IDENTICAL TWINS have identical DNA sequences. Yet in most cases where one twin develops a complex disease known to have a genetic component, such as schizophrenia, bipolar disorder or childhood diabetes, the other twin does not. Environmental factors may play a role, but increasingly biologists are realizing that important traits can be transmitted epigenetically, through the chromosomes but outside the DNA.

The Unseen Genome: Beyond DNA

BY W. WAYT GIBBS

DNA was once considered the sole repository of heritable information. But biologists are starting to decipher a separate, much more malleable layer of information encoded within the chromosomes. Genetics, make way for epigenetics

"Human Genome Placed on Chip"
read the headline this past October

as the *New York Times* reported that three biotech companies have made thumbnail-size devices that can record the activity of all the genes in a sample of human tissue. Thus is fulfilled one of the promises of the Human Genome Project: by scanning the human DNA sequence, scientists can now guess which bits are the genes that are transcribed into RNA messages and then translated into functional proteins.

When the "final draft" of the sequence was released in April, many said that the string of three billion A, T, G and C bases in human DNA represents—choose your metaphor—the book of inheritance, the source code of cells, the blueprint for a life. But in truth, all these metaphors mislead.

A genome, the sum of heritable information that is held in the chromosomes and that governs how an organism develops, is not a static text passed from one generation to the next. Rather a genome is a biochemical machine of awesome complexity. Like all machines, it operates in three-dimensional space, and it has distinct and dynamic interacting parts.

Protein-coding genes make up just one of those parts—and often a small one at that, accounting for less than 2 percent of the total DNA in each human cell. But for the better part of five decades, those genes were enshrined by the central dogma of molecular biology as the repository of heritable traits. Hence the notion of the genome as a blueprint.

As far back as the 1960s, experimenters had uncovered important information hiding elsewhere in the chromosomes. Some was tucked among the "noncoding" DNA, and some lay outside the DNA sequence altogether. The tools of genetic engineering worked best on conventional genes and proteins, however, so scientists looked hardest where the light was brightest.

In recent years, geneticists have been exploring the less visible parts of the genome more thoroughly, in search of explanations for anomalies that contradict the central dogma: illnesses that run in families but pop up unpredictably, even differing among identical twins; genes that switch on or off in cancers yet harbor no mutations; clones that usually die in the womb. They have found that these second and third layers of information, distinct from the protein-coding genes, connect in surprisingly deep and potent ways to inheritance, development and disease.

In the November issue of *Scientific American,* "The Unseen Genome: Gems among the Junk" described those connections for the second layer, which consists of myriad "RNA only" genes sequestered within vast stretches of noncoding DNA. Science had dismissed such DNA as the useless detritus of evolution, because no proteins are made from it. But it turns out that these unconventional genes do give rise to active RNAs, through which they profoundly alter the behavior of normal genes. Malfunctions in RNA-only genes can inflict severe damage.

The third part to the genomic machine, as fascinating as active RNA genes and probably even more important, is the "epigenetic" layer of information stored in the proteins and chemicals that surround and stick to DNA. Epigenetic marks are so named because they can dramatically affect the health and characteristics of an organism—some are even passed from parent to child—yet they do not alter the underlying DNA sequence.

Geneticists have yet to decipher the complex code by which epigenetic marks interact with the other components of the genome. But in working out some of the critical mechanisms, researchers have noticed that the epigenetic part of the genome seems to play crucial roles in growth, aging and cancer. "Epimutations" are also suspected of contributing to diabetes, schizophrenia, bipolar disorder and many other complex ailments.

Epigenetics may suggest new ways to treat these diseases. Whereas cells doggedly protect their DNA against mutation, they routinely add or erase epigenetic marks. In principle, drugs could tinker with the epigenetic code to turn entire sets of rogue genes on and off. New medicines may be able to reverse some of the genetic damage that accompanies aging and precedes cancer.

