



ALTERNATIVE SPACING

TETRARCHS 4:04 AND 4:16PM 4TH NOVEMBER 2001
FROM *GUEST*, A BLIND SPOT BOOK PUBLISHED BY
POWERHOUSE BOOKS. COURTESY BLIND SPOT ARTIST
REPRESENTATION, NEW YORK / WWW.BLINDSPOT.COM

GENETIC
SPLICE
AND DICE
RESEARCHERS
HAVE LONG
ASSUMED
THAT ONE GENE
USUALLY
CODES FOR
ONE PROTEIN.
BUT THERE'S
EVIDENCE
THAT THE
RULE INSTEAD
IS "ONE
GENE, MANY
PROTEINS."
HOW DOES THAT
HAPPEN?

IN FEBRUARY OF 2000, HHMI INVESTIGATOR

S. Lawrence Zipursky was trying to sort out what he had originally expected would be a relatively simple problem. He and his colleagues at the University of California, Los Angeles (UCLA), had discovered a fruit fly gene that encoded a protein on the surface of nerve cells that helped them migrate and connect to the correct cells. The team was analyzing the makeup of that protein, which they called *Dscam*—for Down syndrome cell adhesion molecule—because of its similarity to a human protein of the same name. They were surprised to find a segment of the protein that did not match the human version and suspected the fruit fly gene might produce a few subtly different proteins. • The situation turned out to be far more complex. The team found that the *Dscam* gene in fact produced numerous variants. And when they pored over the newly released sequence of the fruit fly genome, they realized that *Dscam* had the potential to generate a mind-boggling 38,016 distinct forms of proteins. • It's now

ALTERNATIVE SPLICING

known that *Dscam* is an extreme example of “alternative splicing”—a variable but carefully regulated adaptation of a routine RNA modification process that allows a single gene to give rise to multiple versions of a protein. While most cases are more modest than *Dscam*, the reality of alternative splicing has still managed to turn on its head the “one gene, one protein” principle that has guided genetics for more than half a century. • The ability of alternative splicing to exploit limited genetic information to generate a multitude of proteins is important because each cell in an organism depends on a highly specialized set of proteins to carry out its unique function. Even within a single cell, the set of required proteins varies as a function of the stage of development and changing environmental conditions. With the latest estimates putting the content of the human genome at only 20,000 to 25,000 genes, most scientists believe a one-to-one ratio of genes to proteins just cannot be enough.

BY JENNIFER
MICHALOWSKI
Photograph
by Christopher
Bucklow

Variants Per Gene

DOUGLAS BLACK, another HHMI investigator at UCLA, points out that alternative splicing allows for much more complexity than the size of the genome suggests. It's too soon to know just how big the human *proteome* is, but "it's certainly much larger than the number of genes," he says. "The analyses that have been done seem to say that most genes have two or three splice variants per gene." And while not all of these transcripts (RNA segments) produce functional proteins, there are other genes that generate far greater diversity—hundreds or even thousands of forms.

The first example of alternative splicing in cells—there were earlier examples in viruses—was found in a gene called *IgM* in 1980. It was considered an anomaly. "But if you were paying attention, you started to see more and more examples of genes that were alternatively spliced," says Black. Computational biologists are now trying to assess the frequency of alternative splicing more accurately. "If you now ask how many of the known human genes are producing more than one splice variant, almost everyone says at least 50 percent," Black says. "Some people argue that 70 percent have alternative splicing. And while nothing as complicated as *Dscam* has been discovered in mammalian cells, there are many transcripts that can produce hundreds of proteins."

Cells use alternative splicing to increase protein diversity toward a host of biological ends. Some of the best-

studied examples derive from fruit fly development, where the splicing of several related genes dictates whether an embryo will develop into a male or a female. Similarly, alternative splicing can allow one gene to generate different proteins in different tissues—many of the highly specialized proteins in the brain, for example, come from differential splicing of genes that are also expressed in other tissues. Cells can even modify splicing in response to changing conditions: An ion channel transcript studied by Black's lab produces a protein whose sensitivity to calcium depends on which exons (segments of DNA that encode a protein's amino acid sequence) are included.

