomly from each other, but only in synonymous ways, so that all 154 genes still coded for the same protein.

They found that, in the bacterium *Escherichia coli*, these genes differed in the extent of their expression, with the highest-expressing form producing 250 times as much protein as the lowest-expressing form. Bacterial growth rates also varied. The researchers determined that the choice of synonymous codons affected the folding structure of the resulting RNA transcripts, and this structure then affected the rate of RNA translation into protein.¹³

In short, “synonymous” in the narrow terms of code does not mean “synonymous” as far as the molecular sinews of life are concerned.

Finally and most generally: scientists using computers to scan the several billion nucleotide bases of the human genome in the search for significant features have more and more been using sequence variations as indicators of sculptural and dynamic form at different scales—scales ranging from a few to millions of base pairs. Scans focusing on the DNA sequence alone, abstracted from physical form, have failed to find many of the regulatory elements that now appear so crucial to our understanding of genomic functioning. This search is leading to rapid discovery of new functional aspects of the formerly one-dimensional genome.

The search is also producing a growing awareness that what we inherit (and what makes a difference in evolutionary terms) is as much a matter of three-dimensional structure as it is of nucleotide sequence. Researchers have wondered why the sequences of many functional elements in DNA are not kept more or less constant by natural selection. The standard doctrine has it that functionally important sequences, precisely because they are important to the organism, will generally be conserved across considerable evolutionary distances.

But the emerging point of view holds that architecture can matter as much as sequence. As bioinformatics researcher Elliott Margulies and his team at the National Human Genome Research Institute put it, “the molecular shape of DNA is under selection”—a shape that can be maintained in its decisive aspects despite changes in the underlying sequence. It’s not enough, they write, to analyze “the order of A’s, C’s, G’s, and T’s,” because “DNA is a molecule with a three-dimensional structure.”¹⁴ Elementary as the point may seem, it’s leading to a considerable reallocation of investigative resources.

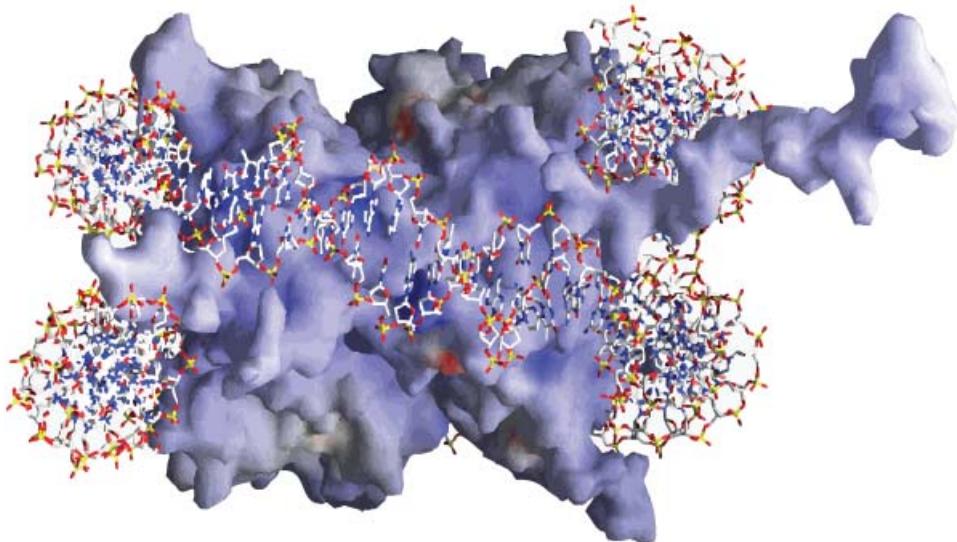
Of course, researchers knew all along that DNA and chromatin were spatial structures. But that didn’t prevent them from ignoring that fact as far as possible. Opportunities to pursue the abstract and determinate lawfulness of a code or mathematical rule have always shown great potential for derailing the scientist’s attention from the world’s full-bodied presentation of itself. Achieving logical and mathematical certainty within

a limited sphere can seem more rigorously scientific than giving attention to the metamorphoses of form and rhythms of movement so intimately associated with life. These latter require a more aesthetically informed approach, and they put us at greater risk of having to acknowledge the evident expressive and highly concerted organization of living processes. When you encounter the meaningful, directed, and well-shaped movements of a dance, it's hard to ignore the active principle—some would say the agency or *being*—coordinating the movements.

