

JUNK DOSSIER

– background materials for „JUNK“ sorted by subject and date

General Information about “Junk DNA”

English Publications

Definition in Wikipedia.org

http://en.wikipedia.org/wiki/Junk_DNA, 12.09.05

“Junk DNA” Creates Novel Proteins

Genome News Network May 30, 2003

Gene deserts bear fruit – Regulatory elements identified in noncoding genomic sequences

The Scientist Oct. 17, 2003

The Unseen Genome: Gems among the Junk

Scientific American v289, n.5 Nov 03

'Junk' throws up precious secret - A collection of mystery DNA segments, which seem to be critical for the are causing great interest among scientists.

BBC NEWS World Edition 12 May, 2004

Junk DNA Yields New Kind Of Gene: Regulates Neighboring Gene Simply By Being Switched On

Science Daily (based on release of Harvard Medical School) 03.06.2004

Mice do fine without 'junk DNA' - Deleting non-coding regions from the genome has no apparent effect.

BioEd Online October 20, 2004

Genomics study highlights the importance of "junk" DNA in higher eukaryotes

Medical News Today 14 July 2005

Helpful junk - Brain development may be influenced by genetic parasites

The economist June 16th 2005

French Publications (ADN-poubelle)

Les zones présumées inutiles de l'ADN conditionneraient certains comportements
Le Monde 10.06.05

German Publications

Rätselhafte Wüsten im Erbgut - Was aus Müll alles werden kann: Steuerung
lebenswichtiger Gene

Frankfurter Allgemeine Zeitung, Nr. 242 18. Oktober 2003

Wie sich Junk-DNA durchsetzt -Forscher: Bei kleinen Populationsgrößen bestimmt der Zufall, bei größeren der Konkurrenzdruck die Genomgröße

Bild der Wissenschaft/ wissenschaft.de 21.11.2003

Schrottgesteuerter Anfang - Junk-DNA-Elemente regulieren früheste Embryonalentwicklung mit

Spektrumdirekt 12. Oktober 2004

Müll, der keiner mehr ist - Eine Neubewertung der sogenannten «junk-DNA»

Neue Zürcher Zeitung 9. Februar 2005

Junk and Patents

Genius of Junk (DNA) – Interview with Dr. Malcolm Simons, who patented the use of junk DNA

Catalyst, ABC TV (Australia) 10 July 2003

Human DNA

Why Do Humans Have So Few Genes?

Science Vol 309 1 July 2005

Bioinformatics

Visualising 'junk' DNA through bioinformatics

Journal of Biological Education 39(2) 2005

Medicine

L'ère post-génomique de la médecine

Pulsion - Bulletin des Hôpitaux universitaire de Genève 02/2005

Criminology/ Forensics

Ein Test aus "Abfall" - Der "genetische Fingerabdruck" ist schon mit nur einer Zelle nachweisbar

Die Welt 13. 03. 2001

Arts

DNA 50 Four Plus: Writing DNA - A Wellcome Trust exhibition at the TwoTen Gallery

British Medical Journal p. 827 12 April 2003

Art after DNA - The double helix has inspired scientists and artists alike

Nature Vol 422 24 April 2003

Junk DNA

From Wikipedia, the free encyclopedia.

In molecular biology, "**junk**" DNA is a collective label for the portions of the DNA sequence of a chromosome or a genome for which no function has been identified. About 97% of the human genome has been designated as junk, including most sequences within introns and most intergenic DNA. While much of this sequence is probably an evolutionary artifact that serves no present-day purpose, some of it may function in ways that are not currently understood. Recent studies have, in fact, suggested functions for certain portions of what has been called junk DNA. The "junk" label is therefore recognized as something of a misnomer, and many would prefer the more neutral term "noncoding DNA".

Broadly, the science of functional genomics has developed widely accepted techniques to characterize protein-coding genes, RNA genes, and regulatory regions. In the genomes of most plants and animals, however, these together constitute only a small percentage of genomic DNA (less than 2% in the case of humans). The function of the remainder, if any, remains under investigation. Most of it can be identified as repetitive elements that have no known biological function (although they are useful to geneticists for analyzing lineage and phylogeny). Still, a large amount of sequence in these genomes falls under no existing classification other than "junk".

It is notable that overall genome size, and by extension the amount of junk DNA, appears to have little relationship to organism complexity: the genome of the unicellular *Amoeba dubia* has been reported to contain more than 200 times the amount of DNA in humans. The *Fugu rubripes* pufferfish genome is only about one tenth the size of the human, yet seems to have a comparable number of genes. Most of the variance appears to lie in what is now known only as junk DNA. This puzzle is known as the "C-value enigma" (7).

Contents

- 1 Hypotheses of origin and function
- 2 Evolutionary conservation of "junk" DNA
- 3 Creation-evolution controversy
- 4 References
- 5 See also
- 6 External links
 - 6.1 Creation-evolution controversy

Hypotheses of origin and function

There are many hypotheses, none conclusively established, for how junk DNA arose and why it

persists in the genome:

- These chromosomal regions could be the remains of ancient pseudogenes, which have been cast aside and fragmented during evolution. A related hypothesis suggests that the junk represents the accumulated DNA of retroviruses.
- Junk DNA may act as a protective buffer against genetic damage and harmful mutations. For example, a high proportion of nonfunctional sequence makes it unlikely that a functional element will be destroyed in a chromosomal crossover event, possibly making a species more tolerant to this important mechanism of genetic recombination.
- Junk DNA might provide a reservoir of sequences from which potentially advantageous new genes can emerge. In this way, it may be an important genetic basis for evolution.
- Some portions of junk DNA could serve presently unknown regulatory functions, controlling the expression of certain genes and/or the development of an organism from embryo to adult.
- Junk DNA may serve other, unknown purposes. For example, some non-coding RNAs have been discovered in what had been considered junk.
- Junk DNA may have no function. For example, recent experiments removed 1% of the mouse genome and were unable to detect any effect on the phenotype³. This result suggests that the DNA is, in fact, non-functional. However, it remains a possibility that there is some function that the experiments performed on the mice was merely insufficient to detect.

Evolutionary conservation of "junk" DNA

Comparative genomics is a promising direction in studying the function of junk DNA. Biologically functional sequences, as the theory goes, tend to undergo mutation at a slower rate than nonfunctional sequence, since mutations in these sequences are likely to be selected against. For example, the coding sequence of a human protein-coding gene is typically about 80% identical to its mouse ortholog, while their genomes as a whole are much more widely diverged. Analyzing the patterns of conservation between the genomes of different species can suggest which sequences are functional, or at least which functional sequences are shared by those species. Functional elements stand out in such analyses as having diverged less than the surrounding sequence.

Comparative studies of several mammalian genomes suggest that approximately 5% of the human genome has evolved under purifying selection⁵ since the divergence of the mammals. Since known functional sequence comprises less than 2% of the human genome, it *appears* that there may be more functional "junk" DNA in the human genome than there is known functional sequence.

A surprising recent finding was the discovery of nearly 500 *ultraconserved* elements², which are shared at extraordinarily high fidelity among the available vertebrate genomes, in what had previously been designated as junk DNA. The function of these sequences is currently under intense scrutiny, and there are preliminary indications^{2,6} that some may play a regulatory role in vertebrate development from embryo to adult.

It must be noted that all present results concerning evolutionarily conserved human "junk"

DNA are expressed in highly preliminary, probabilistic terms, since only a handful of related genomes are available. As more vertebrate, and especially mammalian, genomes are sequenced, scientists will develop a clearer picture of this important class of sequence. However, it is always possible, though highly unlikely, that there are significant quantities of functional human DNA that are not shared among these species, and which would thus not be revealed by these studies.

On a theoretical note, it is often observed that the presence of high proportions of truly nonfunctional "junk" DNA would seem to defy evolutionary logic. Replication of such a large amount of useless information each time a cell divides would waste energy. Organisms with less nonfunctional DNA would thus enjoy a selective advantage, and over an evolutionary time scale, nonfunctional DNA would tend to be eliminated. If one assumes that most junk DNA is indeed nonfunctional, then there are several hypotheses for why it has not been eliminated by evolution: (1) the energy required to replicate even large amounts of nonfunctional DNA is in fact relatively insignificant on the cellular or organismal scale, so no selective pressure results; (2) the aforementioned possible advantage of having extra DNA as a reservoir of potentially useful sequences; and (3) retroviral or transposon insertions of nonfunctional sequence occurring faster than evolution can eliminate it. These are all hypotheses for which the time scales involved in evolution may make it difficult for humans to rigorously investigate.

Creation-evolution controversy

The question of whether junk DNA is really junk has played a minor role in the creation-evolution controversy. Some proponents of evolution hold (1) that at least some junk DNA is truly nonfunctional and (2) that this is evidence for common descent, since the hierarchy of nonfunctional genetic similarities mimics the phylogenetic tree. [1] (<http://www.talkorigins.org/faqs/comdesc/section4.html#pseudogenes>) Advocates of creationism and intelligent design typically contend that no DNA is junk, or that such junk DNA demonstrates only deterioration rather than macroevolution. Another claim commonly made by creationists is that the theory of evolution caused scientists to assume most DNA was functionless, stifling research into the functions of junk DNA. [2] (<http://www.ideacenter.org/contentmgr/showdetails.php?id=1155>)

References

1. Gibbs W.W. (2003) "The unseen genome: gems among the junk", *Scientific American*, **289**(5): 46-53. (A review, written for non-specialists, of recent discoveries of function within junk DNA.)
2. G. Bejerano et al. "Ultraconserved Elements in the Human Genome (<http://www.soe.ucsc.edu/%7Ejill/ultra.html>) ". *Science* **304**:1321-1325, May 2004. Discussed in "Junk' DNA reveals vital role (<http://www.nature.com/nsu/040503/040503-9.html>) ", *Nature* (2004).
3. M.A. Nobrega, Y. Zhu, I. Plajzer-Frick, V. Afzal and E.M. Rubin (2004) "Megabase deletions of gene deserts result in viable mice", *Nature*, **431**: 988-993.
4. Mattick, John S. (2004) "The Hidden Layer of Noncoding RNA: a Digital Control System Underpinning Mammalian Development and Diversity", HGM Symposium 2004 Session 4/16.