Overview/*Epigenetics*

- Most traits are transmitted by genes in the DNA that encode proteins. But a separate code, written in chemical marks outside the DNA sequence, also has dramatic effects on the health and appearance of organisms.
- The epigenetic code may explain why some diseases skip generations and affect only one in a pair of identical twins. Epigenetic mistakes seem to play a role in cancer.
- A genome operates like a machine with several complex interacting parts. The epigenetic part should be easier to modify with medicine than the DNA sequence has been.

RANDY HARRIS PHOTOGRAPHY (*preceding page*)

HUGE HINDQUARTERS distinguish a *callipyge* ewe (*far left*) and ram (*right center*) from their normal siblings. The bizarre pattern of inheritance of the *callipyge* trait can be explained only by the interaction of three distinct layers of information in the genome.

SAM JACKSON *Texas Tech University*

Beautiful Buttocks

THE STORY OF SOLID GOLD illustrates how the three parts of the genome can interact to confound conventional notions of inheritance. Born in 1983 on an Oklahoma sheep ranch, a young ram was christened "Solid Gold" after its rear end grew to prodigiously meaty proportions. Sensing a moneymaking mutation, the rancher promptly put the ram out for stud.

Big-bottomed sons of Solid Gold were crossed with normal ewes. Half the offspring, both male and female, took after dad. Researchers called them *callipyge* (pronounced "kalipeezh"), Greek for "beautiful buttocks." A 50–50 split is just what one would expect from a mutation on a dominant gene. "But then things got more interesting," recalls Michel Georges, a researcher at the University of Liège in Belgium who had been called in as a consultant.

When female *callipyge* sheep were mated with normal males, not a single lamb of any sex showed the maximal gluteus so characteristic of its mother, even though some did inherit the mutation. It seemed as if *callipyge* had suddenly switched from a dominant to a recessive characteristic.

The geneticists next tried crossing normal-looking rams who were carriers of the mutation with completely normal ewes. *Et voilà,* half the lambs were *callipyge*. So the trait was appearing only when sheep inherited the mutation from their sires.

"Things got really bizarre," Georges recalls, when breeding yielded sheep bearing two *callipyge* alleles (in other words, the same mutation on both copies of the chromosome). If *callipyge* were a standard gene, then animals inheriting the mutant form from both parents should have been guaranteed thunderous thighs. Yet all the doubly mutated sheep looked perfectly normal [*see illustration on next page*]. What was going on?

Ten years of experiments have finally answered that question. In May, Georges and his co-workers published a recipe for the *callipyge* trait and pedigree: a standard protein-making gene, one or more RNA-only genes, plus two epigenetic effects. The final ingredient is a tiny mutation, a G base where an A normally appears, at a particular spot "in the middle of a gene desert, 30,000 bases from the nearest known gene," Georges says. Somehow the DNA at that spot controls the activity of the recipe's protein-coding and RNA-only genes on the same chromosome.

