

BATTLEFIELD GENOME



SCIENTISTS STILL ARGUE WHETHER MOST OF OUR DNA HAS A PURPOSE.
BY **DYANI LEWIS**

A little over a year ago it looked like Australian geneticist John Mattick had won a bet against his English colleague, Ewan Birney, over the way the human genome works. Like many others, Birney maintained that our genome was mostly comprised of “junk”, excess DNA that padded it out. Mattick, director of Sydney’s Garvan Institute, had long believed otherwise. In his view, so-called junk DNA would prove to be a code, our genome’s equivalent of a high-level operating system. In 2007 the two made a bet that at least 20% of the “junk” would be found to have a function. The stakes were a case of



Australian geneticist John Mattick, main picture, bet his English colleague Ewan Birney, above, that ‘junk DNA’ was our genome’s operating system. He is still to collect.

good Australian red. It was a well-timed wager. A worldwide project known as ENCODE was gearing up to examine the output of every one of the three billion letters of DNA that comprise the human genome. The results were announced in September 2012 with great fanfare. At a worldwide media conference, Birney declared that 80% of our DNA code was “functional”. Sometime, somewhere, one cell or another in the body was reading almost every bit of the genome.

So can we call it quits on the debate over junk DNA? Far from it. As critics were quick to point out, simply reading out the DNA code is not proof that the code is functional. It might just be the cells’ equivalent of web surfing: a lot of useless sites get perused before anything useful is found. Mattick’s case of wine

suddenly wasn’t looking quite such a sure thing.

How to settle this argument? One way to decide whether junk DNA is useful would be to get rid of it and see what happens. Not an experiment you can do on people. But last year, Victor Albert at the State University of New York in Buffalo reported that nature might have done the experiment for us.

Like us, the genomes of plants, insects and other animals also consist of vast amounts of DNA, much of which we can’t decipher. Albert claimed he had found a carnivorous plant, the bladderwort, which has a virtually junk-free genome and does just fine. Could the debate soon be settled?

The term “junk DNA” was originally coined in 1972 by Japanese American evolutionary biologist Susumo Ohno.

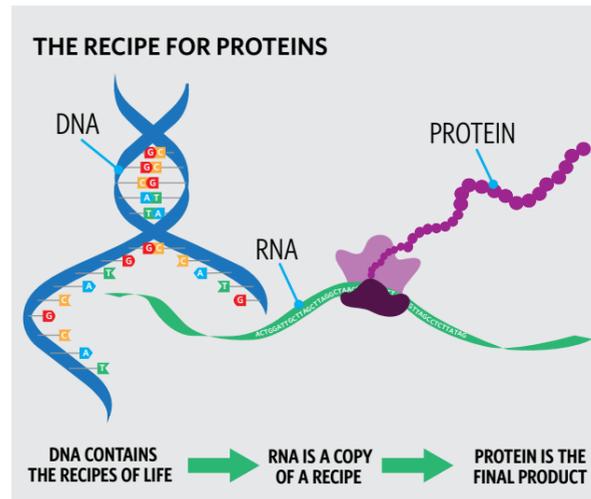
It's easy to forget how little was known about genomes just four decades ago. In 1972 scientists could only speculate about what a whole genome might look like – how a four-letter DNA code of As, Ts, Gs and Cs might be strung together to write an instruction manual. But even without reading it, scientists knew that ours was big. The way Ohno saw it in the early 1970s, with a genome the size of ours, only a small percentage could possibly be made up of genes or we would suffer dangerous mutations that would quickly accrue over the generations.

For decades, scientists focused on genes and ignored the junk.

As many early geneticists found, if you mutate a gene, important developmental processes could be disrupted. At the time, a gene was thought of as a recipe for a protein. Proteins are the construction-site workers charged with turning the information in a one-dimensional DNA code into a living organism. They do it all, forming the bricks and mortar of our cells, the enzymes that drive our metabolism and the components of cell communications systems. But junk DNA could not be deciphered into any protein and the term became shorthand for any stretch of DNA that was not a protein-coding gene.

