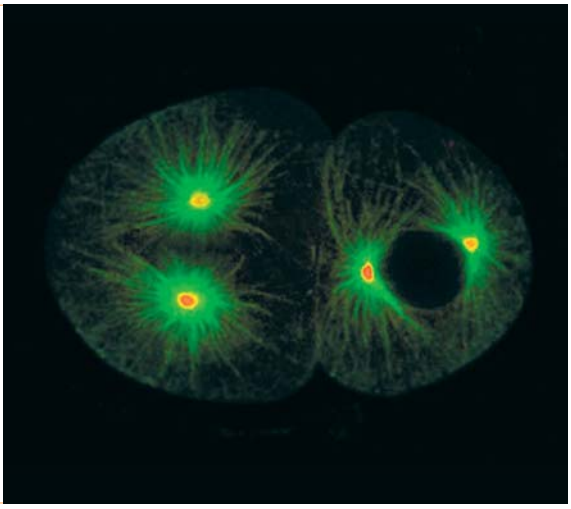
SMALL RNA: BIG NEWS

Are genes the only interesting bits of the human genome? Protein-coding genes have hogged our attention, but now tiny RNAs are where the action is.



Above: Cell division in the nematode worm *C. elegans*. The first regulatory microRNA was discovered in the worm in the 1990s.
J Ahringer

Scientists studying the human genome have always focused on the genes that encode proteins. This is because proteins are required for nearly every conceivable biological process. So we often tend to think of the genome as a kind of human 'component list'.

Indeed, protein-coding genes are in the majority, accounting for more than 99 per cent of the known genes in the genome. These genes are first transcribed, or copied, to produce a messenger RNA (mRNA) of analogous sequence, and the mRNA is then translated to make the protein.

Remarkably, however, protein-coding genes account for only 1.5 per cent of the human genome – but the remaining 98.5 per cent of the DNA is not merely junk. In among it are further genes that are copied into RNA but are not intended to make proteins. This small contingent produces the so-called non-coding RNAs; they represent about 1 per cent of known human genes, but more are being found rapidly.

Many of these RNA molecules have been known about for some time, such as ribosomal RNA (rRNA) and transfer RNA (tRNA) molecules (see 'The new RNA world' box). More recently, it has become clear that lurking quietly in the genome are genes that produce much smaller non-coding RNAs. The number of small RNA genes is unknown, but hundreds have been discovered in the last two years and the final number in the human genome may run into thousands. What is the function of this hidden army?

The big role of small RNAs

The human genome contains 25 000 or so protein-coding genes, which sounds like a lot, but is scarcely more than a worm has. The number of genes in the human genome does not account for the complexity of our bodies.

Additional complexity must be generated in other ways, such as through the process of gene regulation. This means that different genes are switched on and off at different times and at different places, or expressed either strongly or weakly, or are controlled in multiple ways. So, the complement of proteins in a cell, and the amounts of each, can be very finely controlled throughout the cell's life.

For many years, gene regulation was thought to be controlled almost entirely by proteins that bind to DNA and RNA. Then, in 1993, Victor Ambros and colleagues discovered a regulatory worm gene, *lin-4*, that produced a small RNA, not a protein. The RNA interacted with the mRNA of another gene and prevented that protein from being synthesised. However, the involvement of small RNAs in gene regulation was initially thought to be very much an exception to the rule – it was seven years before Gary Ruvkun's group found another example, *let-7*, also in the worm.

THE HUMAN GENOME IN NUMBERS

1.5	Percentage of the genome translated into proteins
27	Percentage of the genome transcribed as part of protein-coding gene expression but not translated into proteins
25	Percentage of the genome that is transcribed but not translated, and is not associated with protein-coding genes
250	Number of microRNAs currently identified (as of June 2005)
10 000	Number of protein-coded genes estimated to be regulated by microRNAs; each microRNA can target several genes, and a particular gene may be regulated by several microRNAs
98	Percentage of genomic output that is non-coding RNA
29	Percentage of genes that appear to have associated antisense transcripts
20 000	Estimated number of 'non-functional' pseudogenes in the genome

By Richard Twyman, a science writer based in York.

In parallel, a second curious RNA-based phenomenon was emerging. In 1998, Andrew Z Fire and Craig C Mello discovered that by injecting worms with double-stranded RNA (artificial RNA molecules with a similar double-helical structure to DNA), any gene with the same sequence as the injected RNA could be repressed, a process they termed RNA interference (RNAi; see page 10).