5. Mouse Genome Sequencing Consortium. "Initial sequencing and comparative analysis of the mouse genome". *Nature* 420:520-562, December 2002.
6. Sandelin, A., et. al. "Arrays of ultraconserved elements span the loci of key development genes in vertebrate genomes". *BMC Genomics* 5(1):99, December 2004.
7. Gregory, T.R. (Ed.) (2005) The Evolution of the Genome, Elsevier, San Diego.
8. Wahls, W.P., et. al. "Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells" (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract>) "Cell" 60(1):95-103 (1990).

See also

- intron
- centromere
- telomere
- Alu repeat
- repeated sequence (DNA)
- satellite DNA
- molecular evolution
- selfish DNA

External links

- ABC News: Genius of Junk (DNA) (<http://www.abc.net.au/catalyst/stories/s898887.htm>)
- Focus Online - News from Harvard Schools - Junk DNA Yields New Kind of Gene (http://focus.hms.harvard.edu/2004/June4_2004/genetics.html)
- noncodingDNA.com (<http://www.noncodingdna.com>)

Creation-evolution controversy

- Junk DNA from Evowiki (http://www.evowiki.org/index.php/Junk_DNA) This article argues that much DNA is truly junk and that this proves an atheistic version of evolution. Much of the evidence is based in science, but is not very conclusive at this point.
- DNA: marvelous messages or mostly mess? (<http://www.answersingenesis.org/creation/v25/i2/dna.asp>) by Jonathan Sarfati. A creationist argument that junk DNA is functional. The article argues that much "junk DNA" is probably functional, presenting a number of likely uses. It further argues that for "junk DNA" with currently unknown function, we should therefore not be so quick to go from "we don't know the function" to "there is no function". Much of the science is scientists' speculations, published in leading science journals, that happen to support the article's viewpoint.

Retrieved from "http://en.wikipedia.org/wiki/Junk_DNA"

Categories: DNA | Genomics

“Junk DNA” Creates Novel Proteins

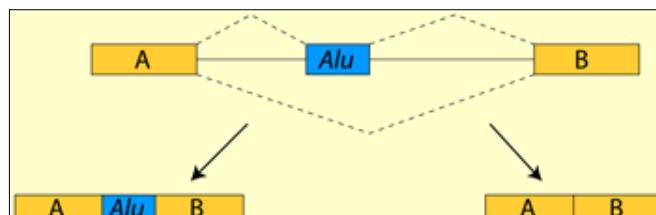
By [Nancy Touchette](#)

DNA sequences long considered genomic garbage are finally getting a little respect. Researchers have figured out how short stretches of DNA that do not normally code for proteins worm their way into genes.

May 30, 2003

This can result in the production of abnormal proteins and lead to genetic diseases, such as Alport Syndrome, a rare kidney disease. But the sequences, sometimes called “junk DNA,” have also allowed humans and other species to create new proteins in a process that has dramatically influenced evolution.

Gil Ast and his colleagues at Tel Aviv University in Israel have figured out how the sequences, known as *Alu* elements, are incorporated into genes to create novel proteins. More than 300,000 sequences are poised for insertion into genes—all that’s needed is a single mutation.



Alu sequences lying between coding regions of genes can be spliced in (left) or out right during RNA processing to create many versions of the same gene.

Through a process called alternative splicing, humans create multiple versions of a gene and, consequently, multiple proteins. It's a way of constructing a new protein, while keeping a backup copy of the original version.

“This is a way to experiment with new structures,” says Wojciech Makalowski of the Pennsylvania State University in University Park. “We create two versions of a protein and check which is better.”

For example, the researchers found that the ADAR2 enzyme contains 40 amino acids in its active site that are derived from an *Alu* element. The addition changes the activity of the enzyme.

“Without this enzyme, we would die,” says Ast. “The incorporation of an *Alu* into the enzyme changed its function, and we have evolved to rely on it.”

One of the biggest surprises to come from the sequencing of the human genome was that we have about 30,000 genes but produce approximately 90,000 proteins. And 99 percent of our DNA codes for no protein at all. The new research provides a clue as to why we have so much “junk DNA.” It also suggests an explanation of how so few genes can produce so many proteins.

Alu elements are short sequences of DNA that are peppered throughout the genome. They comprise approximately ten percent of the entire genome—ten times more than all the genes put together. Until recently, their function had remained a mystery.

But a few years ago, researchers studying splicing—a process by which pieces of RNA are cut and pasted together—discovered that many *Alu* sequences get inserted into existing coding sequences.

In the new study, published in *Science*, Ast and his colleagues discovered a unique sequence within most *Alu* elements that can be mutated at a single base to create a new splice site. Splice sites are special sequences recognized by the cellular machinery that cuts and pastes together

RNA.

Although mice and human have the same number of genes, and many genes share the same functions, only primates have *Alu* sequences. Ast speculates that these sequences have played a key role in our evolution.

"We believe that *Alus* allowed the shuffling of genetic information that may have led to the evolution of primates," says Ast. "They may contribute to a lot of disorders we don't even know about yet. But they have also created genetic diversity."

...

[Lev-Maor, G. et al. The birth of an alternatively spliced exon: 3' splice-site selection in *Alu* exons. *Science* 300, 1288-1291 \(May 23, 2003\).](#)

[W. Makalowski. Not junk after all. *Science* 300, 1246-1247 \(May 23, 2003\).](#)

Genome News Network is an editorially independent online publication of the J. Craig Venter Institute.
© 2000 - 2004 J. Craig Venter Institute.
All rights reserved. This material may not be published, broadcast, rewritten or redistributed.

Oct. 17, 2003

[Previous](#) | [Archive](#) | [Next](#)

► **DAILY NEWS**

Gene deserts bear fruit

Regulatory elements identified in noncoding genomic sequences

By **Cathy Holding**

The evidence that "junk" DNA contains highly conserved sequences that have a regulatory function is mounting: their genetic and (controversial) commercial value had been guessed at as long ago as 1989. In the October 17 *Science*, Marcelo Nobrega and colleagues at the US Department of Energy Joint Genome Institute examined the "gene deserts" flanking the human **DACH gene** and report that they contain several important enhancer functions that have been conserved across species in over a billion years of parallel evolution (*Science*, 302:413, October 17, 2003).

Nobrega et al. compared human DACH flanking sequences with mouse genomic DNA and by combining additional genome comparison information from distantly related vertebrates such as frog, zebrafish, and pufferfish, narrowed the number of conserved sequences from 1098 to 32. Nine of these were cloned upstream of the mouse heat shock protein 68 minimal promoter driving beta-galactosidase expression. These were used to create transgenic mice whose subsequent development revealed enhancer effects of the elements consistent with DACH endogenous gene expression. Genes flanking the DACH orthologues vary with species, but the adjacent gene deserts were found to be maintained.

"The size of genomic regions believed to be functionally linked to a particular gene may need to be expanded to take into account the possibility of essential regulatory sequences acting over near-megabase distances," conclude the authors.



▲ **CURRENT ISSUE**

Sep 12, 2005

Links for this article

C. Dennis, "Geneticists question fees for use of patented 'junk' DNA," *Nature*, 423:105 May 8, 2003. [[PubMed Abstract](#)]

M. Nobrega et al., "Scanning human gene deserts for long-range enhancers," *Science*, 302:413, October 17, 2003. <http://www.sciencemag.org>

Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/>

X. Caubit et al., "Mouse Dac, a novel nuclear factor with homology to Drosophila dachshund shows a dynamic expression in the neural crest, the eye, the neocortex, and the limb bud," *Developmental Dynamics*, 214:66-80, September 16, 1999.

<http://www3.interscience.wiley.com/cgi-bin/abstract/64500801/ABSTRACT>

[Previous](#) | [Archive](#) | [Next](#)

The Unseen Genome: Gems among the Junk

W. WAYT GIBBS / Scientific American v289, n.5 Nov03

Just when scientists thought they had DNA almost figured out, they are discovering in chromosomes two vast, but largely hidden, layers of information that affect inheritance, development and disease.



FLECKS OF DARK BROWN in an iris may be a telltale sign of the hidden genome at work. Certain traits are transmitted not through ordinary genes but rather through chemical modifications to the chromosomes, changes that are regulated in part by bits of "junk" DNA. Unlike genetic mutations, these heritable traits are often reversible and appear in some cells but not others. (The white sphere on the iris is a reflection of the light shining on the eye.)

About 20 years ago, astronomers became convinced that distant galaxies were moving in ways that made no sense, given the laws of gravity and the fabric of celestial objects visible in the sky. Gradually they were forced to conclude that the universe is not as empty as it appears, that in fact it must be dominated by some dark kind of matter. Although no one knew what the stuff is made of or how it works, scientists could see from its effects that it is out there. The quest to understand dark matter (and more recently, dark energy) meant revising or replacing theories, but it reenergized astrophysics and cosmology.

A similar revelation is now unfolding in molecular genetics. This year biologists celebrated the 50th anniversary of the discovery of the double helix, and the Human Genome Project announced its completion of a "final draft" of the DNA sequence for *Homo sapiens*. Scientists have clearly mastered DNA in the lab. Yet as they compare the DNA of distantly related species and look

more closely at how chromosomes function in living cells, they are increasingly noticing effects that current theories cannot explain.

Journals and conferences have been buzzing with new evidence that contradicts conventional notions that genes, those sections of DNA that encode proteins, are the sole mainspring of heredity and the complete blueprint for all life. Much as dark matter influences the fate of galaxies, dark parts of the genome exert control over the development and the distinctive traits of all organisms, from bacteria to humans. The genome is home to many more actors than just the protein-coding genes.