Almost immediately the term seemed doomed. It was imprecise, and ignored growing evidence that some DNA sequences had other essential biological functions. For instance, researchers in the 1960s had already found that small tracts of DNA, known as “promoters”, lie directly ahead of protein-coding genes and act as helipads – landing sites for enzymes that read genes. These enzymes “transcribe” stretches of the DNA code into an almost identical stringy molecule called RNA.

During the 1980s and 1990s, scientists managed to decipher even more novel functions for junk DNA. Other types of helipads, called “enhancers”, were identified, often located thousands of letters away from



++ WE HAVE 20,000 PROTEIN-CODING GENES, THE SAME NUMBER AS THE ROUNDWORM.

John Rasko of the University of Sydney, above, hates the term 'junk DNA'. He believes it holds the key to how complex an organism is.

the gene they controlled. Yet other stretches of DNA carried instructions not for protein recipes, but for RNA recipes alone.

Like a photocopy of a page from a recipe book, RNA was thought to be produced *only* for the purpose of instructing protein synthesis (see figure above). But it turns out that each transcribed RNA molecule could have a function. Some of these functional RNA molecules, dubbed “ribozymes”, work like enzymes to catalyse cellular reactions. Others, known as “microRNAs”, interfere with the RNA copies of other genes, effectively switching them off by preventing proteins from being made from the RNA recipe.

Although these discoveries were momentous, they did not blow away the concept of junk. When the Human

Genome Project finally unveiled its completed 3.2 billion letters of genetic code in 2003, the mystery of our undeciphered genome hit prime time. The idea that only 1.5% of our DNA coded for genes seemed to fire the public imagination

Our total number of genes was also humiliating. Ours was not the first genome to be unveiled: a microbe, a roundworm and a fruit fly all preceded us and revealed gene numbers ranging from 4,000 to 20,000. Surely our vastly more complex species would have at least an order of magnitude more.

Not so. It turns out we have 20,000 protein-coding genes, the same number as the roundworm, a one millimetre long transparent creature boasting just 1,000 cells.

“It was a great shock to everyone,” says University of Sydney haematologist John Rasko. Perhaps what set us apart from simpler organisms lay not in the genes, but in the 98.5% of our DNA still waiting to be decoded – the view firmly held by Mattick. He believes that the complexity of an organism does not relate to the number of genes, but to what’s in the junk DNA. Indeed there is a modest correlation between an organism’s complexity and the amount of junk DNA it carries: the bacterium *E. coli* contains little more than 10% non-protein coding DNA; roundworms 75%; for humans it’s 98.5%.

Rasko hates the term “junk DNA”. “It still riles a lot of people in the field that the term ‘junk’ even took up traction,” he says. It’s not surprising that he is unimpressed with the phrase. Rasko’s “current obsession” is introns, the sort of DNA sequences Ohno would have dismissed as junk. Introns, as their name hints, are found interspersed within protein-coding genes and range in size from 10 to thousands of letters long. When a protein is made, the gene is first transcribed into an RNA copy with introns intact. But before the RNA molecule is finally

translated into protein, the introns are edited out. Should that editing fail, the RNA molecule bearing an intact intron is sent to what Rasko calls “the molecular trash can” (see figure below).

Rasko and his team have found that during the development of white blood cells, many RNA molecules actually hang on to their introns; a perplexing observation since these transcripts are made only to be trashed. “Why would a cell go to all of that trouble?” asks Rasko.

The answer, he says, is “complexity”. Just as in the performance of a symphony orchestra, each instrument must play or be silent at precisely the right time, so too in the development of cells. Particular proteins need to be turned on and off at the stroke of a baton. By making transcripts that are destined for the shredder, Rasko believes that the genome has come up with “an elegant system” for orchestrating protein levels during the development of white blood cells. What’s more, entire suites of proteins can be orchestrated using the same molecular baton.

Rasko identified 86 genes involved in white blood cell development that were all diminished in concert. And it turns out shredding the RNA instructions, rather than making unnecessary proteins, is much easier on the cell’s energy budget. “The energy costs on a cell by controlling the editing of introns are tens-fold less than it would be if you had to use a protein degradation mechanism,” he says.