And nowhere do we find the dance more evident than in the focal performance of the nucleosome.

The Sensitive Nucleosome

We have spoken of nucleosomes as spools around which DNA is wrapped, but they are not at all like the smooth cylinders that sewing thread is wound on; the image of an irregularly shaped pine cone might be more appropriate (see illustration). Hundreds of distinct points of contact, with countless possible variations, define the relationship between the histone proteins of the spool and the approximately two turns of DNA wrapped around them. As previously noted, this relationship affects access to the enwrapped DNA by the transcription factors that bind to it and promote or repress gene expression. One aspect of the dynamic has to do with the electrical forces that come into play between the (for the most part)



A nucleosome spool with DNA wrapped around it. Illustration courtesy of Karolin Luger, Colorado State University.

positively charged histone surfaces and the negatively charged outer regions of the double helix.

Here it is well to remember one of the primary lessons of twentieth-century physics: we are led disastrously astray when we try to imagine atomic- and molecular-level entities as if they were tiny bits of the stuff of our common experience. The histone spool of nucleosomes, for example, is not some rigid thing. It would be far better to think of its “substance,” “surface,” “contact points,” and “physical interactions” as forms assumed by mutually interpenetrating forces in intricate and varied play.

In any case, the impressive enactments of form and force about the nucleosome are surely central to any understanding of genes. The nucleosome is rather like a maestro directing the genetic orchestra, except that the direction is itself orchestrated by the surrounding cellular audience in conversation with the instrumentalists.

The canonical nucleosome spool is a complex of histone proteins, each of which has a flexible, filamentary “tail” (not shown in the illustration). This tail can be modified through the addition of several different chemical groups—acetyl, methyl, phosphate, ubiquitin, and so on—at any of many different locations along its length. A great variety of enzymes can apply and remove these chemical groups, and the groups themselves play a role in attracting a stunning array of gene regulatory proteins that restructure chromatin or otherwise help choreograph gene expression.

After a few histone tail modifications were found to be rather distinctly associated with active or repressed genes, the forlorn hope arose that we would discover a precise, combinatorial “histone code.” It would provide a kind of fixed, digital key enabling us to predict the consequences of any arrangement of modifications.

But this was to ignore the nearly infinite variety of all those other factors that blend their voices in concert with the histone modifications. In the plastic organism, what goes on at the local level is shaped and guided by a larger, coherent context. As Shelley Berger of Philadelphia’s Wistar Institute observes:

Although [histone] modifications were initially thought to be a simple code, a more likely model is of a sophisticated, nuanced chromatin “language” in which different combinations of basic building blocks yield dynamic functional outcomes.¹⁵

And (leaving aside the jarring reference to building blocks) how could it be otherwise? Each histone tail modification reshapes the physical and electrical structure of the local chromatin, shifting the pattern of

interactions among nucleosome, DNA, and associated protein factors. To picture this situation concretely is immediately to realize that it cannot be captured in purely digital terms. A sculptor does not try to assess the results of a stroke of the hammer as a choice among the possibilities of a digital logic. Berger envisions histone modifications as participating in “an intricate ‘dance’ of associations.”

There is much more. The histones making up a nucleosome spool can themselves be exchanged for noncanonical, or variant, histones, which also have recognizable—but not strictly encoded—effects upon the expression of genes. Histones can even be removed from a spool altogether, leaving it “incomplete.” And certain proteins can slide spools along the DNA, changing their position. As we have seen already, a shift of position by as little as two or three base pairs can make the difference between gene activation or repression, as can changes in the rotational orientation of the DNA on the face of the histone spool. And the tails—no doubt depending at least in part on the various modifications and protein associations mentioned earlier—can thread themselves through the encircling double helix, perhaps either loosening it from the spool or holding it more firmly in place. But those same tails are also thought to establish nucleosome-to-nucleosome contacts, helping to compact a stretch of chromatin and repress gene expression.