Recent work in the lab of HHMI investigator Robert B. Darnell at the Rockefeller University shows that not only can alternative splicing tweak the structure of a single protein, but it may also be a means of regulating entire pathways. In the first genome-wide screen for the targets of a tissue-specific splicing factor, Darnell showed that Nova, a protein found only in the brain, controls the splicing of 49 mRNAs to produce proteins not found in other tissues. Almost all of these proteins help nerve cells transmit their signals across the synapse.

"We don't know this yet, but one can guess that different exons may change the way these proteins interact or send messages," Darnell says. "If that is so, by changing these exons, you can

CELLS USE ALTERNATIVE SPLICING TO INCREASE PROTEIN DIVERSITY TOWARD A HOST OF BIOLOGICAL ENDS.

modulate the quality of the synapse in a very powerful way—'powerful' meaning very regulated." Darnell expects that as similar studies are done with other splicing factors, they too will be found to regulate similarly coherent groups of proteins.

Mutations that alter splice sites often cause entire exons to be excluded, severely damaging the encoded protein; these types of mutations are often associated with human disease. In the case of alternative splicing, however, a cell usually splices a single transcript in multiple ways to generate an assortment of proteins. So some mutations that alter a splice site or a nearby regulatory sequence have subtle effects—shifting the ratio of the resulting proteins without entirely eliminating any form. Several human diseases illustrate that these sorts of mutations still have the potential to be devastating. Alternative splicing errors are known to contribute to some growth deficiencies; a urogenital disorder known as Frasier syndrome; a type of cystic fibrosis; and a condition known as frontotemporal dementia and Parkinsonism. In the latter case, mutations disrupt the normal splicing pattern of a transcript called *tau*, which produces a protein that helps give a nerve cell its shape. Normally, the cell uses the transcript to produce six different forms of tau protein. Interfering with splicing, however, can result in an excess of some of those forms, causing tau to clump together in the brain and bring about progressive dementia.

Rules of Regulation

AS IT BECOMES clear that alternative splicing is more the rule than the exception, scientists are realizing that, to make effective use of the enormous amount of data being generated by genome-sequencing projects, they must understand how the cell's splicing machinery processes that information. "Proteins are the major workhorse of the organism. If we want to have a really big-picture view of how organisms develop and function, we need to know what all the proteins are," says Brenton R. Graveley, an associate professor at the

THREE OF THE HHMI INVESTIGATORS INVOLVED IN SPLICING RESEARCH



1. Robert Darnell
Robert Darnell studies degenerative brain disorders that are provoked by an immune response to certain cancers.