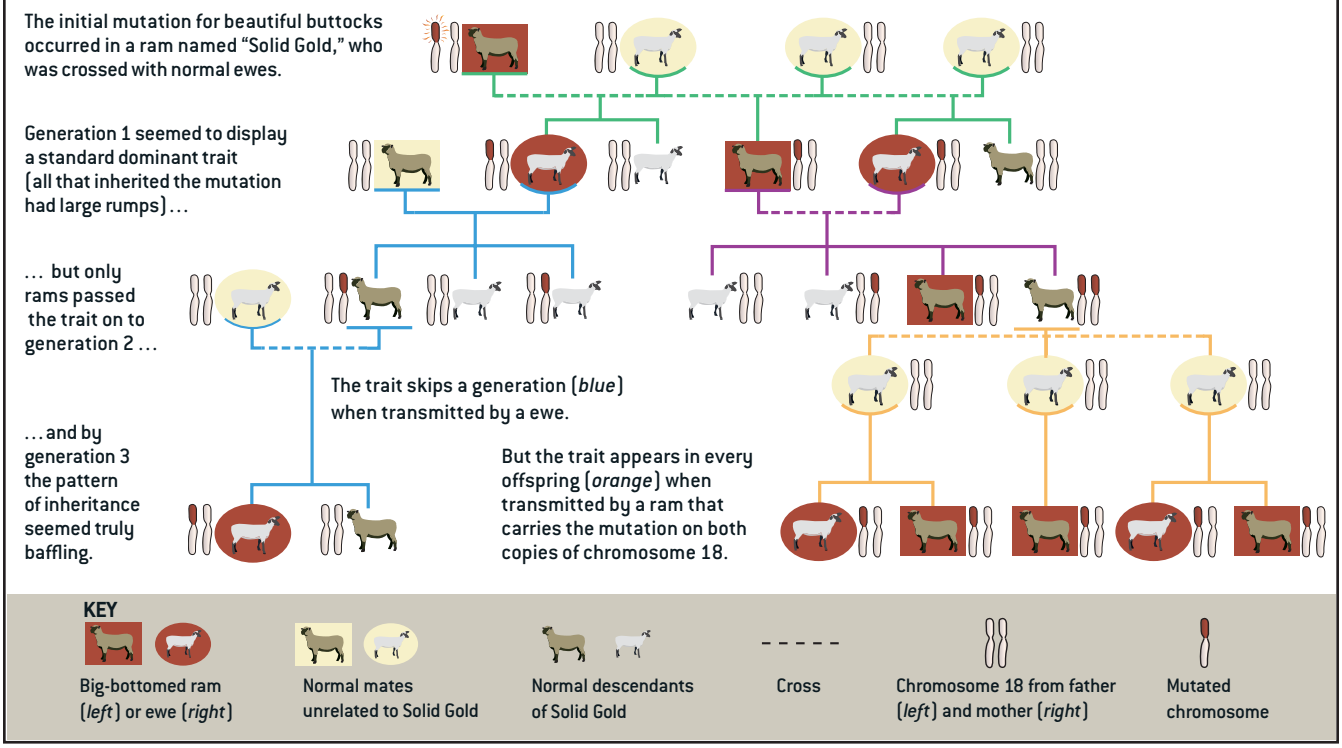
The A-to-G alteration can make those genes hyperactive, so that too much protein or active RNA is made in muscle cells. Excess protein explains the huge hindquarters—but not the odd inheritance pattern. Georges and others see an epigenetic phenomenon, imprinting, at work in the family tree.

For most genes, both the maternal and paternal alleles turn on or off at the same time. Imprinting disrupts this balance. For some imprinted genes, only the copy that came from dad is expressed; the allele inherited from mom is silenced. The protein-making, rump-plumping gene involved in *callipyge* works this way. That is why sheep receiving the A-to-G mutation from mom look normal; the mutation cannot override the selective censorship imposed by imprinting.

The opposite form of imprinting affects the *callipyge* gene (or genes) that makes active RNAs. Those RNAs are produced only from the allele on the maternal chromosome. This second bit of epigenetic wizardry helps to explain why the trait disappears from animals carrying two *callipyge* alleles.

TWISTS AND TURNS IN A FAMILY TREE

TWENTY YEARS AGO a sheep named "Solid Gold" was born with a mutation on chromosome 18 that caused its rump to grow unusually large. Solid Gold passed the trait to about half its offspring (*green*), the typical pattern for a dominant gene. Later generations revealed, however, that sheep inheriting the mutation from their mother look normal (*blue*)—even if they get a second copy of the gene from their father (*purple*). Because of epigenetic effects, the only sheep that develop big bottoms are those that receive just one copy of the mutation, from their father (*orange*).



In these double-mutant sheep the mutation on dad's chromosome throws the protein-making gene into overdrive. At the same time, the copy of the A-to-G mutation on mom's chromosome boosts levels of active RNAs from the RNA-only genes. Somehow the amplified RNAs block the amplified growth signal, and so the lamb looks svelte.