Everything depends on contextual configurations that we can reasonably assume are as nuanced and expressively manifold as the gestural configurations available to a stage actor. Further, the nucleosome positioning pattern and other dynamics vary throughout a genome depending on tissue type, stage of the cell life cycle, and the wider physiological environment. They vary between genes that are more or less continuously expressed and those whose expression level changes with environmental conditions. They vary between open chromatin and gene-repressive condensed chromatin. And they vary for any one gene as the actual process of transcription takes place—this because appropriate DNA regulatory sequences must become nucleosome-free before transcription can start, and also because DNA in the body of the gene must be disengaged from nucleosome spools as the transcribing enzyme passes along, only to be (often) re-engaged behind the enzyme.

The Nucleosome as Mediator

Seemingly in the grip of the encircling DNA with its relatively fixed and stable structure, yet responsive to the varying flow of life around

it, the nucleosome holds the balance between gene and context—a task requiring flexibility, a “sense” of appropriate rhythm, and perhaps we could even say “grace.”

Nucleosomes will sometimes move—or be moved (the distinction between actor and acted upon is obscured in the living cell)—rhythmically back and forth between alternative positions in order to enable multiple transcription passes over a gene. In stem cells, a process some have called “histone modification pulsing” results in the continual application and removal of both gene-repressive and gene-activating modifications of nucleosomes. In this way, a delicate balance is maintained around genes involved in development and cell differentiation. The genes are kept, so to speak, in a finely calculated state of “suspended readiness,” so that when the decision to specialize is finally taken, the repressive modifications can be quickly lifted, leading to rapid gene expression.

But quite apart from their role in stem cells, it is increasingly appreciated that nucleosomes play a key role in holding a balance between the active and repressed states of many genes. As the focus of a highly dynamic conversation involving histone variants, histone tail modifications, and innumerable chromatin-associating proteins, decisively placed nucleosomes can (as biologist Bradley Cairns writes) maintain genes “poised in the repressed state,” and “it is the precise nature of the poised state that sets the requirements for the transition to the active state.” Among other aspects of the dynamism, there is continual turnover of the nucleosomes themselves—a turnover that allows transcription factors to gain access to DNA sequences “at a tuned rate.”¹⁶

With another sort of rhythm the DNA around a nucleosome spool “breathes,” alternately pulling away from the spool and then reuniting with it, especially near the points of entry and exit. This provides what are presumably well-gauged, fractional-second opportunities for gene-regulating proteins to bind to their target DNA sequences during the periods of relaxation.

During the actual process of transcription, RNA polymerase appears to take advantage of this “breathing” in order to move, step by step and with significant pauses, along the gene it is transcribing. The characteristics of nucleosomes—whether firmly anchored to the DNA or easily dislodged—affect the timing and frequency of these pauses. And the rhythm of pauses and movements in turn affects the folding of the RNA being synthesized: a proper music is required for correct folding, which finally in its turn affects the structure and function of the protein produced from the RNA molecule.

Such, then, is the sort of intimate, intricate, well-timed choreography through which our genes come to their proper expression. And the plastic, shape-shifting nucleosome in the middle of it all—with its exquisite sensitivity to the DNA sequence on the one hand, and, on the other hand, its mobile tails responding fluidly to the ever-varying signals coming from the surrounding life context—provides an excellent vantage point from which to view the overall drama of form and movement.

Facing Up to Life

The topics covered in this essay represent just a small sample of the findings of genetic and epigenetic research, and we can be sure that, as the field develops, more discoveries will be made that will continue to undermine the doctrine that a genetic code defines the “program of life.” But this is enough, I hope, to suggest why researchers are so energized and excited today. A sense of profound change seems to be widespread.

Meanwhile, the epigenetic revolution is slowly but surely making its way into the popular media—witness the recent *Time* magazine cover story, “Why DNA Isn’t Your Destiny.” The shame of it is that most of the significance of the current research is still being missed. Judging from much that is being written, one might think the main thing is simply that we’re gaining new, more complex insights into how to treat the living organism as a manipulable machine.

The one decisive lesson I think we can draw from the work in molecular genetics over the past couple of decades is that life does not progressively contract into a code or any kind of reduced “building block” as we probe its more minute dimensions. Trying to define the chromatin complex, according to geneticists Shiv Grewal and Sarah Elgin, “is like trying to define life itself.” Having plunged headlong toward the micro and molecular in their drive to reduce the living to the inanimate, biologists now find unapologetic life staring back at them from every chromatogram, every electron micrograph, every gene expression profile. Things do not become simpler, less organic, less animate. The explanatory task at the bottom is essentially the same as the one higher up. It’s rather our *understanding* that all too easily becomes constricted as we move downward, because the contextual scope and qualitative richness of our survey is so extremely narrowed.