The extent of this unseen genome is not yet clear, but at least two layers of information exist outside the traditionally recognized genes. One layer is woven throughout the vast "noncoding" sequences of DNA that interrupt and separate genes. Though long ago written off as irrelevant because they yield no proteins, many of these sections have been preserved mostly intact through millions of years of evolution. That suggests they do something indispensable. And indeed a large number are transcribed into varieties of RNA that perform a much wider range of functions than biologists had imagined possible. Some scientists now suspect that much of what makes one person, and one species, different from the next are variations in the gems hidden within our "junk" DNA.

Above and beyond the DNA sequence there is another, much more malleable, layer of information in the chromosomes. "Epigenetic" marks, embedded in a

Overview/Hidden Gene

- Geneticists have long focused on just the small part of DNA that contains blueprints for proteins. The remainder in humans, 98 percent of the DNA- was often dismissed as junk. But the discovery of many hidden genes that work through RNA, rather than protein, has overturned that assumption.
- These RNA-only genes tend to be short and difficult to identify. But some of them play major roles in the health and development of plants and animals.
- Active forms of RNA also help to regulate a separate "epigenetic" layer of heritable information that resides in the chromosomes but outside the DNA sequence.

mélange of proteins and chemicals that surround, support and stick to DNA, operate through cryptic codes and mysterious machinery. Unlike genes, epigenetic marks are routinely laid down, erased and rewritten on the fly. So whereas mutations last a lifetime, epigenetic mistakes-implicated in a growing list of birth defects, cancers and other diseases-may be reversible with drugs. In fact, doctors are already testing such experimental treatments on leukemia patients.

Researchers are also coming to realize that just about anything that can happen in the genome does happen, says Carmen Sapienza of Temple University, who started investigating epigenetic phenomena back when they were dismissed as minor anomalies. "There may even be fundamental mechanisms still to discover," Sapienza considers. "I think we are entering the most interesting time yet in genetics."

The Perils of Dogma

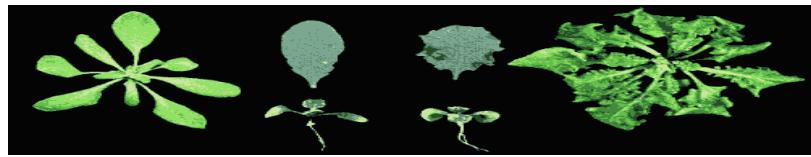
IT WILL TAKE YEARS, perhaps decades, to construct a detailed theory that explains how DNA, RNA and the epigenetic machinery all fit into an interlocking, self-regulating system. But there is no longer any doubt that a new theory is needed to replace the central dogma that has been the foundation of molecular genetics and biotechnology since the 1950s.

The central dogma, as usually stated, is quite simple: DNA makes RNA, RNA makes protein, and proteins do almost all the real work of biology. The idea is that information is stored in the twisted ladders of DNA, specifically in the chemical bases (commonly labeled A, T, G and C) that pair up to form the rungs of the ladders. A gene is just a particular sequence of bases on one side of the ladder that specifies a protein.

The dogma holds
that genes express
themselves as
proteins, which are

made in four steps: First an enzyme docks to the chromosome and slides along the gene, transcribing the sequence on one strand of DNA into a single strand of RNA. Next, any introns-noncoding

parts of the initial RNA transcript-are snipped out, and the rest is spliced together to make a piece of messenger RNA. The RNA message then moves out of the nucleus to the main part of the cell, where molecular machines translate it into chains of amino acids. Finally, each chain twists and folds into an intricate three-dimensional shape.



BIG DIFFERENCES in the appearance and health of organisms can arise from small changes to tiny, unconventional genes. *Arabidopsis* plants, for example, normally have spoon-shaped leaves (*left*). But when scientists interfered with the action of a microRNA, produced by an RNA-only gene, the mutant *arabidopsis* plants developed gross defects (*right*). The microRNA appears to control the activity levels of numerous genes.

It is their shapes that make proteins so remarkably versatile. Some form muscles and organs; others work as enzymes to catalyze, metabolize or signal; and still others regulate genes by docking to specific sections of DNA or RNA. No great wonder, then, that many biologists (and journalists) have taken the central dogma to imply that, with very few exceptions, a DNA sequence qualifies as a gene only if it can produce a protein.

"Typically when people say that the human genome contains 27,000 genes or so, they are referring to genes that code for proteins," points out Michel Georges, a geneticist at the University of Liège in Belgium. But even though that number is still tentative-estimates range from 20,000 to 40,000-it seems to confirm that there is no clear correspondence between the complexity of a species and the number of genes in its genome. "Fruit flies have fewer coding genes than roundworms, and rice plants have more than humans," notes John S. Mattick, director of the Institute for Molecular Bioscience at the University of Queensland in Brisbane, Australia. "The amount of noncoding DNA, however, does seem to scale with complexity."

In higher organisms (such as humans), genes "are fragmented into chunks of protein-coding sequences separated by often extensive tracts of nonprotein-coding sequences," Mattick explains. In fact, protein-coding chunks account for less than 2 percent of the DNA in human chromosomes. Three billion or so pairs of bases that we all carry in nearly every cell are there for some other reason. Yet the introns within genes and the long stretches of intergenic DNA between genes, Mattick says, "were immediately assumed to be evolutionary junk."

That assumption was too hasty. "Increasingly we are realizing that there is a large collection of 'genes' that are clearly functional even though they do not code for any protein" but produce only RNA, Georges remarks. The term "gene" has always been somewhat loosely defined; these RNA-only genes muddle its meaning further. To avoid confusion, says Claes Wahlestedt of the Karolinska Institute in Sweden, "we tend not to talk about 'genes' anymore; we just refer to any segment that is transcribed [to RNA] as a 'transcriptional unit.'"

Based on detailed scans of the mouse genome for all such elements, "we estimate that there will be 70,000 to 100,000," Wahlestedt announced at the International Congress of Genetics, held this past July in Melbourne. "Easily half of these could be noncoding." If that is right, then for every DNA sequence that generates a protein, another works solely through active forms of RNA-forms that are not simply intermediate blueprints for proteins but, rather, directly alter the behavior of cells.

What is true for mice is probably true for people and other animals as well. A team of scientists at the National Human Genome Research Institute (NHGRI) recently compared excerpts from the genomes of humans, cows, dogs, pigs, rats and seven other species. Their computer analysis turned up 1,194 segments that appear with only minor changes in several species, a strong indication that the sequences contribute to the species' evolutionary fitness. To the researchers' surprise, only 244 of the segments sit inside a protein-coding

stretch of DNA. About two thirds of the conserved sequences lie in introns, and the rest are scattered among the intergenic "junk" DNA.

"I think this will come to be a classic story of orthodoxy derailing objective analysis of the facts, in this case for a quarter of a century," Mattick says. "The failure to recognize the full implications of this—particularly the possibility that the intervening noncoding sequences may be transmitting parallel information in the form of RNA molecules—may well go down as one of the biggest mistakes in the history of molecular biology."

More than a Messenger

NOW THAT BIOLOGISTS have turned their attention back to RNA, they are finding it to be capable of impressive feats of cellular chemistry. Like proteins, some RNA transcripts can interact with other bits of RNA, with DNA, with proteins and even with small chemical compounds. Proteins are analog molecules, however; they bind to targets in much the way keys fit in locks. "The beauty of RNA is that it has a specific sequence, so it's digital, like a zip code," Mattick points out. A bit of RNA can float around until it bumps into a DNA (or another RNA) that has a complementary sequence; the two halves of the ladder then join rungs. (Two segments are complementary when all C bases mate with G's and all T or U bases join to A's.)

As an example of the unappreciated power of RNA, consider pseudogenes. Surveys of human DNA have found in it almost equal numbers of genes and pseudogenes—defective copies of functional genes. For decades, pseudogenes have been written off as molecular fossils, the remains of genes that were broken by mutation and abandoned by evolution. But this past May a group of Japanese geneticists led by Shinji Hirotune of the Saitama Medical School reported their discovery of the first functional pseudogene.

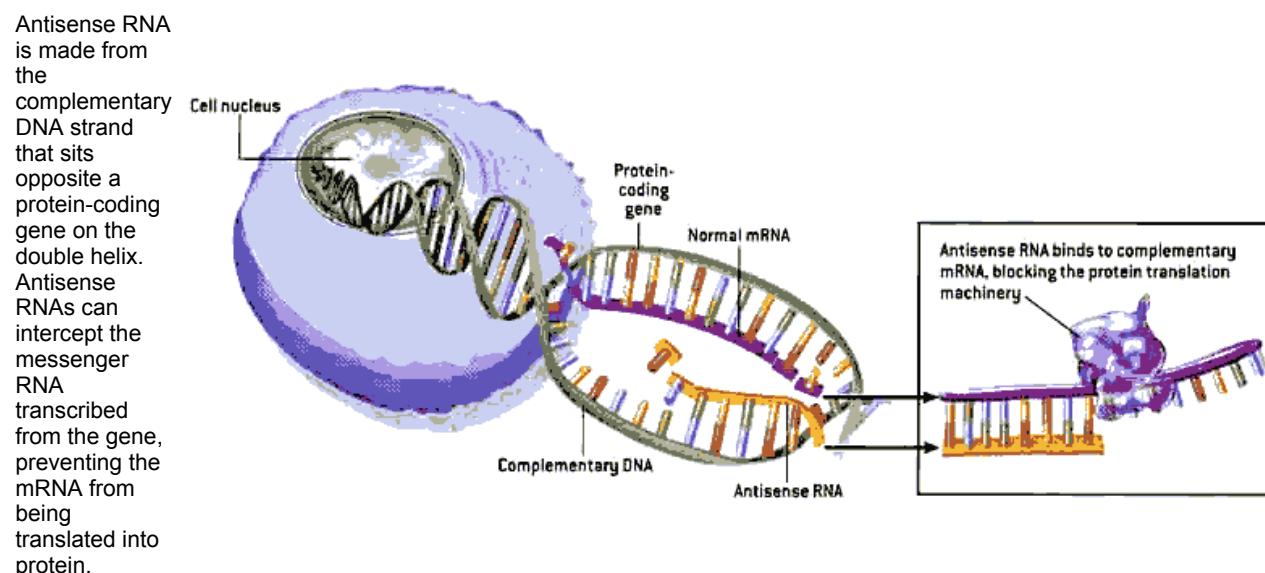
Hirotune was genetically engineering mice to carry a fruit fly gene called *sex-lethal*. Most mice did fine with this foreign gene, but in one strain *sex-lethal* lived up to its name; all the mice died in infancy. Looking closer,

the scientists discovered that in those mice *sex-lethal* happened to get inserted right into the middle of a pseudogene, clobbering it. This pseudogene (named *makorin1-p1*) is a greatly shortened copy of *makorin1*, an ancient gene that mice share with fruit flies, worms and many other species. Although researchers don't know what *makorin1* does, they do know that mice have lots of *makorin1* pseudogenes and that none of them can make proteins. But if pseudogenes do nothing, why were these mice dying when they lost one?

