FLECKS OF DARK BROWN in an iris may be a telltale sign of the hidden genome at work. Certain traits are transmitted not through ordinary genes but rather through chemical modifications to the chromosomes, changes that are regulated in part by bits of "junk" DNA. Unlike genetic mutations, these heritable traits are often reversible and appear in some cells but not others. (The white sphere on the iris is a reflection of the light shining on the eye.)

Just when scientists thought they had DNA almost figured out, they are discovering in chromosomes two vast, but largely hidden, layers of information that affect inheritance, development and disease.

The Unseen Genome:
Gems among the Junk

BY W. WAYT GIBBS
that distant galaxies were moving in ways that made no sense, given the laws of gravity and the fabric of celestial objects visible in the sky. Gradually they were forced to conclude that the universe is not as empty as it appears, that in fact it must be dominated by some dark kind of matter. Although no one knew what the stuff is made of or how it works, scientists could see from its effects that it is out there. The quest to understand dark matter (and more recently, dark energy) meant revising or replacing theories, but it reenergized astrophysics and cosmology.

A similar revelation is now unfolding in molecular genetics. This year biologists celebrated the 50th anniversary of the discovery of the double helix, and the Human Genome Project announced its completion of a “final draft” of the DNA sequence for *Homo sapiens*. Scientists have clearly mastered DNA in the lab. Yet as they compare the DNA of distantly related species and look more closely at how chromosomes function in living cells, they are increasingly noticing effects that current theories cannot explain.

Journals and conferences have been buzzing with new evidence that contradicts conventional notions that genes, those sections of DNA that encode proteins, are the sole mainspring of heredity and the complete blueprint for all life. Much as dark matter influences the fate of galaxies, dark parts of the genome exert control over the development and the distinctive traits of all organisms, from bacteria to humans. The genome is home to many more actors than just the protein-coding genes.

**Overview/Hidden Genes**

- Geneticists have long focused on just the small part of DNA that contains blueprints for proteins. The remainder—in humans, 98 percent of the DNA—was often dismissed as junk. But the discovery of many hidden genes that work through RNA, rather than protein, has overturned that assumption.
- These RNA-only genes tend to be short and difficult to identify. But some of them play major roles in the health and development of plants and animals.
- Active forms of RNA also help to regulate a separate "epigenetic" layer of heritable information that resides in the chromosomes but outside the DNA sequence.

The extent of this unseen genome is not yet clear, but at least two layers of information exist outside the traditionally recognized genes. One layer is woven throughout the vast “noncoding” sequences of DNA that interrupt and separate genes. Though long ago written off as irrelevant because they yield no proteins, many of these sections have been preserved mostly intact through millions of years of evolution. That suggests they do something indispensable. And indeed a large number are transcribed into varieties of RNA that perform a much wider range of functions than biologists had imagined possible. Some scientists now suspect that much of what makes one person, and one species, different from the next are variations in the gems hidden within our “junk” DNA.

Above and beyond the DNA sequence there is another, much more malleable, layer of information in the chromosomes. “Epigenetic” marks, embedded in a melange of proteins and chemicals that surround, support and stick to DNA, operate through cryptic codes and mysterious machinery. Unlike genes, epigenetic marks are routinely laid down, erased and rewritten on the fly. So whereas mutations last a lifetime, epigenetic mistakes—implicated in a growing list of birth defects, cancers and other diseases—may be reversible with drugs. In fact, doctors are already testing such experimental treatments on leukemia patients.

Researchers are also coming to realize that just about anything that can happen in the genome does happen, says Carmen Sapienza of Temple University, who started investigating epigenetic phenomena back when they were dismissed as minor anomalies. “There may even be fundamental mechanisms still to discover,” Sapienza considers. “I think we are entering the most interesting time yet in genetics.”