Such "overdominance" effects seem to be rare. Imprinting, however, is quite common, especially in flowering plants. Randy L. Jirtle of Duke University keeps a running list of imprinted human genes; the number is now up to 75. Many more may await discovery. Maxwell P. Lee of the National Cancer Institute reported in August that a scan of 602 genes in seven people found one allele to be significantly more active than the other in half the genes. For 170 of those genes, the difference between the alleles' expression exceeded a factor of four.

In the first few days after conception, nearly all imprinting is removed from the chromosomes. How this happens is a mystery. But sometime between then and mid-gestation, the epigenetic state is reestablished, says Emma Whitelaw of the University of Sydney. Reprogramming mistakes do happen, however.

The human gene for insulin growth factor 2 (*IGF2*), for instance, normally is imprinted; the maternal copy is deactivated. Yet about one person in 10 has no imprinting at the *IGF2* gene. "When we go into the clinic, we find that defect in 40 percent of people who have sporadic colon cancer," notes Carmen Sapienza of Temple University. "It is just an association, but it is very interesting," he says. A blood test that detects the loss of *IGF2* imprinting is already being evaluated as a way to predict the risk of colon cancer. Faulty imprinting is also a prime suspect in several rarer genetic diseases, such as Prader-Willi, Angelman and Beckwith-Wiedemann syndromes. The last causes facial deformities and an elevated risk of childhood cancer.

Epigenetic variations "could explain the odd discordance of diseases among identical twins," Whitelaw suggests. Identical twins share identical DNA sequences. But when one acquires a disease with a genetic component, such as schizophrenia, bipolar disorder or childhood diabetes, the other "identical" twin usually does not. Last year a group led by Rosanna Weksberg of the Hospital for Sick Children in Toronto compared twins discordant for Beckwith-Wiedemann syndrome and found that in every case the affected twin had lost imprinting within a critical area on chromosome 11, whereas the healthy twin had not.

"Clearly for cancer, for development, for birth defects, it is a very important phenomenon," says Francis Collins, director of the National Human Genome Research Institute. "How imprinting works is still not entirely understood. But DNA methylation seems to play a very significant role."

SOURCE: CAROLE CHARLIER AND MICHEL GEORGES *University of Liège, Belgium*; NADIA STRASSER (*illustration*)

This Be Madness, Yet There Is Methyl in It

SIMPLE YET POWERFUL, methyl consists of a carbon, three hydrogens and a hankering to bond to—to methylate—something else. Methyl has a special affinity for the C (cytosine) bases of DNA. Special-purpose enzymes take methyl molecules derived from basic nutrients, such as folic acid and vitamin B_{12}, and stick them onto certain C bases throughout the genome.

In general, the more methylated a stretch of DNA, the less likely it is to be transcribed to RNA and to carry out its function. The silent allele of an imprinted gene is almost always highly methylated, for example. But imprinting may be a side job for DNA methylation; it mainly seems to defend the genome against parasitic genetic elements called transposons.

"We like to think of the genome as this pristine endowment," observes Timothy H. Bestor of Columbia University. "But revolting as it may seem, our DNA is filled with genetic parasites." Roughly 45 percent of the human DNA sequence consists of viral genes (or gene fragments) that have copied themselves into the genome during the course of evolution. Fortunately for us, nearly all of this selfish DNA is heavily methylated and rendered inactive.

Jirtle's lab at Duke demonstrated the tight link between methyls and transposons this summer in a fascinating experiment with agouti mice, whose fur color varies from yellow to black under the control of a parasitic element. One group of pregnant agouti mice ate a normal diet; about 60 percent of their offspring grew yellow coats. But another group was fed chow enriched with vitamin B_{12}, folic acid and other good sources of methyl. The high-methyl diet changed the hair color of the resulting litter; now 60 percent developed brown coats. The shift appeared to be the result solely of increased methylation (and reduced expression) of the agouti transposon DNA.

