

Master & commander

New Scientist vol 184 issue 2475 - 27 November 2004, page 36

In the land where DNA is king, is a lowly servant pulling the strings? Philip Cohen exposes the revolution inside your cells

LIFE as we know it is about to be transformed. Until recently, we have been pretty confident we understood the cast of molecular characters that rule life on Earth. Now it seems that a veritable army has simply evaded detection.

The roots of this once invisible throng pre-date even the iconic double helix of DNA, and its members could be the most versatile entities in our cells. This discovery, still hotly contested by many biologists, promises to overturn much of what we thought we knew - and could even force us to rethink what genes really are. In spite of our sophisticated genome sequencing projects, it seems we have missed at least half the story of how life works.

"Right now only a few people are aware of the rolling thunder in the distance," says John Mattick of the University of Queensland in Brisbane, Australia. "But it will soon be clear to everyone what's happening when they get hit by a tsunami of new data."

According to the status quo, proteins are the centre of the biological universe. Proteins and the machines they build have long been considered the prime movers and shakers that keep life's gears spinning: they copy our chromosomes, control genes, digest our food and hold cells together, among myriad vital tasks. DNA's big role in all this is to serve as the blueprint for proteins. So it is no surprise that modern biology's most ambitious undertaking, the human genome project, was largely focused on uncovering the 35,000 or so genes that encode human proteins.

But Mattick and a few other researchers are boldly predicting that an unlikely challenger may soon join proteins on their lofty pedestal: RNA, DNA's less stable chemical cousin. Unlike DNA, which normally exists as a double helix of two intertwined molecules, RNA usually comes as a single strand of genetic letters and can fold into a range of complex shapes. It has long been known to perform several lowly tasks within the cell, the best known being contained in the formal definition of a gene: any segment of DNA that is transcribed into a "messenger RNA", a template the cell reads to make the protein coded for by that gene.

But recent discoveries are calling this job description into question. The findings hint that for all the genes that encode proteins, an equal number - and perhaps many times more - don't produce proteins at all. Rather, a surprisingly large chunk of our genome appears dedicated to the manufacture of pieces of RNA for their own sake. What the vast majority of these "non-coding" or ncRNA molecules do for the cell is still up for grabs. However, RNA is now constantly revealing powers that go far beyond its traditional role as data courier between genes and proteins. It could be that RNA's role in the cell is equal to, or even greater than, that of proteins.

The work has huge implications for researchers who hunt down molecules that might underpin diseases and, crucially, look for ways to target those molecules to treat disease. "We live in a world where protein-coding genes are the most important aspect genomes have to offer, and that's where biologists and big pharma have looked for therapeutic targets," says Tom Gingeras of genomics firm Affymetrix in Santa Clara, California, whose team discovered the new hoards of human ncRNAs. "But if this is right, there are many, many new RNA targets to be used. That's why even though this is under the radar for most people, the pharma people are already quite interested."

But if RNA's role in biology is really so overwhelming, how could it have been underestimated for so long? RNA enthusiasts maintain there has always been a strange bias against their favourite molecule. "Every time you find an interesting RNA a biologist's brain is trained to think it is a freak, an outlier," says Larry Gold, an RNA biochemist at biotech company SomaLogic in Boulder, Colorado.

Perhaps for that reason, RNA has traditionally attracted fewer researchers and research dollars than DNA and protein. And it doesn't help that RNA is harder to work on than DNA, being less stable. Proteins also seem better equipped for complex roles in the cell, being built up from 20 different amino acid units that offer a wide range of chemical powers. In contrast, DNA and RNA, each just a backbone studded with four simple letters or nucleotides, seem suited for information storage rather than chemical wizardry.

But in the 1980s people's perception of RNA's abilities began to shift. Sidney Altman at Yale University and Tom Cech at the University of Colorado, Boulder, shared a 1989 Nobel prize for years of work demonstrating that two natural RNAs could act as catalysts and accelerate chemical reactions - a task only protein enzymes were thought up to.