For some reason, *makorin1-and* apparently only *makorin1-all* but shuts down when its pseudogene p1 is knocked out. RNA made from the pseudogene, in other words, controls the expression of the "real" gene whose sequence it mimics, even though the two lie on different chromosomes. There is nothing pseudo about that.

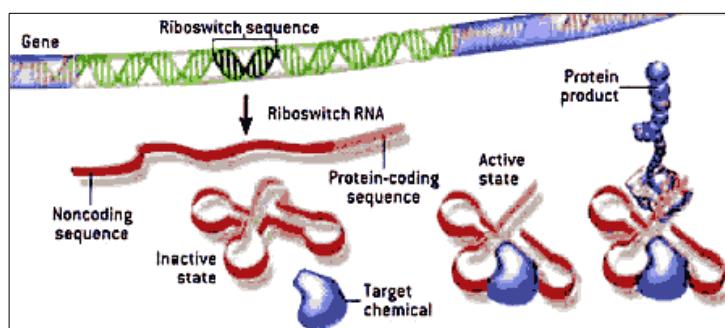
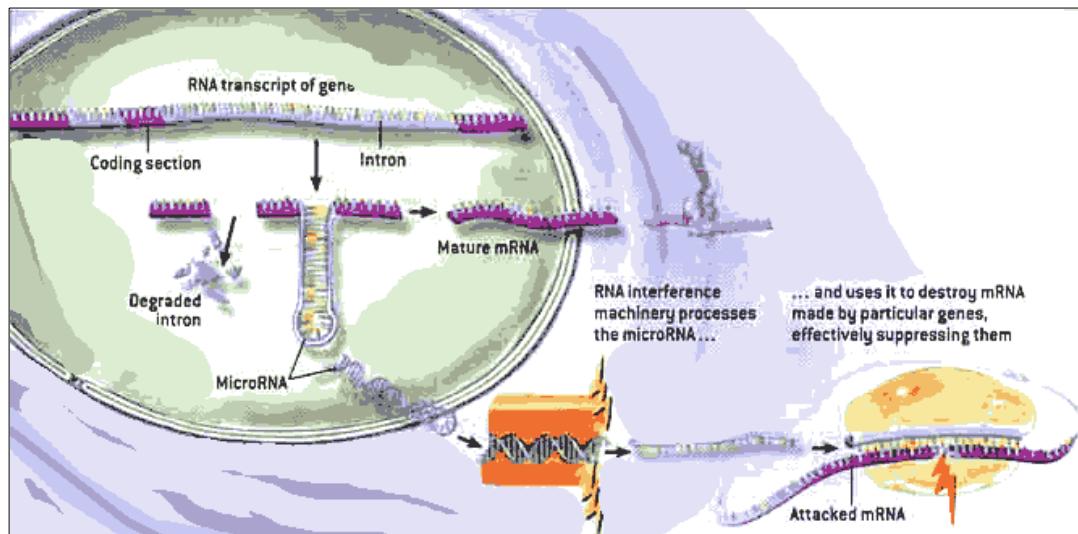
A BESTIARY OF UNCONVENTIONAL GENES

GENES, according to conventional wisdom, are those sections of the DNA that encode functional proteins. Such sequences make up only about 2 percent of the human genome, however. The rest of the human genome is filled with DNA that is "noncoding"-but not useless. Scientists are discovering many noncoding genes that give rise to surprisingly active RNAs, including varieties that can silence or regulate conventional genes.



Protein-coding genes contain noncoding sections called introns. Introns are snipped out of the initial RNA transcript; the coding sections are then spliced to create a mature mRNA. Although many introns degrade, some contain active elements, such as microRNAs that can exploit the "RNA interference" effect to control other genes

Riboswitches are a newly discovered form of RNA that act as precision genetic switches. Produced in many cases from noncoding DNA between known genes, a riboswitch folds into a complex shape. One part of the folded RNA can bind to a specific target protein or chemical. Another part contains the RNA code for a protein product. The riboswitch turns "on" and produces the protein it encodes only when in the presence of its target.



It is too early to say whether many pseudogenes give rise to active RNA. But there are plenty of other sources scattered about the dark parts of the genome. Every normal protein-making gene, for instance, has a complementary DNA sequence that sits on the other side of the ladder and usually is not transcribed

into RNA. Biologists like to think of this as a backup copy, because the cell can use it to repair damage to the gene.

In some cases, however, the backup has its own agenda. While the gene is producing a sensible RNA message, its alter ego can churn out an "antisense" RNA that has a complementary sequence. Whenever matched sense and antisense RNAs meet, they mesh to form their own double-stranded ladders—effectively interfering with the gene's ability to express its protein.

Biologists knew that bacteria and plants can produce antisense, but most thought that mammals rarely do. In April, Galit Rotman and her co-workers at CompuGen, a biotech firm in Tel Aviv, dashed that assumption. They screened human genome databases and concluded that at least 1,600 human genes (and probably many more) have a mate that yields antisense RNAs.

These competing RNAs may suppress a gene just by tying up the gene's messenger RNA. But Rotman speculates that *they* employ a built-in genome censor, known as the RNA interference machinery. Scientists are still enthralled by the discovery several years ago of this scheme for selectively silencing individual genes. When double-stranded RNA appears in a cell, enzymes dice it up, peel the two strands apart, and use one RNA fragment to seek out and destroy any other RNA messages that stick to its sequence. The system protects cells against viruses, which often deliver their payloads in the form of double-stranded RNA. But the censor also provides a handy way for scientists to shut off any gene at will [see "Censors of the Genome," by Nelson C. Lau and David P. Bartel; SCIENTIFIC AMERICAN, August].

Neither pseudogenes nor antisense RNAs, however, can explain the crinkled leaves that Detlef Weigel of the Max Planck Institute for Developmental Biology in Tübingen, Germany, and his collaborators saw in their *arabidopsis* plants this summer. These weeds of the mustard family normally have smooth, spoon-shaped leaves. The plants owe their gentle symmetrical curves, Weigel's group showed in *Nature* this past August, in part to a kind of active RNA called microRNA.

MicroRNAs, first observed a few years ago in roundworms, are short noncoding RNAs that fold back on themselves, like hairpins. In *arabidopsis*, the JAW microRNA doubles over and is then captured by the RNA interference machinery, just as if it had come out of a virus. But the JAW sequence matches a handful of different protein-making genes, members of a family that control the shape and size of the plant. The censor dutifully represses each of them by chopping up much (but not all) of the messenger RNA they produce. Thus, JAW, a tiny RNA-only gene, serves as the main lever by which *arabidopsis* cells adjust the volume of a suite of crucial protein genes. When Weigel's group engineered plants in which the microRNA could not do its job, the plants became sick and deformed.

In just the three years since researchers started looking in earnest, they have found hundreds of microRNAs-more than 150 in humans alone. They seem to be a well-established means for organisms to wrangle genes; about half the microRNAs in humans also appear, in nearly identical form, in the DNA of pufferfish, even though the two species went their separate ways some 400 million years ago.

Just what those 150-plus microRNAs do in people is a mystery. Anna M. Krichevsky of Harvard Medical School suspects that, among other things, they play an important role in brain development. Her lab used a "gene chip" to screen mouse neurons for 44 different kinds of microRNA. Krichevsky reported in September that levels of at least nine distinct microRNAs are precisely regulated in the mice as their brains grow. The link is still indirect, but as Diya Banerjee of Yale University noted last year in a review of microRNA science, "it seems that we are on the verge of an explosion of knowledge in this area."

Moving Genetics Forward

EVER SINCE THE INVENTION of recombinant DNA technology made genetic engineering feasible, most research in genetics has been run in "reverse." Reverse genetics begins with a particular gene of interest.

The scientist fiddles with that gene in a cell culture or a living organism, watches what happens, and then tries to deduce the gene's function. It is a classic reductionist approach, and it can be very powerful.

But the gradual realization that the genome includes hidden genes functional sequences that were misclassified as junk highlights a major problem with reverse genetics: it can lead to tunnel vision. So recently a number of geneticists have been returning to the older practice of "forward" genetics as a way to identify the genes, both conventional and unconventional, that they don't know about.

Phenomix, a biotechnology company in La Jolla, Calif., founded last year by several prominent genetics teams, hopes to make a business out of the approach. The firm has set up a kind of production line for making mutant mice. In each group of mice, mutations to random points in the genome disable not just standard protein-coding genes but also hidden genes that make only active forms of RNA.

Phenomix starts with both healthy mice and mice that have diseases analogous to common human illnesses, such as diabetes, asthma, arthritis and Parkinson's disease. Some mutations induce or alleviate symptoms of these disorders in the mice. Researchers then do genetic screening to determine which mutations accounted for the effects. Whether the approach will inspire better drug designs remains to be seen. But forward genetics has already unearthed genetic phenomena, such as a functional pseudogene (*see main text*), that no one knew were possible. W.W.G.