But what happens if the methyl defense falters? In a famous study five years ago, genetic engineers disabled one of the methyl-adding enzymes in embryonic stem cells. With the methyl guard lowered, many transposons became active. The rate of DNA mutations in the cells shot up 10-fold. Such experiments raised an intriguing possibility: Could epigenetic abnormalities accelerate—perhaps even initiate—the genetic chaos that leads to cancer?

After all, tumor cells often contain both too little methylation in the genome overall and, confusingly, too many methyl molecules attached to certain genes that normally prevent deranged cells from becoming malignant. "In colon polyps [benign growths from which tumors often arise], we can already see a genome-wide reduction in DNA methylation" occurring even before mutations knock out key antigrowth genes on the road to cancer, says Stephen B. Baylin of Johns Hopkins University.

No one knows why so many methyls fall off the DNA in the first place—no methyl-removing enzyme has been positively identified. But researchers suspect that methyl-poor chromosomes are more likely to malfunction during cell division, taking a step toward malignancy.

Work this year by Rudolph Jaenisch of the Whitehead Institute at M.I.T. reinforced that suspicion. His group created mice with an inborn deficiency of a methylating enzyme. In most of the mice, at least one of the undermethylated chromosomes became unstable. Mutations accumulated quickly, and 80 percent of the mice died from cancer within nine months.

The idea that a lack of methyl on the DNA can lead to human cancer is still just a hypothesis, and in any case oncologists have no drugs that can correct genome-wide undermethylation. But doctors are testing several anticancer drugs that attack the other methyl problem: too much of it stuck on some cancer-related genes. Until recently, many scientists believed that a tumor could take hold only after mutation had knocked tumor suppressor genes out of commission. Yet in many tumor cells these cancer-fighting genes have a perfectly normal DNA sequence. Methylation mistakes, not mutations, lay the genes low.

> "When the drug works, the leukemia goes away: 99.9 percent of the cancerous cells are gone."

Drugs such as the anesthetic procaine, the mood stabilizer valproic acid and the chemotherapy agent decitabine all seem to either strip methyl groups from DNA or prevent methyl tags from being attached to newly formed cells. Jean-Pierre Issa has been testing decitabine in patients with advanced leukemia at his clinic at the M. D. Anderson Cancer Center at the University of Texas. Like most chemotherapies, the compound is quite toxic. But "when the drug works," Issa says, "the leukemia goes away: 99.9 percent of the cancerous cells are gone." Eight of 130 patients had such good fortune, in a controlled trial Issa published in August, and in 22 others the demethylating medicine sent the disease into partial remission.

"These drugs are quite promising," avers Sabine Maier of Epigenomics, a biotech company in Berlin that is working with Roche in Basel, Switzerland, to develop methylation-based diagnostics for cancer. "But there is one problem," she adds. "The drugs all lead to demethylation of the whole genome. This probably causes side effects."