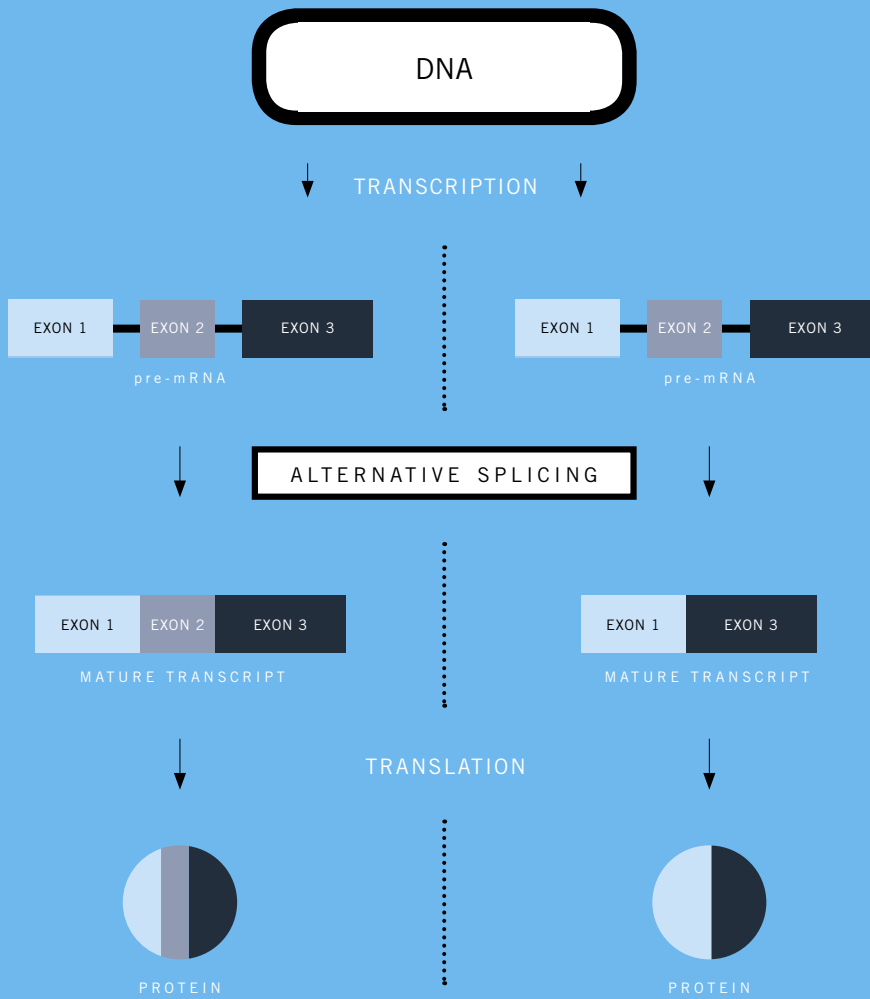


2. Douglas Black
Douglas Black researches the regulation of pre-mRNA splicing in differentiated cells, particularly neurons.



3. Lawrence Zipursky
Lawrence Zipursky is interested in uncovering the mechanisms by which neurons make highly specific patterns of connections during development.

SPlicing AND TRANSCRIPTION



FUNDAMENTALS

Two Types of Splicing

Fundamentally, alternative splicing is little different from routine, or constitutive, splicing.

Scientists have known since the 1970s that, in eukaryotic organisms, gene sequences that encode proteins are interrupted by segments of noncoding DNA known as introns. These regions, which can make up more than 90 percent of a gene, are transcribed into RNA along with the interspersed coding segments, or exons.

Because they cannot be used during the translation of RNA to protein, they must be removed.

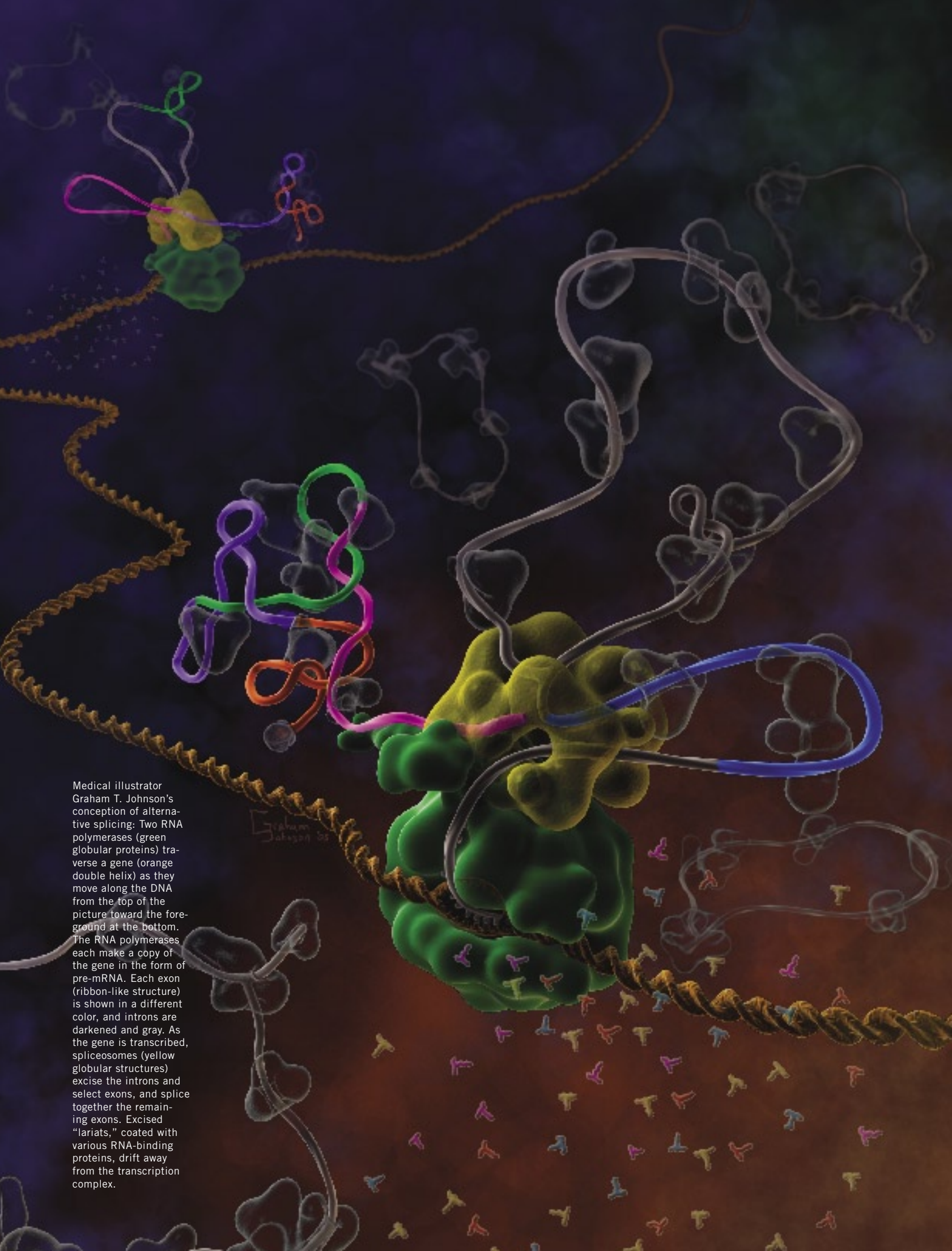
A complex of proteins and RNA known as the spliceosome is responsible for identifying splice sites on an RNA molecule, snipping out the appropriate segments, and reconnecting the severed transcript.