**IT WILL TAKE YEARS, perhaps decades, to construct a detailed theory that explains how DNA, RNA and the epigenetic machinery all fit into an interlocking, self-regulating system. But there is no longer any doubt that a new theory is needed to replace the central dogma:that: has been the foundation of molecular genetics and biotechnology since the 1950s.**

The central dogma, as usually stated, is quite simple: DNA makes RNA, RNA makes protein, and proteins do almost all the real work of biology. The idea is that information is stored in the twisted ladders of DNA, specifically in the chemical bases
BIG DIFFERENCES in the appearance and health of organisms can arise from small changes to tiny, unconventional genes. Arabidopsis plants, for example, normally have spoon-shaped leaves (left). But when scientists interfered with the action of a microRNA, produced by an RNA-only gene, the mutant Arabidopsis plants developed gross defects (right). The microRNA appears to control the activity levels of numerous genes.

(percent of the DNA in human chromosomes. Three billion or so pairs of bases that we all carry in nearly every cell are there for some other reason. Yet the introns within genes and the long stretches of intergenic DNA between genes, Mattick says, "were immediately assumed to be evolutionary junk.

That assumption was too hasty. "Increasingly we are realizing that there is a large collection of 'genes' that are clearly functional even though they do not code for any protein" but produce only RNA, Georges remarks. The term "gene" has always been somewhat loosely defined; these RNA-only genes muddle its meaning further. To avoid confusion, says Claes Wahlestedt of the Karolinska Institute in Sweden, "we tend not to talk about 'genes' anymore; we just refer to any segment that is transcribed [to RNA] as a 'transcriptional unit.'"

Based on detailed scans of the mouse genome for all such elements, "we estimate that there will be 70,000 to 100,000," Wahlestedt announced at the International Congress of Genetics, held this past July in Melbourne. "Easily half of these could be noncoding." If that is right, then for every DNA sequence that generates a protein, another works solely through active forms of RNA—forms that are not simply intermediate blueprints for proteins but, rather, directly alter the behavior of cells.

What is true for mice is probably true for people and other animals as well. A team of scientists at the National Human Genome Research Institute (NHGRI) recently compared excerpts from the genomes of humans, cows, dogs, pigs, rats and seven other species. Their computer analysis turned up 1,194 segments that appear with only minor changes in several species, a strong indication that the sequences contribute to the species' evolutionary fitness. To the researchers' surprise, only 244 of the segments sit inside a protein-coding stretch of DNA. About two thirds of the conserved sequences lie in introns, and the rest are scattered among the intergenic "junk" DNA.

"I think this will come to be a classic story of orthodoxy detailing objective analysis of the facts, in this case for a quarter of a century," Mattick says. "The failure to recognize the full im-
NOW THAT BIOLOGISTS have turned their attention back to RNA, they are finding it to be capable of impressive feats of cellular chemistry. Like proteins, some RNA transcripts can interact with other bits of RNA, with DNA, with proteins and even with small chemical compounds. Proteins are analog molecules, however; they bind to targets in much the way keys fit in locks. “The beauty of RNA is that it has a specific sequence, so it’s digital, like a zip code,” Mattick points out. A bit of RNA can float around until it bumps into a DNA (or another RNA) that has a complementary sequence; the two halves of the ladder then join rungs. (Two segments are complementary when all C bases mate with G’s and all T or U bases join to A’s.)

The failure to recognize the importance of introns

in the history of molecular biology.”

As an example of the unappreciated power of RNA, consider pseudogenes. Surveys of human DNA have found in it almost equal numbers of genes and pseudogenes—detective copies of functional genes. For decades, pseudogenes have been written off as molecular fossils, the remains of genes that were broken by mutation and abandoned by evolution. But this past May a group of Japanese geneticists led by Shinji Hirotsume of the Saitama Medical School reported their discovery of the first functional pseudogene.

Hirotsume was genetically engineering mice to carry a fruit fly gene called sex-lethal. Most mice did fine with this foreign gene, but in one strain sex-lethal lived up to its name; all the mice died in infancy. Looking closer, the scientists discovered that in those mice sex-lethal happened to get inserted right into the middle of a pseudogene, clobbering it. This pseudogene (named makorin1-p1) is a greatly shortened copy of makorin1, an ancient gene that mice share with fruit flies, worms and many other species. Although researchers don’t know what makorin1 does, they do know that mice have lots of makorin1 pseudogenes and that none of them can make proteins. But if pseudogenes do nothing, why were these mice dying when they lost one?