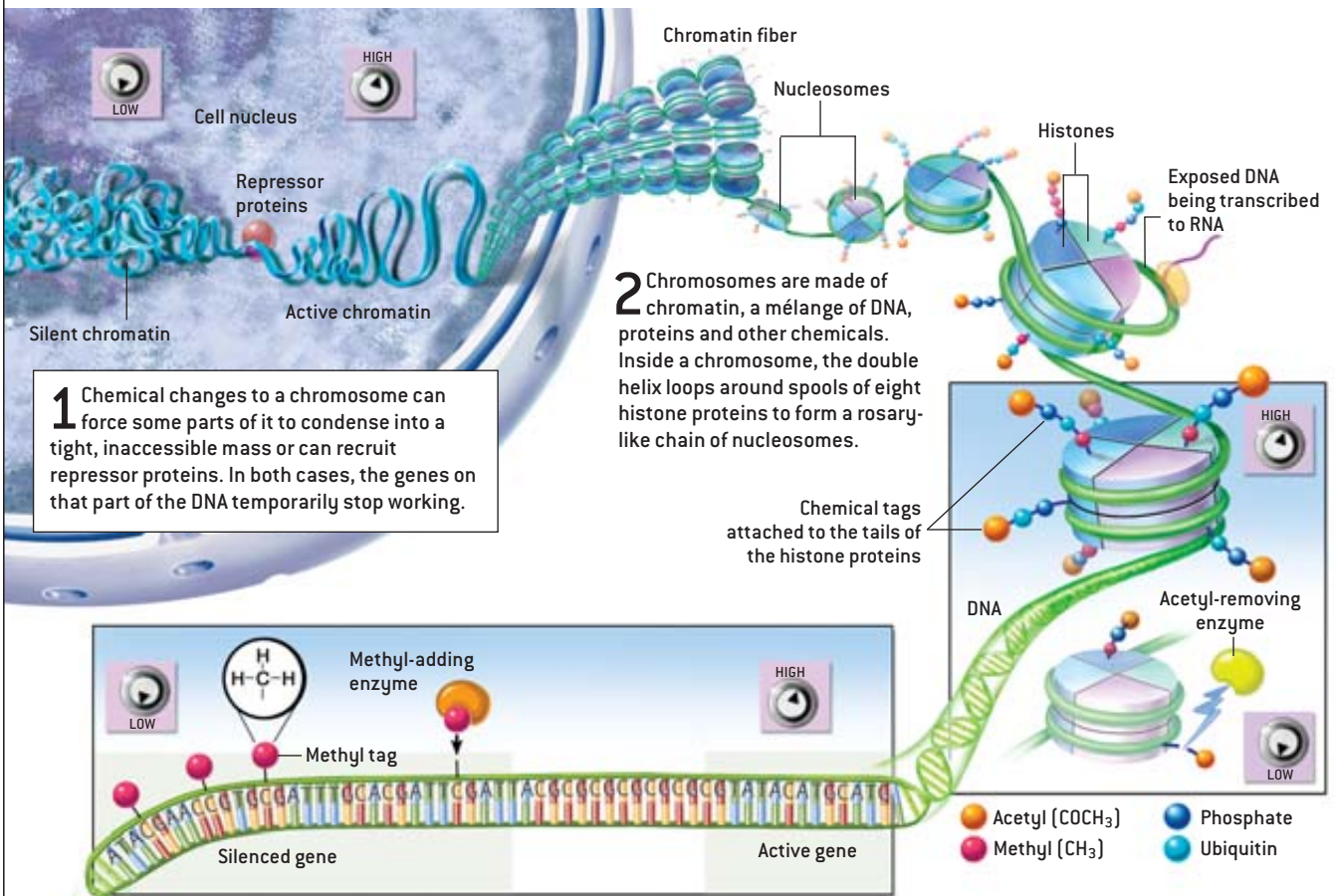
Another worry is that the effect is temporary: methyl tags soon start popping up again, and the tumor suppressor genes switch back off. "The drug-induced change in gene expression may not be permanent," Issa acknowledges, "but if it changes in such a way that the immune system can identify the tumor cell or that induces apoptosis [cell suicide], then the cell is still dead."

Breaking the Code

THE REEMERGENCE of a distinctive DNA methylation pattern after drugs wipe it clean strangely echoes the reprogramming of an embryo's imprinting marks shortly after conception. What directs the methyl-adding enzymes back to those tumor-suppressing genes or to those few alleles that should be imprinted?

VOLUME CONTROLS FOR GENES

THE DNA SEQUENCE is not the only code stored in the chromosomes. So-called epigenetic phenomena of several kinds can act like volume knobs to amplify or mute the effect of genes. Epigenetic information is encoded as chemical attachments to the DNA or to the histone proteins that control its shape within the chromosomes. Among their many functions, the epigenetic volume controls muffle parasitic genetic elements, called transposons, that riddle the genome.