For some transcripts, this process is always the same: Introns are removed and the remaining exons are pieced back together. But for alternatively spliced transcripts, things are more complicated. A subset of exons is often removed along with the introns (the most prevalent form of alternative splicing in mammals); in other cases (most commonly in plants and lower animals), introns are retained in the final

transcript. The inclusion or exclusion of each segment determines the structure of the resulting protein, and whether to remove an alternatively spliced exon is a decision the spliceosome must make each time a gene is transcribed, with the assistance of a sizable collection of regulatory molecules.

To begin the conversion of genetic information into proteins, cells produce a precise RNA copy of a gene using the DNA sequence as a template. In all but the simplest organisms, these transcripts are not functional mRNAs until they undergo a molecular editing process, when the cell snips out unintelligible or unwanted bits of genetic sequence. This splicing process often occurs while an RNA molecule is still being assembled, and for alternatively spliced transcripts, how quickly the cell's transcription machinery pieces together the RNA can influence which bits are discarded.

- The spliceosome (the mass of proteins and RNA that controls genetic splicing) recognizes some splice sites more readily than others, ordinarily ignoring weaker sites in favor of stronger ones located nearby. But according to work by Alberto R. Kornblihtt, an HHMI international research scholar at the University of Buenos Aires, slowing down the transcription of certain genes can alter this selection process.
- The effect is seen in genes where a weak splice site is transcribed before a stronger one. Ordinarily, the spliceosome would opt for the stronger site, but when transcription slows, there is a delay during which the weak site has been transcribed, but the stronger one does not yet exist. A spliceosome working on an incomplete transcript has no choice but to use the weaker splice site and consequently produces a different protein than is generated when transcription proceeds more rapidly.
- Cells alter transcription rates to modulate the amount of proteins they produce. For genes whose splicing is coupled with transcription, this strategy may alter not just the quantity of those proteins but also some important aspect of their structure, Kornblihtt says. He also thinks this effect can explain why splicing is often coordinated between distant regions of the same gene. It's common for the inclusion of one exon to relate directly to the inclusion of another—which may mean that the splicing of each region depends on how quickly transcription proceeds. One transcription rate may mean that both exons are included, for example, while at a faster rate, both will be excluded.