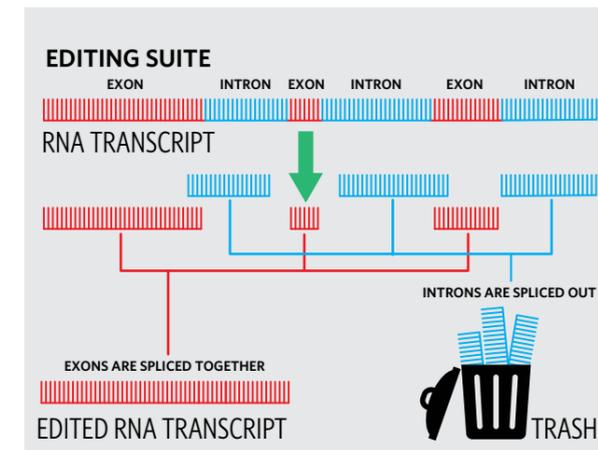
Introns are just one example of DNA sequences once viewed as superfluous, but now thought to be critical to the development of a complex organism such as a human. Disrupt intron editing and, as Rasko found, you disrupt the entire symphony. White blood cells unable to wield the baton failed to develop into the cells of the immune system. Rasko’s work illustrates how a once-overlooked component of the genome can turn out to be vital. The question is, how many other parts of the genome, once dubbed junk,

are essential? That’s where ENCODE comes in. A small army of researchers joined forces in the wake of the Human Genome Project’s completion in 2003 to systematically sift through the vast tracts of mystery DNA. The purpose was to find which bits have a biological function.

The massive international undertaking aimed to create the Encyclopaedia of DNA Elements (ENCODE’s full name) and brought together 442 scientists from around the globe. In September 2012, in an event that typifies the coordination required of such an immense project, their initial results were unveiled in a clutch of 30 scientific papers simultaneously published in three different scientific journals.

The bottom line, as Birney – ENCODE’s lead analysis coordinator – announced to the media, is that 80% of the genome has a “biochemical function”. To arrive at this estimate, 147 types of cells were subjected to 24 different experiments to search for meaning in the oceans of DNA. What was surprising was the number of potentially useful sequences dotted throughout the genome. Instead of an immense ocean of junk DNA punctuated with occasional islands of protein-coding genes, the genome began to look like a thick soup, packed with active ingredients.

Promoters and enhancers were known to be important residents of



Plant geneticist Jeffrey Bennetzen, above, believes most DNA is useless.

Introns are trashed, below. Rasko finds unspliced RNA also ends up in the trash.

the mysterious non-coding DNA. But ENCODE found more than four million of them, many more than had previously been recognised. Combined with the 1.5% of protein-coding DNA, that takes the proportion of our genome with known function up to around 10%.

ENCODE then measured other hints of function by looking at where proteins dock on to the long strands of DNA, finding three million of these sites. But the vast majority of “function” was inferred from the fact that in some cell somewhere in the body, at some time, DNA was being read, that is, transcribed into RNA.

The ENCODE fanfare was answered with a storm of criticism. A “meaningless measure of functional significance”, tweeted Michael Eisen from the US Howard Hughes Medical Institute. The definition of “function” was “so loose as to be all but meaningless”, opined T. Ryan Gregory from the University of Guelph in Canada. The conclusions were “absurd” and full of “logical and methodological transgressions”, wrote Dan Graur from the University of Houston. Jeffrey Bennetzen, a plant geneticist from the University of Georgia, summarised the feeling: “I don’t think there’s anybody who believes that because something is transcribed, that means it has a function.”

Mattick, who was involved in the pilot phase of ENCODE, disagrees. “I personally think it’s intellectually lazy to say it’s noisy transcription.” If it were noisy transcription, he says, then ENCODE would have seen random patterns of transcription. Instead it found precisely orchestrated patterns, tuned to particular cell types. Mattick believes that while gene number does not relate to complexity, those orchestrations of RNA transcribed from “junk” DNA do. As analysis of ENCODE continues, he predicts that the percentage of the human genome with proven function will edge towards 100%.

For Magdalena Skipper, the editor at *Nature* who shepherded the

publication of ENCODE's *Nature* papers, arguments over the numbers are missing the point. "The value of ENCODE goes so much beyond this discussion of what is the percentage of the genome that is functional and in what way we define function."