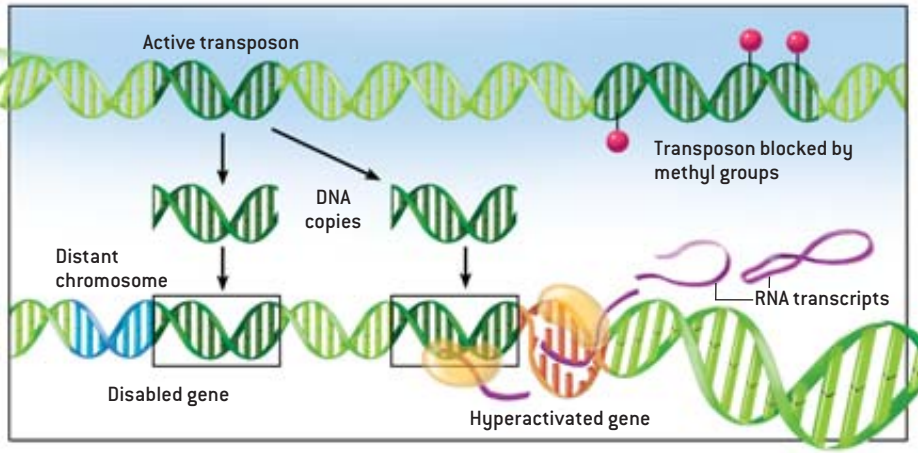


1 Chemical changes to a chromosome can force some parts of it to condense into a tight, inaccessible mass or can recruit repressor proteins. In both cases, the genes on that part of the DNA temporarily stop working.

2 Chromosomes are made of chromatin, a mélange of DNA, proteins and other chemicals. Inside a chromosome, the double helix loops around spools of eight histone proteins to form a rosary-like chain of nucleosomes.

3 An intricate histone code—written in chemical tags stuck to the histones' tails (*above*)—governs gene expression as well. Acetyl tags usually amplify nearby genes, whereas acetyl-removing enzymes mute them. But the rest of the code remains to be deciphered.

4 Genes can also be suppressed by methyl tags that stick directly to the DNA, usually at places where a C base is followed by a G. Whether DNA methylation turns down genes independently or only in combination with histone tags is still a mystery.



5 Transposons, also called jumping genes, can clone themselves and then insinuate the copies into distant sections of the genome, sometimes disabling or hyperactivating genes. One major function of DNA methylation seems to be the suppression of transposons, which make up almost half the human genome.

TERESE WINSLOW

That question must be answered if cloning is ever to become routine. Currently epigenetic reprogramming goes terribly awry in clones that are made by replacing the DNA in a fertilized egg with DNA from an adult cell. "The majority of such clones display abnormal patterns of methylation and gene expression," says David Wells, a cloning expert at AgResearch in Hamilton, New Zealand. Even though their DNA sequence may be fine, 90 percent of the animals die before birth; half of those born alive never make it to adulthood. The few that survive to maturity tend to suffer obesity and diseases of the immune system.

To permanently prevent or reverse the methylation errors so common in clones, tumors and imprinting disorders, researchers must decipher a related epigenetic code—one altogether separate from DNA. "Methylation alone doesn't silence the genes," Baylin of Johns Hopkins says, "it just locks in the silent state." The methyl-adding enzymes seem to take their orders from elsewhere.

Zoom in on a chromosome, and you will find that it is not (as often drawn) a haphazard tangle of DNA, nor even a single object, but a dynamic assembly of DNA, proteins and other chemicals. This filamentlike assemblage, called chromatin, does more than support the DNA. It also controls access to it.