For some reason, makorin1—and apparently only makorin1—all but shuts down when its pseudogene p1 is knocked out. RNA made from the pseudogene, in other words, controls the expression of the “real” gene whose sequence it mimics, even though the two lie on different chromosomes. There is nothing pseudo about that.

It is too early to say whether many pseudogenes give rise to active RNA. But there are plenty of other sources scattered about the dark parts of the genome. Every normal protein-making gene, for instance, has a complementary DNA sequence that sits on the other side of the ladder and usually is not transcribed into RNA. Biologists like to think of this as a backup copy, because the cell can use it to repair damage to the gene.

In some cases, however, the backup has its own agenda. While the gene is producing a sensible RNA message, its alter ego can churn out an “antisense” RNA that has a complementary sequence. Whenever matched sense and antisense RNAs meet, they mesh to form their own double-stranded ladders—effectively interfering with the gene’s ability to express its protein.

Biologists knew that bacteria and plants can produce antisense, but most thought that mammals rarely do. In April, Galir Rotman and her co-workers at CompuGen, a biotech firm in Tel Aviv, dashed that assumption. They screened human genome databases and concluded that at least 1,600 human genes (and probably many more) have a mate that yields antisense RNAs.

These competing RNAs may suppress a gene just by tying up the gene’s messenger RNA. But Rotman speculates that they employ a built-in genome censor, known as the RNA interference machinery. Scientists are still enthralled by the discovery several years ago of this scheme for selectively silencing individual genes. When double-stranded RNA appears in a cell, enzymes dice it up, peel the two strands apart, and use one RNA fragment to seek out and destroy any other RNA messages that stick to its sequence. The system protects cells against viruses, which often deliver their payloads in the form of double-stranded RNA.

But the censor also provides a handy way for scientists to shut off any gene at will [see “Censors of the Genome,” by Nelson C. Lau and David P. Barrell; SCIENTIFIC AMERICAN, August].

Neither pseudogenes nor antisense RNAs, however, can explain the wrinkled leaves that Detlef Weigel of the Max Planck Institute for Developmental Biology in Tubingen, Germany, and his collaborators saw in their Arabidopsis plants this summer. These weeds of the mustard family normally have smooth, spoon-shaped leaves. The plants owe their gentle symmetrical curves, Weigel’s group showed in Nature this past August, in part to a kind of active RNA called microRNA.

MicroRNAs, first observed a few years ago in roundworms, are short noncoding RNAs that fold back on themselves, like hairpins. In Arabidopsis, the JAW microRNA doubles over and is then captured by the RNA interference machinery, just as if it had come out of a virus. But the JAW sequence matches a handful of different protein-making genes, members of a family that control the shape and size of the plant. The censor dutifully represses each of them by chopping up much (but not all) of the messenger RNA they produce. Thus, JAW, a tiny RNA-only gene, serves as the main lever by which Arabidopsis cells adjust the volume of a suite of crucial protein genes. When Weigel’s
A BESTIARY OF UNCONVENTIONAL GENES

GENES, according to conventional wisdom, are those sections of the DNA that encode functional proteins. Such sequences make up only about 2 percent of the human genome, however. The rest of the human genome is filled with DNA that is “noncoding”—but not useless. Scientists are discovering many noncoding genes that give rise to surprisingly active RNAs, including varieties that can silence or regulate conventional genes.

Antisense RNA is made from the complementary DNA strand that sits opposite a protein-coding gene on the double helix. Antisense RNAs can intercept the messenger RNA transcribed from the gene, preventing the mRNA from being translated into protein.

Protein-coding genes contain noncoding sections called introns. Introns are snipped out of the initial RNA transcript; the coding sections are then spliced to create a mature mRNA. Although many introns degrade, some contain active elements, such as microRNAs that can exploit the “RNA interference” effect to control other genes...

Riboswitches are a newly discovered form of mRNA that act as precision genetic switches. Produced in many cases from noncoding DNA between known genes, a riboswitch folds into a complex shape. One part of the folded RNA can bind to a specific target protein or chemical. Another part contains the RNA code for a protein product. The riboswitch turns "on" and produces the protein it encodes only when in the presence of its target.
group engineered plants in which the microRNA could not do its job, the plants became sick and deformed.