Ironically though, this astonishing discovery didn't convince biologists that RNA's role in the cell had been underestimated. Rather, they were inclined to write it off as an oddity of our evolutionary past. That's because the discovery suddenly suggested an answer to an intriguing problem: how life began. The fact that RNA could both store information and catalyse the reactions needed to copy itself became the basis of a very attractive hypothesis called the "RNA world" - the idea that life got started when RNA or a molecule much like it emerged and learned to replicate itself.

It seems we have somehow missed at least half the story of how life works

But, the story goes, RNA's reign ended when DNA and proteins evolved. DNA proved to be more reliable genetic material: its greater chemical stability allowed it to store and transmit information more effectively. And proteins, with their greater suite of building-blocks and wider range of chemical abilities, ended RNA's rule as biochemist extraordinaire.

So for many biologists, catalytic RNAs were not proof of RNA's unsung role in biology, they were simply dinosaurs whose more interesting jobs had been outsourced to better molecules. "To most people, the RNA world means that's how life got started," says Stephen Holbrook of the Lawrence Berkeley National Laboratory in Berkeley, California. "But it's supposed to be over with. We aren't supposed to still be living in it."

But now it seems the RNA world has been with us all along. The molecule's repertoire of abilities has surprised biologists again and again ([see Table](#)), and in just the past couple of years, researchers have stumbled across many more examples, including as-yet unnamed vertebrate RNAs that can enhance the replication of infectious prion particles (*Nature*, vol 425, p 717) and others that help cells form connections in tissues (*Cell*, vol 117, p 649).

Ancient defender

One aspect of RNA biology called RNA interference, or RNAi, ranks among the hottest topics in medicine ([New Scientist, 14 September 2002, p 28](#)). RNAi is an ancient cellular immune system that uses small pieces of RNA to target and silence protein-coding genes. It was discovered in plant cells, but was recently found to be at work in human cells too. As much as 1 per cent of human genes may encode small RNAs involved in this process. Companies are now scrambling to harness RNAi to shut down viruses and regulate genes in patients - a major

goal of many therapies.

Sean Eddy, a molecular biologist at Washington University School of Medicine in Saint Louis, Missouri, also points out that no one has found any hint of RNAi in bacteria, suggesting it is a relatively new function that evolved in advanced cells. "The RNA world story says RNA was only the way life got started, but then RNA's role got smaller and smaller," he says. "But these are all evolutionary fairy stories. RNA is a perfect molecule whose powers were recognised again and again by evolution."

Yet even with the excitement over RNAi and these other discoveries, non-coding RNAs were still only thought to account for a few hundred genes in the human genome - a drop in the bucket compared with tens of thousands of protein-coding genes. Protein's role as ruler of the cell was hardly under assault.

The first hints of the coming revolution arose a few years ago, when Gingeras, like any good genomic scientist at the time, began using a novel approach to hunt for protein-coding genes. Hundreds of scientists were already putting protein-sniffing computer programs to work on the newly acquired human genome sequence, or using biochemical techniques to pull messenger RNAs from cells and look for sequences that seem to encode proteins. Gingeras decided to conduct a comprehensive search. He isolated pieces of DNA from each section of chromosomes 21 and 22. Then, using gene chips, he tested whether the cell produced a corresponding RNA. "This worked in bacteria, but people warned me that the human genome was too large and too complex for our technology. All we would see was noise," he says. And at first, they seemed to be right. Even far outside any known protein-encoding gene, Gingeras's team found evidence for RNAs. "I looked at this data and thought, well, god, this is going to be really really complicated."

But after ruling out technological glitches, Gingeras's team concluded these regions really were producing RNAs. And since chromosomes 21 and 22 are the two smallest and best studied human chromosomes, the team was confused about why no one had ever mentioned these RNAs in their descriptions of the chromosomes. A few phone calls to various colleagues revealed why. "People told us they *had* seen them," says Gingeras. "But they never included them in the chromosome map because the RNAs obviously didn't code for protein."