The search for precise explanatory mechanisms and codes leads us along a path of least resistance toward the reduction of understanding. A capacity for imagination (not something many scientists are trained for today) is always required for grasping a context in meaningful terms,

because at the contextual level the basic data are not things, but rather relations, movement, and transformation. To see the context is to see a dance, not merely the bodies of the individual dancers.

The hopeful thing is that molecular biologists today—slowly but surely, and perhaps despite themselves—are increasingly being driven to enlarge their understanding through a reckoning with genetic contexts. As a result, they are writing “finis” to the misbegotten hope for a non-lifelike foundation of life, even if the fact hasn’t yet been widely announced.

It is, I think, time for the announcement.

There is a frequently retold story about a little old lady who claims, after hearing a scientific lecture, that the world is a flat plate resting on the back of a giant tortoise. When asked what the turtle is standing on, she invokes a second turtle. And when the inevitable follow-up question comes, she replies, “You’re very clever, young man, but you can’t fool me. It’s turtles all the way down.”

As a metaphor for the scientific understanding of biology, the story is marvelously truthful. In the study of organisms, “It’s life all the way down.”

NOTES

1. Jacques E. Dumont, Frédéric Pécasse, and Carine Maenhaut, “Crosstalk and Specificity in Signalling: Are We Crosstalking Ourselves into General Confusion?,” *Cellular Signalling* 13 (2001): 457-63.
 2. Christophe Lavelle, “Forces and Torques in the Nucleus: Chromatin under Mechanical Constraints,” *Biochemistry and Cell Biology* 87 (2009): 307-22.
 3. Marieke Simonis, Petra Klous, Erik Splinter, *et al.*, “Nuclear Organization of Active and Inactive Chromatin Domains Uncovered by Chromosome Conformation Capture-on-chip (4C),” *Nature Genetics* 38, no. 11 (Nov. 2006): 1348-54.
 4. Valerie Brown, “Environment Becomes Heredity,” *Miller-McCune*, July 14, 2008, 50-9.
 5. Courtney A. Miller and J. David Sweatt, “Covalent Modification of DNA Regulates Memory Formation,” *Neuron* 53, no. 6 (March 15, 2007): 857-69.
 6. Joseph H. Nadeau, “Transgenerational Genetic Effects on Phenotypic Variation and Disease Risk,” *Human Molecular Genetics* 18, no. 2 (2009): R202-10.
 7. David G. Dineen, Andreas Wilm, Pádraig Cunningham, and Desmond G. Higgins, “High DNA Melting Temperature Predicts Transcription Start Site Location in Human and Mouse,” *Nucleic Acids Research* 37, no. 22 (2009): 7360-7.
 8. Remo Rohs, Sean M. West, Alona Sosinsky, *et al.*, “The Role of DNA Shape in Protein-DNA Recognition,” *Nature* 461 (Oct. 29, 2009): 1248-53.
 9. Tom Tullius, “DNA Binding Shapes Up,” *Nature* 461 (Oct. 29, 2009): 1225-6.
-

10. Sebastiaan H. Meijzing, Miles A. Pufall, Alex Y. So, *et al.*, “DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity,” *Science* 324 (April 17, 2009): 407–10.
11. Christoph Geserick, Hellmuth-Alexander Meyer, and Bernard Haendler, “The Role of DNA Response Elements as Allosteric Modulators of Steroid Receptor Function,” *Molecular and Cellular Endocrinology* 236, no. 1–2 (May 31, 2005): 1–7.
12. Rocco Moretti, Leslie J. Donato, Mary L. Brezinski, *et al.*, “Targeted Chemical Wedges Reveal the Role of Allosteric DNA Modulation in Protein-DNA Assembly,” *ACS Chemical Biology* 3, no. 4 (2008): 220–9.
13. Grzegorz Kudla, Andrew W. Murray, David Tollervy, and Joshua B. Plotkin, “Coding-sequence Determinants of Gene Expression in *Escherichia coli*,” *Science* 324 (April 10, 2009): 255–8.
14. Stephen C. J. Parker, Loren Hansen, Hatice Ozel Abaan *et al.*, “Local DNA Topography Correlates with Functional Noncoding Regions of the Human Genome,” *Science* 324 (April 17, 2009): 389–92.
15. Shelley L. Berger, “The Complex Language of Chromatin Regulation During Transcription,” *Nature* 447 (May 24, 2007): 407–12.
16. Bradley R. Cairns, “The Logic of Chromatin Architecture and Remodelling at Promoters,” *Nature* 461 (Sep. 10, 2009): 193–8.