It soon became clear that these phenomena were related, and that a new biological mechanism had been discovered. As interest in RNAi as a tool for the deliberate silencing of specific genes began to grow, other small regulatory RNAs began to appear ever more frequently. These were named microRNAs.

MicroRNAs appear to work either by binding to the ends of messenger RNA produced from normal protein-coding genes and stopping these being translated, or by causing those mRNAs to be cut into smaller pieces and destroyed. This appears to depend on how closely the microRNA matches its target messenger RNA, with perfect matches usually causing the messenger RNA to be destroyed and imperfect matches causing the inhibition of protein synthesis.

Although microRNAs can be hard to detect, they have begun to spring up all over the biological world. They have been found to have a variety of roles, ranging from the control of brain development in fish to guiding the maturation of blood cells in mammals.

Because they appear to regulate cell growth and differentiation, microRNAs could be involved in the unregulated growth of cells: cancer. Indeed, in June 2005, microRNAs were shown to be involved in several types of cancer (see page 3).

Exploiting small RNAs

The discovery of RNA interference heralded a revolution in research, as it is now possible todisable genes selectively in many organisms simply by ensuring that double-stranded RNA with the same sequence is available in the cell. RNAi may also yield new therapies (see page 10).

Could microRNAs be exploited in the same way? Preliminary analysis of potential microRNA target sites in human genes suggests that as many as 30 per cent of our protein-coding genes may be regulated in this way. Therefore, a new way to tackle disease might be to 'regulate these regulators', or to add additional microRNAs into the cell to manipulate the activity of genes. On a wider scale, small RNAs have been detected in many different organisms, including microbes and plants, so the use of small RNAs could have huge implications for medicine, agriculture and biotechnology.

It's getting complex…

The discovery of the importance of small RNAs was completely unexpected. Some even describe RNA regulation as 'a new genetics'. That may be premature, as we still do not know the full extent of the microRNA toolkit, nor what most of the RNAs actually do. But it is certainly a profoundly exciting development.

And one of its most significant contributions may actually have been to provide the key biochemical tool needed to generate complex life forms, including us. John Mattick in Brisbane, Australia, has argued that the complexity of genetic regulation needed to generate multicellular organisms – which appeared only after 3 billion years of evolution – is too great to be explained solely by DNA- and protein-based approaches. Only when RNA regulation appeared, he suggests, did a sufficiently sophisticated system emerge to create the wonderful diversity of life seen today.

THE NEW RNA WORLD

Old favourites:

- Messenger RNAs (mRNAs) – carry sequence information from DNA to protein-building ribosome
- Transfer RNAs (tRNAs) – deliver amino acids to the protein-synthesising ribosome
- Ribosomal RNAs (rRNAs) – essential structural and functional component of the ribosome.

Teenagers:

- Spliceosomal RNAs – part of the machinery that chops non-coding RNA (intron sequences) out of mRNA
- Small nucleolar RNAs (snoRNAs) – mainly involved in processing rRNA in the nucleolus.

New kids on the block:

- MicroRNAs (miRNAs) – short (*c.* 21 nucleotide) RNAs that inhibit translation of target mRNAs
- Small interfering RNAs (siRNAs) – similar to miRNAs but trigger destruction of targeted mRNAs.

Longer non-coding RNAs:

While many are of unknown function, some have clear – and biologically very interesting – regulatory roles:

- Various RNAs have been implicated in gene inactivation on sex chromosomes (e.g. whereby one X chromosome in XX females is turned off completely) and in regulation of imprinted genes.
- Numerous 'antisense RNAs' have been detected (messenger RNA from a gene is always produced from one strand – the 'sense' strand of the DNA double helix; if transcription runs in the opposite direction, an 'antisense' RNA is produced). These have turned out to be surprisingly common, and in some cases they seem to be functional.
- RNAs from pseudogenes. Pseudogenes have been assumed to be nonfunctional genes that sit passively in the genome accumulating mutations. This picture may be too simplistic, however, as many are transcribed and may have some biological role.

Further reading

- Mattick JS. RNA regulation: a new genetics? Nat Rev Genet 2004;5(4):316–23.
- Mattick JS, Makunin IV. Small regulatory RNAs in mammals. Hum Mol Genet 2005;14(R1):R121–32.
- Huttenhofer A et al. Non-coding RNAs: hope or hype? Trends Genet 2005;21(5):289–97.