Digital and Analog

PROTEINS MAY BE the draft horses of the cell, but active RNA sometimes wields the whip. And several kinds of RNA have turned up doing mules' work as well: catalyzing, signaling and switching as competently as any protein. In fact, some inherited diseases have stumped researchers because, in their diligent search for a mutant protein, the investigators ignored the active RNA right under their noses.

Doctors struggled for more than nine years, for example, to nail down the gene responsible for cartilage hair hypoplasia. This recessive disease was first identified in the Amish, one in 19 of whom carries a copy of the defective gene, which causes an unusual kind of dwarfism. People with CHH are not only small in stature but also at high risk for cancer and immune disorders. Geneticist Maaret Ridanpää of the University of Helsinki tracked the gene to chromosome nine, sequenced a large region and then proceeded to check all 10 protein-making genes in the area, one by one. None caused the disease.

Finally, in 2001, Ridanpää and his co-workers identified the culprit, an RNA-only gene called *RMRP*. The RNA transcribed from *RMRP* links up with proteins to form an enzyme that works inside a cell's energy generators, the mitochondria. A change to just a single base at a critical spot on this RNA can

mean the difference between a full-size, healthy life and a short, abbreviated one (if the same mutation is inherited from both parents). Such "analog" RNAs, which fold up into complex shapes just as proteins do, have also been discovered recently to be essential to the function of enzymes that protect the chromosomes and that escort secreted protein signals out of cells' portholes.

Perhaps the most intriguing form of RNA yet discovered is the riboswitch, isolated last year by Ronald R. Breaker's lab at Yale. He and others have long wondered how, billions of years ago, the very earliest chemical precursors to life got along in the RNA world before DNA and proteins existed. They speculated that such proto-organisms would need to use RNA as sensors and switches to respond to changes in the environment and in their metabolism. To test the idea, they tried to create RNAs with such capabilities.

"Our laboratory successfully produced a number of synthetic RNA switches," Breaker recalls. Dubbed riboswitches, these long RNAs are both coding and noncoding at once. As the RNA folds up, the noncoding end becomes a sensitive receptor for a particular chemical target. A collision with the target flips the switch, causing the other end, which contains a standard blueprint for a protein, to change shape. The riboswitch thus gives rise to a protein, much like a normal gene does—but only when it senses its target.



CLONES IN ALL BUT NAME, these littermates from a highly inbred strain of mice share practically identical DNA. Yet their coat colors run the spectrum from golden yellow to mahogany brown because of variations in the "epigenetic" chemical attachments each has to a particular segment of DNA that lies outside any known gene. The hair color of these mice cannot be predicted by current theories of genetics.

Breaker's group started hunting for riboswitches in the wild and soon found them hiding in intergenic DNA. These precision genetic switches have been extracted now from species in all three kingdoms of life. "This implies that they were probably present in the last common ancestor," not long after the dawn of evolution, Breaker argues.

In August, Breaker and his co-workers reported that one family of riboswitches regulates the expression of no fewer than 26 genes in *Bacillus subtilis*, a common kitchen bacterium. These are not once-in-a-blue-moon genes, either, but genes that the bacterium relies on to metabolize such basic staples as sulfur and amino acids. Breaker estimates that *B. subtilis* has at least 68 genes, nearly 2 percent of its total, under the control of riboswitches. His lab has already begun engineering the hybrid digital-analog molecules to do useful things, such as selectively kill germs.

The Big Picture

AS BIOLOGISTS SIFT more and more novel kinds of active RNA genes out of the long-neglected introns and intergenic stretches of DNA, they are realizing that science is still far from having a complete parts list for humans or any other higher species. Unlike protein-making genes, which have standard "start" and "stop" codes, RNA-only genes vary so much that computer programs cannot reliably pick them out of DNA sequences. To spur the technology on, the NHGRI is launching this autumn an ambitious \$36-million project to produce an "Encyclopedia of DNA Elements." The goal is to catalogue every kind of RNA and protein made from a select 1 percent of the human genome-in three years.

No one knows yet just what the big picture of genetics will look like once this hidden layer of information is made visible. "Indeed, what was damned as junk because it was not understood may, in fact, turn out to be the very basis of human complexity," Mattick suggests. Pseudogenes, riboswitches and all the rest aside, there is a good reason to suspect that is true. Active RNA, it is now coming out, helps to control the large-scale structure of the chromosomes and some crucial chemical modifications to them—an entirely different, epigenetic layer of information in the genome.

The exploration of that epigenetic layer is answering old conundrums: How do human beings survive with a genome horribly cluttered by seemingly useless, parasitic bits of DNA? Why is it so hard to clone an adult animal yet so easy to

clone an embryo? Why do certain traits skip generations in an apparently unpredictable way? Next month the conclusion to this article will report on the latest discoveries about how the chromosomal layer of epigenetic phenomena works and on the initial attempts to exploit epigenetics in medicine and biotechnology. m

W. Wayt Gibbs is senior writer.

MORE TO EXPLORE

- Non-Coding RNA Genes and the Modern RNA World. Sean R. Eddy in *Nature Reviews Genetics*, Vol. 2, pages 919-929; December 2001.
- An Expanding Universe of Noncoding RNAs. Gisela Storz in *Science*, Vol. 296, pages 1260-1263; May 17, 2002.
- Widespread Occurrence of Antisense Transcription in the Human Genome. Rodrigo Yelin et al. in *Nature Biotechnology*, Vol. 21, pages 379-385; April 2003.
- Challenging the Dogma: The Hidden Layer of Non-Protein-Coding RNAs in Complex Organisms. John S. Mattick in *BioEssays*, Vol. 25, pages 930-939; October 2003.

[News Front Page](#)

Last Updated: Wednesday, 12 May, 2004, 13:56 GMT 14:56 UK

[E-mail this to a friend](#)[Printable version](#)

'Junk' throws up precious secret

By Julianna Kettlewell
BBC News Online science staff

A collection of mystery DNA segments, which seem to be critical for the survival of many animals, are causing great interest among scientists.

Researchers inspecting the genetic code of rats, mice and humans were surprised to find they shared many identical chunks of apparently "junk" DNA.



Humans and rats share large amounts of DNA

This implies the code is so vital that even 75 million years of evolution in these mammals could not tinker with it.

But what the DNA does, and how, is a puzzle, the journal *Science* reports.

Excess baggage?

Before scientists began laboriously mapping several animal life-codes, they had a rather narrow opinion about which parts of the genome were important.

According to the traditional viewpoint, the really crucial things were genes, which code for proteins - the "building blocks of life". A few other sections that regulate gene function were also considered useful.

The rest was thought to be excess baggage - or "junk" DNA.

“ It absolutely knocked me off my chair ”

David Haussler, University of California

But the new findings suggest this interpretation was somewhat wanting.

David Haussler of the University of California, Santa Cruz, US, and his team compared the genome sequences of man, mouse and rat. They found - to their astonishment - that several great stretches of DNA were identical across the three species.

To guard against this happening by coincidence, they looked for sequences that were at least 200 base-pairs (the molecules that make up DNA) in length. Statistically, a sequence of this length would almost never appear in all three by chance.

Not only did one sequence of this length appear in all three - 480 did.

Vital function

The regions largely matched up with chicken, dog and fish sequences, too; but are absent from sea squirt and

fruit flies.

"It absolutely knocked me off my chair," said Professor Haussler. "It's extraordinarily exciting to think that there are these ultra-conserved elements that weren't noticed by the scientific community before."

The really interesting thing is that many of these "ultra-conserved" regions do not appear to code for protein. If it was not for the fact that they popped up in so many different species, they might have been dismissed as useless "padding".

But whatever their function is, it is clearly of great importance.

We know this because ever since rodents, humans, chickens and fish shared an ancestor - about 400 million years ago - these sequences have resisted change. This strongly suggests that any alteration would have damaged the animals' ability to survive.

"These initial findings tell us quite a lot of the genome was doing something important other than coding for proteins," Professor Haussler said.

He thinks the most likely scenario is that they control the activity of indispensable genes and embryo development.

Nearly a quarter of the sequences overlap with genes and may help slice RNA - the chemical cousin of DNA involved in protein production - into different forms, Professor Haussler believes.

The conserved elements that do not actually overlap with genes tend to cluster next to genes that play a role in embryonic development.

"The fact that the conserved elements are hanging around the most important development genes, suggests they have some role in regulating the process of development and differentiation," said Professor Haussler.

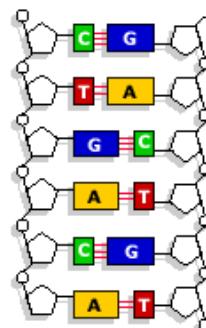
Rethinking "junk" DNA

The next step is to pin down a conclusive function for these chunks of genetic material.

One method could be to produce genetically engineered mice that have bits of the sequences "knocked out". By comparing their development with that of normal mice, scientists might be able to work out the DNA's purpose.

Despite all the questions that this research has raised, one thing is clear: scientists need to review their ideas about junk DNA.

DNA: THE CODE OF LIFE



The double-stranded DNA molecule is held together by chemical components called bases

Adenine (A) bonds with thymine (T); cytosine (C) bonds with guanine (G)

These letters form the "code of life"; there are close to 3 billion base pairs in mammals such as humans and rodents

Written in the DNA of these animals are 25,000-30,000 genes which cells use as templates to start the production of proteins; these sophisticated molecules build and maintain the body

Professor Chris Ponting, from the UK Medical Research Council's Functional Genetics Unit, told BBC News Online: "Amazingly, there were calls from some sections to only map the bits of genome that coded for protein - mapping the rest was thought to be a waste of time.

"It is very lucky that entire genomes were mapped, as this work is showing."

He added: "I think other bits of 'junk' DNA will turn out not to be junk. I think this is the tip of the iceberg, and that there will be many more similar findings."