Medical illustrator
Graham T. Johnson's
conception of alternative
splicing: Two RNA
polymerases (green
globular proteins) tra-
verse a gene (orange
double helix) as they
move along the DNA
from the top of the
picture toward the fore-
ground at the bottom.
The RNA polymerases
each make a copy of
the gene in the form of
pre-mRNA. Each exon
(ribbon-like structure)
is shown in a different
color, and introns are
darkened and gray.
As the gene is transcribed,
spliceosomes (yellow
globular structures)
excise the introns and
select exons, and splice
together the remain-
ing exons. Excised
"lariats," coated with
various RNA-binding
proteins, drift away
from the transcription
complex.

University of Connecticut Health Center. “In cases like *Dscam*, where you have 38,000 different proteins that are made from that gene, knowing just one of them is not terribly useful.”

Christopher Burge, an associate professor at the Massachusetts Institute of Technology, likens the state of the splicing field to an earlier phase in the ongoing quest to decipher the genetic code. “Back in the early sixties,” he says, “they didn’t know whether it was a triplet code, a doublet code, a tetramer code, or not of fixed length. And I think that’s where we are with splicing. We don’t yet know what the code looks like.”

To delve into that code, Burge has scoured human and mouse transcript databases—enormous collections of gene sequences generated from mRNAs—for features that distinguish alternatively spliced exons. The presence of a splice site, Burge says, is not sufficient to know that splicing will occur; nearby enhancer and repressor sequences—short segments of RNA that serve as landing pads for regulatory proteins—are equally crucial. The splicing of a single exon, he estimates, is likely promoted by at least three to seven enhancer sequences.

By analyzing the complete sequence information of genes known to be alternatively spliced in both mice and humans, Gene Yeo, a graduate student in Burge’s lab (now a fellow at the Salk Institute for Biological Studies), developed a profile of a “typical” alternatively spliced exon and later identified another 2,000 exons that fit the description. Of these, at least 70 percent appear to undergo alternative splicing in both humans and mice.

Burge has focused his analyses on those alternative splicing events that are conserved between species, because he believes they are most likely to be functionally important. “It’s a diverse collection of genes,” he says, “but there are significant biases in what kinds of genes undergo conserved alternative splicing.” These exons are more likely to be in genes that are involved in development, he says. “They’re more likely to be transcription factors. They’re more likely to be expressed in different brain regions.”

THE GOAL OF BEING ABLE TO PREDICT SPLICING FROM A GENE SEQUENCE IS STILL TANTALIZINGLY OUT OF REACH.

he says, “presumably both forms would produce stable, folded proteins that would have the same enzymatic activity. They would just differ in these regions that might affect other properties of the protein—maybe its localization or its regulation.” By altering proteins’ locations inside a cell or their sensitivity to activators or inhibitors, most conserved alternative splicing events, Burge expects, produce “subtle, but perhaps very important differences.”

While Burge explores the sequence clues to how transcripts are spliced, others are trying to unravel the convoluted system of regulatory proteins that contribute to splicing decisions. It’s an extensive network of positive and negative regulators, some ubiquitous and some expressed only in specific cells, and Black thinks that easily hundreds of proteins could be involved.

Connecticut’s Graveley, for example, focuses on the fruit fly *Dscam* gene, which he considers the perfect model for studying the intricate systems that regulate alternative splicing. Graveley saw Zipursky’s first paper about *Dscam* the day it was published in the journal *Cell* (June 9, 2000); thinking “this is too good to be true,” he started designing experiments that very day.

Eliminate the Proteins

SINCE THEN, Graveley and his colleagues have found 47 proteins that alter *Dscam* splicing, and they expect to turn up more. Their strategy has been to systematically eliminate each of the 250 or so proteins in the fruit fly that bind to RNA—good candidates for splicing regulators—and examine the effects of such knockouts. Some of the molecules they’ve identified in their screen are generic splicing factors, thought to be required for the splicing of many genes in a variety of cell types; others are novel, perhaps controlling a more defined set of spliced exons. Most of the factors regulate the inclusion of a single exon within *Dscam*.