To be fair, researchers had a good excuse for ignoring these unexpected RNAs: they had no known function. And there was a perfectly good explanation for their existence. The RNAs might simply have been the result of random mistakes, where the molecular machines that produce messenger RNA simply overshot their target or started in the wrong place.

Nonetheless Gingeras's team decided to study them more closely, and their surprising findings were published this year (*Cell*, vol 116, p 499). This time, rather than look for the RNA directly, they looked for another physical hallmark of bona fide genes, proteins called transcription factors. These generally attach to DNA strands at the start of genes and help initiate the process of making the messenger RNA copy. The researchers isolated sections of DNA that were tagged by three transcription factors, known as cMyc, p53 and Sp1, and then traced the locations of those DNA regions on chromosomes 21 and 22.

Surprisingly, only about a quarter of these transcription factors were where they were supposed to be, near the start of known genes. More than a third were inside or at the end of genes and, even more dramatically, another quarter sat completely outside any known gene. In most of these cases, the researchers were able to find an ncRNA that appeared to be produced by that region.

Death to the dogma

Some of these ncRNAs appear in the mouse genome as well, suggesting they perform a function important enough to be preserved over millions of years of evolution. What's more, some of them seem to be switched on and off during embryonic development by the same molecular signal that controls the activity of certain protein-coding genes (*Cell*, vol 116, p 499). Gingeras says his team has now discovered similar densities of ncRNAs on 10 other human chromosomes.

Eddy describes the new data as impressive. Still, while he thinks Gingeras has proven these ncRNAs aren't simple accidents, he is not yet willing to accept the idea that every ncRNA - or even the majority of those that Gingeras detects - have important functions. For instance, the mechanics of transcribing DNA into RNAs may make it very difficult for the cell to avoid producing redundant transcripts such as these ncRNAs. Transcription factors can sometimes stick to DNA sequences that look similar to the target site, so unintended binding sites could arise pretty much randomly. And if the resulting RNAs were non-coding, there wouldn't be any pressure to shut down these production sites.

Even so, Eddy admits that the sheer number of ncRNAs Gingeras has discovered is remarkable. He concedes that a small number of them could have a function, pointing out that natural selection would be sure to act on such a vast palette of "useless" RNAs. "I can't prove it, but I believe nature is smarter than that," says Holbrook. "When we find a protein-coding gene that is this controlled, that's what we assume. I don't see a compelling case to say that should be different for RNA." Peter Good, programme director for the Encyclopedia of DNA Elements (ENCODE) project at the National Human Genome Research Institute, part of the US National Institutes of Health, agrees that the jury is still out. But he also points out that with the discovery that humans only have slightly more protein-coding genes than microscopic worms, scientists are wondering if other kinds of genome information accounts for the greater complexity of human biology. "It suggests the genome isn't as simple as we thought it was," he says.

Gingeras knows that convincing the naysayers will be an uphill battle. "When you challenge a field in a fundamental way, the bar is raised in terms of the proof you need." The bar, in this case, is to find functions for as many of those RNAs as he can, as quickly as possible. It's a daunting task. To begin with, he is looking at similar ncRNAs in the fruit fly *Drosophila melanogaster*, where genetic tests can quickly determine whether a segment of DNA is important or not.

Until more evidence emerges from these and other labs, whether RNA really rules the roost remains an open question. But the new insights have already reshaped something as profound as the DNA-RNA-protein definition of the gene, which has dominated biology for the past half-century. The late Francis Crick, co-discoverer of DNA's double helix structure, even famously called this definition "the central dogma of molecular biology".

Of course, we have known about the existence of a few non-coding RNAs for decades. But Eddy thinks the discovery that our cells may be churning out just as many or more non-coding RNAs than protein-coding ones is death to the dogma. "If you have this mass of RNA coming from a zillion different parts of the genome, do we call these genes or are they something else? The language we have starts to fail us."

Philip Cohen