[E-mail this to a friend](#)

[Printable version](#)

[Back to top](#)

Source: [Harvard Medical School](#)

Date: 2004-06-03

Junk DNA Yields New Kind Of Gene: Regulates Neighboring Gene Simply By Being Switched On

BOSTON -- In a region of DNA long considered a genetic wasteland, Harvard Medical School researchers have discovered a new class of gene. Most genes carry out their tasks by making a product-a protein or enzyme. This is true of those that provide the body's raw materials, the structural genes, and those that control other genes' activities, the regulatory genes. The new one, found in yeast, does not produce a protein. It performs its function, in this case to regulate a nearby gene, simply by being turned on.

Joseph Martens, Lisa Laprade, and Fred Winston found that by switching on the new gene, they could stop the neighboring structural gene from being expressed. "It is the active transcription of another gene that is regulating the process," said Martens, HMS research fellow in genetics and lead author of the June 3 *Nature* study .

"I cannot think of another regulatory gene such as this one," said Winston, HMS professor of genetics. The researchers have evidence that the new gene, SRG1, works by physically blocking transcription of the adjacent gene, SER3. They found that transcription of SRG1 prevents the binding of a critical piece of SER3's transcriptional machinery.

The discovery raises tantalizing questions. How does this gene-blocking occur? Do other regulatory genes work in this fashion? Does the same mechanism occur in mammals, including humans?

At the same time, SRG1 provides clues to a recent puzzle. Researchers have lately begun to suspect that the long stretches of apparently useless, or junk, DNA might possess a hidden function. In the past year, evidence has been pouring in, not just from yeast but from mammals, that these apparently silent regions produce RNAs, which are associated with transcriptional activity (see Focus, March 5, 2004 http://focus.hms.harvard.edu/2004/March5_2004/biological_chemistry.html). Yet no one has found associated protein products. "For us it is easy to look at those findings and say, 'Well maybe those are examples of what is going on here in yeast,'" said Martens.

If so, the findings would carry an important message for the post-human genome era--namely, that researchers' attempts to turn the masses of data churned out by the Human Genome Project into an understanding of what actually happens in the human body may be even more complex than they anticipated. One of the main challenges for that effort is to figure out how and when genes are turned on and off during normal development and disease. Rather than look only at how genes are regulated by proteins, they would have to turn their attention to a new, and possibly more-difficult-to-detect form of control. And given that junk DNA makes up 95 percent of chromosomes, the mechanism could be fairly common.

"I think if nothing else, this sends up an alert that this likely occurs in other cases," said Winston. "When people are looking to understand regulation of genes from whatever organism--humans, flies, mice, yeast--they cannot just look for proteins that are acting there. It might be that it is simply the act of transcribing that is causing regulation."

Like many researchers, Winston and his colleagues may have known in the back of their minds that someday they would have to contend with junk DNA, but it was not their intention to map a new gene in those wild and relatively uncharted regions of the chromosome. If anything, the yeast SER3 gene was their lodestar. What intrigued them about the gene, which is involved in the synthesis of the amino acid serine, was its unusual expression pattern. To be turned on, genes must first be bound by an activator molecule. A common activator in yeast is a molecule called Switch/Sniff. While most genes are turned on by Switch/Sniff, SER3 is turned off by the complex.

In the course of exploring how this repression happens, Martens came across an even more surprising result. "The usual story when a gene is transcriptionally repressed is that RNA polymerase, TATA binding protein and

Related News Stories



[UCR Chemist Part Of Team Identifying New Areas Of Gene Regulation](#) (May 7, 2005) --

Researcher Kangling Zhang at the University of California, Riverside is part of a team that has discovered a new way that yeast governs genetic expression and repression, a finding that could be ... > [full story](#)

[Yale Yeast Genome Study Could Lead To Further Understanding Of Diseases Like Cancer](#)

(December 1, 1999) -- In the largest genome project of its kind to date, Yale scientists have succeeded in fully characterizing the function of the yeast gene, furthering understanding of how human cells work. ... > [full story](#)

[Novel Means For Stopping Transcription Found, Hopkins Scientists Report](#) (September 26, 2001)

-- Researchers have discovered what is believed to be a novel method in yeast for governing gene expression at the end, rather than the beginning, of transcription, the process of reading DNA to make ... > [full story](#)

[Family Of Life-extending Genes Discovered; Discovery Sheds Light On How A Low-calorie Diet Extends Lifespan](#) (July 29, 2005) -- Mice, rats, worms, flies, and yeast all live longer on a low-calorie diet, which also seems to protect mammals against aging-related diseases. A gene called SIR2 is thought to control this process. ... > [full story](#)

> [more related stories](#)

a host of other factors associated with active transcription, will not be there," he said. He, Laprade, a research associate, and Winston conducted a series of experiments and found that the factors were all present and active, and they were located just upstream of the SER3 promoter-as was a jot of DNA needed for the onset of transcription, the TATA element.

Thinking that the TATA element might signify the beginning of a new gene, one associated with both the active RNA polymerase and SER3 repression, Martens mutated it. "We no longer saw the RNA, and we found transcription of SER3 was de-repressed," he said. "That is when we thought, 'OK, we have got a new regulatory gene.'" After characterizing SRG1, which turned out to be 550 base pairs long, they tackled the question, How is it regulating SER3? They put the question on the table during a lab retreat atop a downtown skyscraper. "Everybody talks, and they are not allowed to show any data," said Winston. Out of that intellectual free-for-all, three models emerged.

The first held that RNA transcripts produced from SRG1 were being recruited to SER3 and were somehow repressing transcription. The researchers assumed that if this were true, it would not matter where the RNA came from. As it turned out, SER3 was repressed only when the RNA was produced by an adjacent SRG1. The second model, which proposed that the SRG1 promoter outcompeted the SER3 promoter for transcription factors, also did not hold up to experimental scrutiny.

There had been hints all along favoring the third model. In this one, transcription of the nearby SRG1 somehow prevents an activator from binding the SER3 promotor. Using chromatin immunoprecipitation, a powerful method for imaging the location of molecules in living cells, the researchers found that this was exactly what happened: a well-known activator fell off the SER3 promotor when SRG1 was turned on. In fact, when SER3 was replaced by a reporter gene, the same thing happened-the turning on of SRG1 prevented the activator from binding to that gene as well.

As for how this interference actually occurs, one possibility is that the machinery required to transcribe SRG1 -RNA polymerase, TATA-binding proteins and other factors-somehow spills over to the nearby SER3 promotor, physically preventing it from being approached by an activator. "It is also possible that active transcription alters chromatin structure and modifies things in other ways," said Winston. As for the molecule that got them started in the first place, Switch/Sniff, the researchers now think it may activate SRG1 and in that way bring about SER3's anomalous repression. "That is our current thinking," Winston said. It is a view he expects will be revised. "Every time we thought we understood everything going on here, we have been wrong. There are additional layers of complexity."

This story has been adapted from a news release issued by Harvard Medical School.



October 20, 2004

Mice do fine without 'junk DNA'

Deleting non-coding regions from the genome has no apparent effect.

by Roxanne Khamsi
news@nature.com

Mice born without large portions of their 'junk DNA' seem to survive normally. The result contradicts the beliefs of many scientists who have sought to uncover the function of these parts of the genome.

More than 90% the genome of organisms such as mice and humans does not appear to code for any proteins. And yet this DNA shows striking similarities between species. If they had no function, over time mutations would scramble the sequences. Why have these bits of the genome remained so highly conserved?

One study, published this month in *Developmental Cell*, reports that parts of the non-coding DNA may be involved in embryonic development. Barbara Knowles and her colleagues from the Jackson Laboratory in Bar Harbor, Maine, found that non-coding regions known as transposable elements, which can regulate genes, are highly active in mouse embryos.

"Survival in the laboratory for a generation or two is not the same as successful competition in the wild for millions of years."

David Haussler

University of California, Santa Cruz

Knowles speculates that transposable elements could control embryonic differentiation, activating or reprogramming parental chromosomes. "I think they contain controlling sequences," she says.

Take out the trash

But transposable elements are only a small part of the non-coding regions. And now Edward Rubin's team at the Lawrence Berkeley National Laboratory in California has shown that deleting large sections of non-coding DNA from mice appears not to affect their development, longevity or reproduction.

The team created mice with more than a million base pairs of non-coding DNA missing - equivalent to about 1% of their genome. The animals' organs looked perfectly normal. And of more than 100 tests done on the mice tissues to assess gene activity, only two showed changes. The results are reported in this week's *Nature*.

The group has now created mice missing three million base pairs. "We can see no effect in them," Rubin says.

Tough test

Knowles cautions that the study doesn't prove that non-coding DNA has no function. "Those mice were alive, that's what we know about them," she says. "We don't know if they have abnormalities that we don't test for."

David Haussler of the University of California, Santa Cruz, who has investigated why genetic regions are conserved, says that Rubin's study gives no hint that the deleted DNA has a function. But he also believes that non-coding regions may have an effect too subtle to be picked up in the tests to far.

"Survival in the laboratory for a generation or two is not the same as successful competition in the wild for millions of years," he argues. "Darwinian selection is a tougher test."

1 Peaston A. E., et al. *Developmental Cell*, 7. 597 - 606 (2004).

2 Nóbrega M. A., et al. *Nature*, 431. 988 - 993 (2004).

Article Copyright © 2004 MacMillan Publishers Ltd. All rights reserved. This material may not be published, broadcast, rewritten, or redistributed.



Medical News Today



www.medicalnewstoday.com

Genomics study highlights the importance of "junk" DNA in higher eukaryotes

14 Jul 2005

A landmark comparative genomics study appears online today in the journal *Genome Research*. Led by Adam Siepel, graduate student in Dr. David Haussler's laboratory at the University of California, Santa Cruz, the study describes the most comprehensive comparison of conserved DNA sequences in the genomes of vertebrates, insects, worms, and yeast to date.

One of their major findings was that as organism complexity increases, so too does the proportion of conserved bases in the non-protein-coding (or "junk") DNA sequences. This underscores the importance of gene regulation in more complex species.

