

); } //-->

How Science Is Rewriting the Book on Genes

By David Brown Washington Post Staff Writer Monday, November 12, 2007; A08

Everyone who goes to medical school hears this story at some point.

Graduation day comes and the new doctors assemble to get their diplomas. The dean gazes out and announces sheepishly: "I'm sorry to tell you that half of what we taught you is wrong. The problem is, we don't know which half."

Nowhere has this been more evident than in genetics.

The rules of inheritance, and hints of the biological mechanisms behind them, were first elucidated by Gregor Mendel in the 1860s. Over the ensuing 130 years, scientists gained insight at a molecular level into how biological information is recorded, preserved, used and passed on to future generations.

In recent years, however, many of the certainties gained over that long run are being overturned.

Ever better tools for cutting and splicing strands of DNA, combined with the explosion of data from sequencing the genomes -- the inherited genetic instruction books -- of humans and more than a dozen other organisms, are the engines driving this revolution.

Here are some of the recent revisions of what were once canonical rules of genetics.

What Is a Gene?

DNA, or deoxyribonucleic acid, is a molecule made of four chemical compounds, called nucleotides, strung end to end like beads on a string. The nucleotides are designated by the letters A, T, C and G. In fact, they function like letters, carrying instructions that are spelled out in three-letter "words" -- ATG, CAC, etc. -- never two letters or four letters or five letters. The human genome consists of 3.1 billion nucleotide pairs, strung together in a particular order.

Classically, a gene is a stretch of DNA within a cell that contains the instructions for making a protein. A gene has a section of DNA that marks its starting point, a body consisting of coding sequences (called exons) interspersed with filler areas (called introns), and a termination point. The three-letter words in the coding sequences specify the order of the different amino acids that will make up a protein: the gene's product.

When a gene is activated, it is first transcribed into an intermediate molecule called mRNA. The introns are clipped out and the exons spliced together, and the whole thing is then translated into the protein.

Biologists used to think one gene produced one protein. Now it's clear that one gene can produce many different proteins. Under certain conditions, a cell clips out not only the intron fillers but also one or more of the exons. This is like taking a speech and removing many of the sentences. Done in different ways, it can produce many different messages.

"Alternative splicing represents the most significant driving force to diversify the output of a finite genome," says Hal Dietz, a [Johns Hopkins University](#) physician and genetics researcher.

This helps explain a mystery that emerged after researchers completed the human genome in 2001. They found 22,000 genes, but there are more than 100,000 proteins. How could that be?

The discovery of a whole world of post-transcriptional control of a gene's instructions, which allows one gene to provide instructions for more than one protein, is a big part of the answer.

How Does Evolution Change Genes?

About 5 percent of the total human genome appears to be a "message" of one sort or another, the order of the nucleotide letters determined over the course of evolution and passed through the generations with few changes. In those stretches of DNA (there are about 500,000 of them salted through our 23 pairs of chromosomes), any addition, deletion or change of a letter is likely to have a big effect. Often the effect is death.

The function of the remaining 95 percent of the DNA chain is uncertain. Changes creep in and pile up without obvious effect. The whole mass used to be called "junk DNA," but few biologists are willing to call it that anymore.

It was long thought that the active 5 percent of our DNA consisted almost entirely of genes coding the instructions for making proteins. But it turns out that's not true.

It's now clear that more of those evolutionarily preserved stretches of DNA don't code for proteins than those that do. By one estimate, 70 percent of the conserved elements are non-coding.

"The majority of what evolution cared about is stuff we didn't know about a few years ago," says Eric S. Lander, a geneticist and head of the Broad Institute of MIT.

So what are these conserved non-coding elements? They are molecules worthy of the "Star Wars" cantina scene -- insulators, micro-RNAs, exon-splicing enhancers, 3'-untranslated hairpins and other weird characters only now emerging from the shadows.

What they have in common, other than that they are never translated into proteins, is that they regulate the activity of genes that do carry instructions to make proteins. They turn them on and off, tweak them to make one version of a protein rather than another, increase or decrease the efficiency of production, and coordinate the sequential or simultaneous action of genes.

A study in the journal *Science* in August found that these elements are less tolerant of mutation than protein-coding genes. That means they are more likely to be identical among people, mice, fruit flies and worms than are the genes coding for proteins. It turns out that more of evolution's survival-of-the-fittest battles occurred in writing the instruction manual for running the genes than in designing the genes themselves.