GLOSSARY

- **amino acid.** Amino acids are, among other things, constituent elements of protein. There are twenty different amino acids in protein, and any number of amino acid molecules—up to many thousands—are arranged in sequence to form the main body of a particular protein.
- **base pairs.** See “nucleotide base,” below.
- **bind.** To attach chemically; form a chemical bond with. The term **binding site** refers to the particular sequence of nucleotide bases on a DNA or RNA molecule that a protein or RNA molecule can “target” and attach to. In the case of RNA, its affinity for another RNA or DNA is a matter of sequence (base pair) complementarity. But in the case of a protein, its affinity for a binding site is given by its own molecular folded shape, distribution of electrical charges, and perhaps other characteristics.
- **chromatin.** The complex of DNA, proteins, and RNA that constitutes chromosomes. The histones that form nucleosome “spools” are the most abundant proteins in chromatin, but many other proteins also play a role. The chromatin is highly dynamic in form and structure.
- **codon.** The “words” of the genetic code consisting of three successive nucleotide bases, or “letters.”
- **DNA.** Deoxyribonucleic acid, a molecule that figures centrally in inheritance. Constituting part of the material of chromosomes, it is commonly double-stranded in the famous double helix form. Connecting the two strands are base pairs consisting of nucleotide bases.

- **DNA methylation.** The attachment of a methyl chemical group to particular nucleotide bases (usually cytosine) of the DNA molecule. Methylation plays a major role in gene regulation; it tends to repress gene expression.
- **epigenetics.** Literally, that which is “added to” genetics. The term is most commonly taken to refer to heritable changes in gene expression that do not result from changes in actual gene sequences. (“Heritable” here can refer not only to inheritance between parents and offspring, but also between parent and daughter cells in a single organism.) The changes result from the way the larger cellular context interacts with the genes.
- **gene expression.** A gene is generally said to have been “expressed” when it results in a protein or RNA. The term **gene regulation** refers to the cell’s overall management of gene expression—activating genes, silencing them, and so on.
- **genome.** All the DNA in an organism or cell, especially with reference to the total sequence of bases or “letters” of the genetic code.
- **histone.** A family of simple proteins, abundant in the cell nucleus and constituting a substantial part of the chromatin. A group of histones makes up the “spool” of a nucleosome, each with a thin, filamentary **histone tail** extending out.
- **nucleosome.** The “spool,” made up of histones, around which DNA is commonly wrapped about two turns. (The length of DNA wrapped around a “standard” nucleosome is commonly given as 147 base pairs. But many variations upon this standard length are currently being investigated.) There are millions of nucleosomes in the human genome; they are a focus of many different aspects of gene regulation.
- **nucleotide base.** Chemical groups that are constituents of DNA and RNA. The four main bases in DNA are adenine, guanine, cytosine, and thymine (A, G, C, and T, respectively—the “letters” of the genetic code). In RNA, uracil (U) stands in the place of thymine. These bases combine in restricted ways to form complementary base pairs, a fact that is central to DNA replication and gene expression.
- **protein.** Also known as polypeptides, proteins are folded chains of amino acids. They play myriad structural, regulatory, and enzymatic roles in every cell.
- **RNA.** Ribonucleic acid, like DNA, contains a series of nucleotide bases (the “letters” of the genetic code). Although RNA was classically thought of as existing in three primary forms (mRNA, rRNA, and tRNA), more recently, a great variety of RNA types have been discovered. They play a major role in many epigenetic processes.
- **RNA polymerase.** The enzyme (protein) that transcribes DNA into RNA; see “transcription” below.
- **transcription.** The process by which an RNA polymerase (in cooperation with many other cellular elements) uses a DNA gene template to form an RNA molecule. The gene is said to have been “transcribed,” and the RNA is a “transcript.” **Transcription factors** are proteins that play a part in gene expression, activating it or repressing it, by binding directly to DNA.