Chromatin contains half as much DNA as it does protein, most of which is in the form of histones. Histones are nature's answer to the question: How does a cell fit 1.8 meters of DNA into its nucleus? In a word, clever packaging. DNA wraps around histone spools to form a rosarylike chain, which is then twisted into a bundle [*see illustration on opposite page*]. Sections of chromatin can condense and expand independently, effectively hiding whole swaths of the DNA from view while exposing other sections for transcription.

Females, for example, start out life with two active X chromosomes; males inherit just one. A female embryo must muzzle the extra X to prevent its cells from getting a double dose of X-borne genes. To do this, two parts of the genomic machine conspire to shut down the third. A noncoding gene named *Xist* produces an active RNA that coats the unneeded X chromosome. The needed X meanwhile produces "antisense" RNA, which acts like an antidote to protect it from *Xist*. A chain reaction spreads down the unwanted chromosome: methyls tag much of the DNA, histones shed the chemical acetyl from their tails, and the chromatin compacts into an inaccessible, RNA-coated mass. The silent X chromosome is then passed down, inactive, to every genome-bearing cell as the woman grows.

The role of histones in this drama is not clear, but recent work has shown that the protein tails that hang off the histone spools can sport an impressive array of chemical additions. Where acetyls adorn certain spots on the histones, for example, the chromatin is usually open for business, allowing the cell's transcription machinery to read the DNA in that part of the chromosome.

Compact, silent chromatin generally lacks acetyls in the special positions and instead will often have methyl groups stuck at different points on the histone tails. The histones also play host to phosphate groups and to the peptide ubiquitin—and all of these tags appear in a bewildering variety of locations and combinations. The histone code will not be easy to crack.

Unlike the stable genetic code of DNA, many epigenetic marks are in constant flux. When one section of chromatin condenses, the silence can spread along the chromosome until it hits a barrier. Xin Bi of the University of Rochester recently identified boundary elements that recruit acetyl-adding enzymes to histones, ensuring that they stay active. Sometimes a physical gap where the DNA floats free of any histones can halt the spread, Bi says. At other places, there is no boundary, just a continual tug-of-war between the active and silent regions of the chromosome.

Issa thinks this struggle might explain why cancer risk rises so steadily with age. Perhaps the barriers in the chromosomes that separate the tightly condensed, highly methylated and silent portions from the accessible, unmethylated and active portions break down over the years as cells divide or grow old.

"There is a whole new universe out there that we have been blind to. It is very exciting."

The darker parts of the genome are still perceived only dimly. But it is quite clear, Sapienza asserts, that "the Human Genome Project was just the beginning of the job. We now need to produce a similar description of the epigenetic landscape." In October, Epigenomics and the Wellcome Trust Sanger Institute in the U.K. undertook to do just that, launching a five-year Human Epigenome Project to map all the DNA methylation sites. The consortium also announced its completion of a map of more than 100,000 methyl tags attached to the major histocompatibility complex, a section of chromosome 6 linked to many diseases.

The new view of the genomic machine is energizing, because it opens avenues to genomic engineering. Those 30,000-odd protein-coding genes, so important yet so immutable, are not the only instruction set to which cells refer. Noncoding DNA matters. Chemical attachments to DNA and to the histones matter. The shape of chromatin matters. And all of these are subject to manipulation. "There is a whole new universe out there that we have been blind to," Bestor says. "It is very exciting." SA

W. Wayt Gibbs is senior writer.

MORE TO EXPLORE

The Epigenome: Molecular Hide and Seek. Edited by Stephan Beck and Alexander Olek. Wiley, 2003.

Controlling the Double Helix. Gary Felsenfeld and Mark Groudine in *Nature,* Vol. 421, pages 448–453; January 23, 2003.

The Callipyge Locus: Evidence for the *Trans* Interaction of Reciprocally Imprinted Genes. Michel Georges, Carole Charlier and Noelle Cockett in *Trends in Genetics,* Vol. 19, No. 5, pages 248–252; May 2003.

Summaries of recent research, lists of imprinted genes, and other information on epigenetics are available online at **geneimprint.com**