In just the three years since researchers started looking in earnest, they have found hundreds of microRNAs—more than 150 in humans alone. They seem to be a well-established means for organisms to wrangle genes; about half the microRNAs in humans also appear, in nearly identical form, in the DNA of pufferfish, even though the two species went their separate ways some 400 million years ago.

Just what those 150-plus microRNAs do in people is a mystery. Anna M. Krichevsky of Harvard Medical School suspects that, among other things, they play an important role in brain development. Her lab used a “gene chip” to screen mouse neu-
un for 44 different kinds of microRNA. Krincehsky reported in September that levels of at least nine distinct microRNAs are precisely regulated in the mice as their brains grow. The link is still indirect, but as Diya Banerjee of Yale University noted last year in a review of microRNA science, “it seems that we are on the verge of an explosion of knowledge in this area.”

Proteins may be the draft horses of the cell, but active RNA sometimes wields the whip. And several kinds of RNA have turned up doing mule’s work as well: catalyzing, signaling and switching as competently as any protein. In fact, some inherited diseases have stumped researchers because, in their diligent search for a mutant protein, the investigators ignored the active RNA right under their noses.

Doctors struggled for more than nine years, for example, to nail down the gene responsible for cartilage hair hypoplasia. This recessive disease was first identified in the Amish, one in 19 of whom carries a copy of the defect gene, which causes an unusual kind of dwarfism. People with CHH are not only small in stature but also at high risk for cancer and immune disorders. Geneticist Maaret Ridanpää of the University of Helsinki tracked the gene to chromosome nine, sequenced a large region and then proceeded to check all 10 protein-making genes in the area, one by one. None caused the disease.

Finally, in 2001, Ridanpää and his co-workers identified the culprit, an RNA-only gene called RMRP. The RNA transcribed from RMRP links up with proteins to form an enzyme that works inside a cell’s energy generators, the mitochondria. A change to just a single base at a critical spot on this RNA can mean the difference between a full-size, healthy life and a short, abbreviated one (if the same mutation is inherited from both parents). Such “analog” RNAs, which fold up into complex shapes just as proteins do, have also been discovered recently to be essential to the function of enzymes that protect the chromosomes and that escort secreted proteine signals out of cells’ portholes.

Perhaps the most intriguing form of RNA yet discovered is the riboswitch, isolated last year by Ronald R. Breaker’s lab at Yale. He and others have long wondered how, billions of years ago, the very earliest chemical precursors to life got along in the RNA world before DNA and proteins existed. They speculated that such proto-organisms would need to use RNA as sensors and switches to respond to changes in the environment and in their metabolism. To test the idea, they tried to create RNAs with such capabilities.

“Our laboratory successfully produced a number of synthetic RNA switches,” Breaker recalls. Dubbed riboswitches, these long RNAs are both coding and noncoding at once. As the
CLONES IN ALL BUT NAME, these littermates from a highly inbred strain of mice share practically identical DNA. Yet their coat colors run the spectrum from golden yellow to mahogany brown because of variations in the RNA. Due to epigenetic chemical attachments each has to a particular segment of DNA that lies outside any known gene. The hair color of these mice cannot be predicted by current theories of genetics.

RNA folds up, the noncoding end becomes a sensitive receptor for a particular chemical target. A collision with the target flips the switch, causing the other end, which contains a standard blueprint for a protein, to change shape. The riboswitch thus gives rise to a protein, much like a normal gene does—but only when it senses its target.

Breaker’s group started hunting for riboswitches in the wild and soon found them hiding in intergenic DNA. These precision genetic switches have been extracted now from species in all three kingdoms of life. “This implies that they were very present in the last common ancestor,” not long after the dawn of evolution, Breaker argues.

In August, Breaker and his co-workers reported that one family of riboswitches regulates the expression of no fewer than 26 genes in Bacillus subtilis, a common kitchen bacterium. These are not once-in-a-blue-moon genes, either, but genes that the bacterium relies on to metabolize such basic staples as sulfur and amino acids. Breaker estimates that B. subtilis has at least 68 genes, nearly 2 percent of its total, under the control of riboswitches. His lab has already begun engineering the hybrid digital-analog molecules to do useful things, such as selectively kill germs.

MORE TO EXPLORE


W. Wayt Gibbs is senior writer.