It also turns out that a huge number of proteins encoded by the classical genes are used to regulate other genes, and not to be final products such as enzymes, cartilage and hemoglobin. Nearly 1 in 10 code for transcription factors, proteins that help turn genes on and off at the right time.

In practice, many of these transcription factors act like orchestra conductors. One of them, called HIF-1, picks up the baton when a cell's demand for oxygen rises or the supply falls. Depending on the length or severity of the oxygen shortage, HIF-1 can turn on genes that make more red blood cells and blood vessels, or change the cell's metabolism so it needs less oxygen.

This kind of regulation may be what separates mice from men -- or, more likely, what separates mice and men from yeast and flies. Nine percent of all human genes code for transcription factors. For fruit flies, it's 5 percent; for yeast, 3 percent.

Is Efficiency the Goal?

It used to be a rule -- actually, more of an assumption -- that the genetic machinery of living organisms was never intentionally wasteful or inaccurate. It turns out this isn't always true, either.

Three recent discoveries show that what appear to be genetic mistakes can have a purpose.

Some genes, for example, contain mutations that are thought to be "silent" and harmless because the substituted letter is synonymous with the correct letter. It's a little like spelling "skeptical" with a "c" instead of a "k" -- "sceptical" -- the meaning is still the same.

Earlier this year, however, a study in *Science* showed that these synonymous spellings can make a difference. That's because it can be harder to make a protein from the instruction with the unusual, but synonymous, spelling. The construction process takes longer, and the final protein folds up differently. It has a slightly different shape -- and a different function.

A second example involves the gene for the blood-clotting protein called prothrombin. In its most common form, this gene contains a stretch of nucleotides near one end that makes it hard for the cell to

translate the mRNA transcript into the prothrombin protein.

Some people have a mutation that fixes this problem. Their prothrombin gene makes a transcript that pours out prothrombin. The result? Their blood clots too easily. They get clots when they don't need them.

Why would evolution favor this built-in inefficiency?

One theory is that in certain circumstances -- perhaps life-threatening bleeding -- the body can throw a switch that somehow overcomes this natural inefficiency and increases clotting at a moment's notice.

A similar goal may be behind the third example, a phenomenon called nonsense-mediated decay.

When a cell is translating a piece of mRNA into a protein, it stops when it reaches a three-nucleotide segment called a termination codon that causes the just-built protein chain to separate from the construction machinery -- like a car rolling off the assembly line. It now turns out that the mRNA instructions for a few proteins have a termination codon in the middle of the chain, not just at the end.

The protein-building machinery can sometimes "read through" this instruction and complete the protein. (How it does that is complicated and only partially understood). But most of the time, construction just stops. The half-finished protein is broken down, and all the effort goes for naught.

Why would an organism have an assembly line that builds a product partway and then tears it up time after time? This may be another strategy that allows the body to handle certain problems at a moment's notice.

Think of a cell as containing a factory that makes both tractors and tanks. In peacetime, few tanks are made, but the knowledge and capacity is never lost. Most tanks are built halfway and then broken down, with the parts sent back up the assembly line for reuse.

But then comes great stress; say the cell is experiencing too much heat or not enough oxygen or food. That's where nonsense-mediated decay (NMD) comes into play. It's suddenly wartime, but instead of refitting the factory to make tanks, all the cell has to do is give the order to take the half-made tanks to completion.

For example, when an organism is well-fed, it doesn't need a lot of machinery dedicated to capturing and transporting amino acids, the building blocks of proteins. That's because there is an abundance of amino acids floating around and within easy reach.

In starvation times, however, the cell needs all the amino-acid-capturing machinery it can get. So it turns down NMD, and the proteins whose functions are to capture and carry amino acids start rolling off the assembly line in abundance.

Just the right amount of wrong instructions and wasteful habits -- that's what evolution has built into all of us.

Post a Comment

[View all comments](#) that have been posted about this article.

Comments that include profanity or personal attacks or other inappropriate comments or material will be removed from the site. Additionally, entries that are unsigned or contain "signatures" by someone other than the actual author will be removed. Finally, we will take steps to block users who violate any of our posting standards, terms of use or privacy policies or any other policies governing this site. Please review the [full rules](#) governing commentaries and discussions. You are fully responsible for the content that you post.

© 2007 The Washington Post Company