CONTINUED ON PAGE 64

Burge’s group has also observed that alternatively spliced exons tend to fall between the segments of a gene that encode the functional units, or domains, of a protein. For most alternatively spliced transcripts,

SECRETS OF THE SPLICEOSOME

Cells invest significant resources in carrying out and regulating splicing—both constitutive and alternative. In mammals particularly, the machinery for identifying splice sites, snipping out introns, and reconnecting the severed transcript has long been known to be a large complex of regulatory and catalytic components, both protein and RNA. • Melissa J. Moore, an HHMI investigator at Brandeis University, says that not long ago, she estimated that 50 to 70 proteins participated in the splicing process. So it came as a surprise when, in 2002, her lab and others purified the spliceosome in various stages of the process and found about 100 proteins at each phase. The proteins known to come and go during splicing total about 300. • These studies, Moore says, focused on constitutively spliced exons, which are more efficiently recognized and removed than those that are alternatively spliced; many of the molecules regulating the latter process have probably been left out of recent structural models. “My guess is that the list of proteins affecting splicing is not complete yet. We haven’t even begun to scratch the surface of the proteins that affect alternative splicing,” she says. • Scientists are still debating how the components of the massive complex arrive at a splice site, but according to HHMI investigator Michael Rosbash, also at Brandeis, “It’s rather difficult to imagine how alternative splicing would take place if the spliceosome was preassembled.” While some researchers argue for a model in which the spliceosome is at least partially preassembled, recent studies from Rosbash’s lab and another have demonstrated that, at least in yeast, loading of the spliceosome occurs in a stepwise fashion. • Rosbash has found that the five small nuclear ribonuclear particles (snRNPs) at the core of the splicing complex load sequentially onto the transcript, presenting an opportunity for regulation. “A lot of alternative splicing regulation could take place at the level of which snRNPs jump on where, and in what order,” he notes. “It just gives you many more degrees of freedom.” Although there’s no evidence yet of this stepwise assembly in other organisms, “the core mechanisms are so similar that it would be shocking if something as fundamental as this were not conserved between humans and yeast,” Rosbash says.

“When we knock these out, they don’t seem to just globally screw up splicing—they’re very specific alternative splicing events, which was a little bit surprising,” Graveley says. “So we definitely know that there are some exons that are highly regulated.”

And this, Zipursky says, is a system that is in some ways random. Since his initial discovery of the gene, Zipursky has found that *Dscam*’s role is to enable the extending processes of a neuron’s dendrites and axons to distinguish between themselves and their neighbors—and he thinks that by producing so many forms of the protein, the fruit fly has actually minimized the need for elaborate control.

Each neuron expresses 10 to 50 forms of *Dscam* at any given time, and Zipursky’s group has shown that each form has unique recognition properties. What is important, he says, is that cells express forms of *Dscam* that are different from one another. With more than 38,000 to choose from, odds are—even if they’re randomly selected—one cell’s *Dscam* forms won’t be the same as those expressed by a neighboring cell.

“The fly has invested in making many, many different types of forms,” Zipursky concludes, “but it hasn’t invested in the detailed control that would be necessary to make specific forms in specific neurons. That would require a tremendous investment in genetic regulation.”

But in mammals, Black says, “the regulation gets very complicated.” His lab has been teasing out the precise ways in which specific splicing factors, and the sequences they bind to, orchestrate that regulation. “For most exons that have been analyzed, what we’ve found is there are multiple regulators,” he says. Some of these are so globally expressed that additional regulators must be around simply to counteract them when they’re not needed. In some cases, a single molecule can activate splicing in one cell type while repressing it in another. The finely tuned balance of these splicing factors determines how each molecule behaves and how a transcript is ultimately spliced. “There’s this combinatorial system of regulation similar to transcriptional regulation,” Black says.

The average mammalian gene has eight or nine exons, and with most human genes undergoing some form of alternative splicing, virtually all of these are candidates for elaborate control. Most of the molecules and their interactions remain to be elucidated. In that light, the goal of being able to predict splicing from a gene sequence seems attainable, but that knowledge is still tantalizingly out of reach. “We’re not very good at predicting what splicing patterns are going to be, let alone how those splicing patterns will be regulated,” Black says. “People are certainly making progress, but we don’t know enough about the mechanism.” ■