The manuscript also reports exciting biological findings regarding highly conserved DNA elements and the development of a new computational tool for comparing several whole-genome sequences. It was authored by multiple investigators from leading research institutions, including Penn State University (University Park, PA), Washington University School of Medicine (St. Louis, MO), Baylor College of Medicine (Houston, TX), and the University of California, Santa Cruz.

One of the most powerful approaches for pinpointing biologically relevant elements in genomic DNA is to identify sequences that are similar across multiple species. Such approaches are particularly useful for analyzing non-protein-coding sequences - sometimes called "junk" DNA. Although "junk" DNA is poorly understood, the increasing availability of whole-genome sequences is rapidly enhancing the ability of scientists to ascertain the biological significance of these non-protein-coding regions.

"Looking for functional elements in mammalian and other vertebrate genomes is like looking for needles in a haystack," explained Siepel. "By focusing on conserved elements, you get a much smaller haystack. It's not guaranteed to have every needle in it, and not everything in it is a needle, but you're much more likely to find a needle if you look in this smaller haystack than if you look in the big one."

Siepel's team aligned whole-genome sequences for four groups of eukaryotic species (vertebrates, insects, worms, and yeast). The vertebrates included human, mouse, rat, chicken, and pufferfish, and the insects included three species of fruit fly and one species of mosquito. Two worm species and seven yeast species rounded out the set.

To help ease the gargantuan task of identifying conserved elements in multiple alignments of whole-genome sequences, the researchers developed a new computational tool called phastCons. In contrast to traditional tools that compute conservation levels based on sequence similarity at each nucleotide position, phastCons allows for multiple substitutions per site, accounts for unequal rates of substitutions for different nucleotides, and considers the phylogenetic relationships of the species involved.

After applying phastCons to multiple alignments of each of the four groups of eukaryotic species, the researchers estimated that only between 3-8% of the human genome was conserved in the other

vertebrate species. On the other hand, the more compact genomes of insects were more highly conserved (37-53%), as were those of worms (18-37%) and yeast (47-68%).

The scientists also observed that the proportion of conserved sequences located outside of protein-coding regions tended to increase with genome length and with the species' general biological complexity.

Most strikingly, the researchers discovered that two-thirds or more of the conserved DNA sequences in vertebrate and insect species were located outside the exons of protein-coding genes, while non-protein-coding sequences accounted for only about 40% and 15% of the conserved elements in the genomes of worms and yeast, respectively.

"The conserved noncoding story seems to be fairly similar in vertebrates and insects, but looks quite different in worms and yeast," explained Siepel. "These findings support the hypothesis that increased biological complexity in vertebrates and insects derives more from elaborate forms of regulation than from a larger number of protein-coding genes." He noted that the results for the worm group should be interpreted cautiously because the analysis was based on the genomes of only two quite divergent worm species.

"We still understand remarkably little about the function and evolutionary origin of these elements," Haussler added. But the locations of the conserved elements will provide the scientists with some key clues to the potential functions of these sequences.

Some of the strongest sequence conservation in vertebrates was observed in the 3' untranslated regions (3'UTRs) of genes, which indicates that post-transcriptional regulation may be a widespread and important phenomenon in more complex species. The scientists found positive associations between highly conserved elements (HCEs) in known genes and RNA editing, as well as between HCEs and microRNA targets.

Interestingly, the researchers discovered that many HCEs in vertebrates may encode functional RNAs. The HCEs in introns and intergenic regions in vertebrates were significantly enriched for statistical evidence of local RNA secondary structure, which indicates that many may function as RNA genes.

"There really does seem to be a lot more going on at the RNA level than people would have guessed a few years ago," commented Siepel.

HCEs were also associated with "gene deserts" - long regions of the genome that are devoid of protein-coding genes. This indicates that some of the conserved elements may function as long-range transcriptional regulatory elements.

For genomic scientists, the current study is a major contribution to the field. Not only will the new bioinformatics tool phastCons help researchers identify evolutionarily conserved DNA elements, the reported conserved elements are represented as conservation tracks in the widely used UCSC Genome Browser. "With phastCons and with the conservation tracks in the browser," says Siepel, "we're trying to make it as easy as possible for researchers to home in on functionally important DNA sequences."

Adam Siepel, first and corresponding author on the manuscript, has agreed to be contacted by e-mail (acs@soe.ucsc.edu) or by phone (+1-831-423-0863) for further information. David Haussler, Ph.D., is the principal investigator on this work and can be reached at haussler@soe.ucsc.edu or +1-831-429-9472. Information about Haussler's Genome Bioinformatics Group at the University of California, Santa Cruz can be accessed at <http://www.cbse.ucsc.edu/staff/haussler.shtml>.

This work will be published in the August print issue of *Genome Research*. It will also appear online on July 15 as a "Genome Research in Advance" article at <http://www.genome.org/cgi/doi/10.1101/gr.3715005>. Its citation is as follows:

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H. Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., Haussler, D. 2005. Evolutionary conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15: 1034-1050.

Copies of the manuscript can be obtained by contacting Maria A. Smit, Ph.D., Assistant Editor, *Genome Research*, by e-mail (smit@cshl.edu) or by phone (+1-516-422-4013).

Genome Research (www.genome.org) is an international, monthly, peer-reviewed journal published by Cold Spring Harbor Laboratory Press. Launched in 1995, it is one of the five most highly cited primary research journals in genetics and genomics. The journal publishes novel genome-based studies and cutting-edge methodologies in comparative and functional genomics, bioinformatics, proteomics, evolutionary and population genetics, systems biology, epigenetics, and biotechnology.

Cold Spring Harbor Laboratory Press is an internationally renowned publisher of books, journals, and electronic media, located on Long Island, New York. It is a division of Cold Spring Harbor Laboratory, an innovator in life science research and the education of scientists, students, and the public.

For more information, visit <http://www.cshlpress.com>.

Brain science**Helpful junk**

Jun 16th 2005

From The Economist print edition

**Brain development may be influenced by genetic parasites**

THE brain is the most complicated object known. How it gets that complicated is, however, almost completely unknown. But part of the answer may turn out to be junk—at least that is the conclusion of a study led by Fred Gage of the Salk Institute in La Jolla, California, which has just been published in *Nature*.

One of the puzzling features of the human genome is that although genes are numerous they actually form less than 5% of the DNA in a cell nucleus. The rest was thus, rather cavalierly, dubbed “junk DNA” by those who discovered it. Gradually, a role for some of this junk has emerged. In particular, parts of it regulate the activity of genes, and thus which proteins are produced and in what quantities. That has implications for what a cell does—or, to put it another way, what type of cell it is. One of the most puzzling sorts of junk, though, is something known as a LINE-1 retrotransposon. This is junk that won’t stay in one place.

Retrotransposons are sometimes known as “jumping genes”. They pop from

chromosome to chromosome with gay abandon. The assumption has been that they are genetic parasites. They resemble retroviruses, which certainly are parasites (HIV, the cause of AIDS, is a retrovirus). And the effect of a string of irrelevant LINE-1 DNA popping into the middle of a functional gene is indeed traumatic. The gene in question stops working.

The parasite hypothesis is supported by the fact that although bits of DNA that look as if they have once been part of a LINE-1 element make up 20% of the human genome (ie, they are more than four times as abundant as real genes), only 100 retrotransposons are actually able to leap around, and only ten of those leap often. By and large, the parasites have been disabled, suggesting they are such bad news that evolution has eliminated them. Dr Gage and his colleagues, however, suspect that at least some of those that have not been disabled have been allowed to live on for a purpose. Instead of being destroyed, they have been subverted—and what they have been subverted to do is to create complexity in the brain.

Light fantastic

The researchers were led to this idea when they scanned the stem-cell precursors of nerve cells with a device called a gene chip. This detects the activity of genes by measuring the presence of the molecular messengers they send into the cell to do their bidding. To their surprise, the researchers discovered a lot of LINE-1 messengers, suggesting that retrotransposons are active in these precursor cells.

To find out what was going on, Dr Gage and his colleagues built a piece of DNA that included a human LINE-1 retrotransposon; a gene for a molecule called green fluorescent protein (GFP); a genetic switch to turn the whole lot on; and a special sequence of DNA that keeps the switch in the “off” position unless the retrotransposon jumps from one place to another. The result of all this genetic engineering was a system that produces light in cells in which a retrotransposon has jumped. And GFP glows green, as its name suggests, so such cells are easy to spot.

The researchers spliced their creation into the DNA of nerve-cell precursor cells from rats (which they then grew as laboratory cultures). They also spliced it into the DNA of a line of mice, so that it was present in every cell in the mice's bodies.

Nerve-cell precursors can turn into two types of brain cell besides nerve cells. These other two types have supporting, rather than starring roles in the brain, and cannot transmit nerve impulses. The rat-cell work showed that LINE-1 jumping happens only in precursors that turn into nerve cells, and that it seems to be regulated by a protein called Sox2 that is already known to play a crucial role in the formation of nerve cells. The mouse work showed that LINE-1 was not jumping in any other parts of the body (except, oddly, the sex cells—a result that had been seen before). That suggests it is happening in the brain for a purpose.

The mouse work also showed that the retrotransposons were jumping mainly into genes that are active while precursor cells are changing into their destined cell types. The team identified a dozen and a half such genes that were affected by LINE-1, and followed up one of them, *PSD-93*, in detail. *PSD-93* makes a protein

found in the places where nerve cells touch each other and pass their signals on. When LINE-1 jumped to a location in the genome near *PSD-93* it increased production of the protein. That increase, at least in cell cultures, made it likelier that a developing precursor cell would turn into a nerve cell.

So much is observation. This is where the speculation comes in. Brain formation is an incredibly wasteful process. About half of the nerve cells created in a developing brain have died by the time that brain has formed. Many researchers think that which cells live and which die is decided by a process similar to natural selection. Cells with the right properties in the right places flourish; those without wither. But natural selection requires random variation to generate the various properties.