No doubt. But we still want to know what most of our DNA is really doing. The answer might come from an unexpected place.

The floating bladderwort is an unassuming carnivorous pondweed that captures its prey using tiny suction traps that lie beneath the water. But it wasn't the bladderwort's appearance or eating habits that intrigued evolutionary biologist Victor Albert. "It was known to have a tiny genome," he says. "The question was, what's missing?"

Albert and his colleagues found that the bladderwort genome contains a meagre 82 million letters. That's 1/40 the size of our own, and an even punier 1/240 that of its plant relative, the Norway Spruce. But size was only half the story.

"There's essentially no junk DNA," Albert says. The tiny genome contains around 28,500 protein-coding genes, but only 3% is what he would consider junk. "It's an interesting counterpoint to the human genome situation."

Some have suggested that the bladderwort may have rid itself of excess DNA to save on phosphorous, an element that is part of the DNA molecule. Bladderworts live in an environment that is poor in phosphorous, and eat meat to bolster their intake of the element. (Albert himself doesn't buy this explanation as to why they ditched their junk, since other phosphate-hungry carnivorous plants don't have tiny genomes). So if the bladderwort can do all sorts of complex things without its excess genomic baggage, does it follow that junk DNA is irrelevant? Not necessarily. By "junk", Albert was restricting his definition to a particular class of junk DNA known as "transposons" – repeating tracts that

are relics of ancient viruses.

And indeed the bladderwort seems to have dispensed with them. But as Mattick points out, even the minimalist bladderwort genome contains plenty of other non-protein coding sequences in the form of introns and tracts between genes that were traditionally termed junk – by his calculation some 65% of its genome. So, says Mattick, rather than spelling the death knell to junk, the bladderwort actually bolsters the view that no genome can truly go without.

For Mattick, the bladderwort's claim is just a replay of the claims made for the fugu, the highly poisonous Japanese puffer fish. For geneticists, it's best known for having the tiniest genome of any back-boned animal, one-eighth the size of ours. When its genome was first read in 2002 it was similarly billed as a complex creature that had managed to do away with its "junk" DNA. But as Mattick points out, in fact 89% of fugu's DNA does not code for proteins. So bladderworts and fugu still have a very high proportion of non-coding DNA, comparable to that of other complex organisms.



The bladder trap, below, is an unusual structure that gives the carnivorous pondweed bladderwort, left, its name. But a more surprising feature is the plant's shrunken genome.



As for the transposons, the bits of old virus that seem to multiply in genomes, Mattick concedes that they could be padding the genomes of some plants. "But you don't see nearly so much in animals," he says, possibly because they are under greater evolutionary pressure than plants to streamline their genomes, keeping sequences that are useful, and jettisoning the rest.

While no one argues that all non-protein coding DNA lacks function, the question now is how much is, in fact, junk? As Dan Graur cautions, when it comes to thinking about genomes, it's a mistake to think in terms of a "Goldilocks genome" where every bit of DNA is perfectly fit for its function. "Evolution never breeds perfection," he says. But even if a stretch of DNA is not perfectly functional, having some junk DNA to tinker with could be a big plus. As Mattick points out, bacteria with little "junk" have stayed stuck in the single-celled world whereas those with junk-laden genomes have formed the kingdoms of plants, animals and fungi. Perhaps genomes hang on to junk to allow the flexibility to evolve new and complex traits.

But that loose association between junk DNA and complexity still doesn't wash with many biologists. Until the function of the various sequences is demonstrated, biologists such as Albert, Bennetzen and Graur say that we are a long way from relegating the term "junk DNA" to the history books.

Scientists such as Mattick and Rasko continue to pore over the "functional" DNA identified by ENCODE. But how much of the genome will eventually pass muster for the tougher critics is still open to wager. As geneticist Daniel MacArthur at Harvard University's Broad Institute has declared, "I'd still take on Mattick's wager any day, so long as I got to specify clearly what was meant by 'functional'." 📌

Dyani Lewis is a freelance science journalist based in Melbourne, Australia.