Retrotransposons could provide that variation, by affecting gene expression at random, depending on where they pitch up. Changing the quantities of proteins such as the one made by *PSD-93* would probably change the nature of the affected cell quite radically, and might even be responsible for generating the different types of nerve cell that are known to exist. Certainly, the brain has many more different types of cell in it than any other organ.

A similar idea about generating variation has been proposed in the past, to explain the activity of LINE-1 elements in sex cells. This, the theory goes, would bolster variety in an individual's offspring above and beyond the variation already provided by sexual reproduction. That is an interesting idea. But the thought that the complexity of the mammalian brain—and the existence of human intelligence—depends on variety induced by a tamed genetic parasite is truly audacious. Whether it is true remains to be seen.

Des biologistes américains ont mis en évidence le rôle de régions non codantes du patrimoine génétique dans les différences d'organisation sociale de deux groupes de campagnols

Les zones présumées inutiles de l'ADN conditionneraient certains comportements

LE MONDE | 10.06.05 | 13h10 • Mis à jour le 10.06.05 | 15h51

Le point de départ des travaux de Larry Young et Elizabeth Hammock, chercheurs à l'université Emory (Atlanta, Etats-Unis), se résume en une question très simple : quelle est la différence entre le campagnol de prairie et son très proche cousin montagnard ? Le néophyte n'en voit aucune à l'oeil nu. Mais l'éthologue sait que les mâles de ces deux groupes de rongeurs, bien que très semblables, ont des comportements sociaux distincts. Voire opposés.

Le premier, expliquent les deux chercheurs, vendredi 10 juin, dans la revue américaine *Science*, lorsqu'il se met en couple, "*demeure fidèle à une unique partenaire tout au long de sa vie*". Il montre de plus un "*haut niveau d'intérêt social*". Le campagnol mâle montagnard, au contraire, ne se met pas durablement en couple, "*ne contribue pas à l'attention parentale*" portée à la progéniture. Il semble, pour finir, se désintéresser de toute forme d'interaction sociale avec ses congénères.

Selon les travaux des deux chercheurs américains, ces grandes différences comportementales tiennent, partiellement au moins, à des facteurs génétiques. Plus précisément, c'est une partie de l'ADN dit "*poubelle*" dont les biologistes ont longtemps cru qu'il était totalement inutile qui serait impliqué dans les mécanismes physiologiques expliquant une telle variabilité. Or l'"*ADN-poubelle*" (junk-DNA en anglais), plus justement appelé ADN non codant, ne permet pas la synthèse d'une protéine particulière : les deux biologistes n'ont donc pas identifié un hypothétique "*gène de la sociabilité*" qui serait présent dans un cas et absent dans l'autre.

En étudiant le patrimoine génétique de chaque groupe de rongeurs, les chercheurs ont détecté des différences au sein du gène V1aR, présent chez les individus des deux groupes. V1aR code pour la synthèse de récepteurs de la vasopressine une hormone connue pour être impliquée dans les comportements sociaux des rongeurs, mais aussi des primates.

D'un groupe à l'autre, les différences au sein de ce même gène sont ténues. Pour comprendre, il faut savoir que, dans une majorité de séquences génétiques, cohabitent des régions codantes les exons et d'autres non codantes les introns. Chez le campagnol de prairie, les introns du gène V1aR sont formés de séquences beaucoup plus longues que chez son cousin vivant sur les reliefs. Les chercheurs ont observé *in vivo* que ces variations de la longueur des introns influent sur l'expression du gène V1aR dans certaines zones du cerveau.

La mise en évidence d'un effet de l'ADN-poubelle sur le comportement social des individus est originale. "*Les séquences non codantes ne sont pas fortement soumises à la pression de la sélection. Elles sont donc très différentes d'un individu à l'autre, explique Pierre Roubertoux, chercheur (CNRS) au laboratoire Génomique fonctionnelle. Une conséquence très importante est l'effet de ce type de processus sur le plan évolutif. Les variations aléatoires de ces régions non codantes vont servir à créer, en quelque sorte, un réservoir de variabilité au sein d'une même espèce.*"

Comment des séquences non codantes peuvent-elles influencer l'expression de certains gènes ? Le processus mis en cause, explique M. Roubertoux, se nomme "*épissage alternatif*". "*Lors de la transmission de l'information génomique de l'ADN à l'ARN -qui est ensuite directement impliqué dans la synthèse des protéines-, on assiste à un phénomène d'"élimination" des introns, pour ne conserver que les exons, c'est-à-dire les seules régions codantes, ajoute M. Roubertoux. Mais, au cours de ce processus, les introns peuvent entraîner avec eux des séquences codantes.*" L'épissage alternatif introduit donc une déperdition aléatoire de l'information servant à coder les protéines. Le mécanisme induit une variabilité des molécules synthétisées par un gène. "*A partir du gène de la drosophile qui gère la poussée des neurones, on peut, par exemple, obtenir 32 000 protéines différentes*", explique M. Roubertoux.

L'implication de l'ADN non codant dans le développement de certaines pathologies neurologiques est soupçonnée depuis plusieurs années. C'est en effet ce même mécanisme d'épissage alternatif qui est généralement invoqué lorsque certaines protéines sont exprimées de manière tronquée.

Les mécanismes mis en lumière par Larry Young et Elizabeth Hammock ne dévoilent pas tous les secrets de ces régions du génome qui forment la plus grande part du patrimoine génétique des mammifères. Alors que des individus amputés d'une proportion importante de cet ADN parviennent à vivre sans dommages apparents (Le Monde du 23 octobre 2004), certaines séquences, bien que non codantes, présentent une grande stabilité au fil de l'évolution (Le Monde du 4 octobre 2003) et se retrouvent chez de nombreuses espèces. Accréditant ainsi, au contraire, l'idée qu'elles sont très utiles...

Stéphane Foucart

Article paru dans l'édition du 11.06.05

Frankfurter Allgemeine Zeitung
Samstag, 18. Oktober 2003, Nr. 242
Seite 36, Feuilleton
Rätselhafte Wüsten im Erbgut
Was aus Müll alles werden kann: Steuerung lebenswichtiger Gene

Daß die Gene im Erbgut des Menschen und aller anderen höheren Lebewesen gleichsam wie winzige Inseln in einem riesigen Meer vermeintlich nichtssagender Basensequenzen liegen, hatte die Forscher vor vielen Jahren höchst überrascht. Weil man sich nicht vorstellen konnte, daß diese scheinbar informationsleeren, oft monoton hunderttausendfach wiederholten Basenfolgen eine biologische Bedeutung haben könnten, hat man sie seinerzeit

abfällig genetischen "Müll" ("junk") genannt. Diese abwertende Bezeichnung hat viele Wissenschaftler jahrelang davon abgehalten, sich diese erstaunlichen Basensequenzen näher anzuschauen. Nur wenige fragten sich fasziniert, wozu diese nichtkodierenden Sequenzabschnitte gut sein könnten. Daß sie vollends überflüssig sein würden, wollte ihnen nicht in den Sinn. Dann hätte der Müll im Verlauf der Evolution als unnützer Ballast aus dem Genom verloren gehen müssen, argumentierten sie. Doch genau das Gegenteil ist

der Fall, besitzen Bakterien doch nur in den seltensten Fällen kleine Schnipsel nichtkodierender Basensequenzen, höhere Lebewesen hingegen besonders reichlich davon.

Seit einigen Jahren kristallisiert sich nun in der Tat immer deutlicher heraus, daß sich hinter dem vermeintlichen genetischen Müll unter anderem eine Vielzahl von Regelwerken verbirgt, die darüber entscheiden, wann welche

Gene und in welchem Ausmaß aktiviert oder lahmgelegt werden. Über den neuesten Puzzlestein aus dem Schatzkästchen des Genoms berichten Marcelo Nobrega und seine Kollegen vom amerikanischen Department of Energy Joint Genome Institute im kalifornischen Walnut Creek in der aktuellen Ausgabe der

Zeitschrift "Science" (Bd. 302, S. 413).

Die Forscher nahmen ein bestimmtes Gen (das sogenannte "Dach-Gen") genauer unter die Lupe, von dem bekannt war, daß es an seinem vorderen und seinem hinteren Ende jeweils von einem besonders langen nichtkodierenden Abschnitt eingehaumt ist. Während der Sequenzabschnitt vor dem Gen rund doppelt so lang ist wie das Gen selbst, beträgt die nichtkodierende Sequenz im direkten

Anschluß an das Gen mit seinen mehr als tausend Basenpaaren sogar das Dreifache der Gengröße. Um sich im Dickicht der unzähligen nicht kodierenden

Sequenzen einen besseren Überblick zu verschaffen, hatten die Genforscher schon vor längerer Zeit die nichtkodierenden Abschnitte nach verschiedenen Gesichtspunkten geordnet. Bereiche, in die nur wenige echte Gene eingestreut

waren, hätten sie als "Wüsten" bezeichnet. Zu dieser Kategorie gehören auch die vor und hinter dem Dach-Gen liegenden Basensequenzen. Nobrega war aufgefallen, daß in dem üblichen Regulationsbereich unmittelbar vor dem Dach-Gen erstaunlich wenig Strukturen für die genetische Steuerung dieses Gens zu erkennen waren. Das überraschte um so mehr, als das Gen während der Embryonalentwicklung bei den unterschiedlichsten Prozessen mitwirkt, bei der

Entwicklung des Gehirns ebenso wie beim Ausformen der Gliedmaßen sowie bei der Differenzierung der Sinnesorgane. Das ließ auf komplizierte Steuerungsmechanismen für dieses Gen schließen. Die Forscher wagten daher die Hypothese, daß sich in den genetischen Wüsten, die das Gen einrahmen, womöglich weitere Regelwerke befinden.

Bestärkt wurden sie in dieser Vorstellung durch den Nachweis, daß das Dach-Gen bei der Maus, die dem Menschen im Aufbau des Genoms recht ähnlich ist, von ganz ähnlichen Wüsten-Sequenzen umgeben ist. Dagegen ergab der Vergleich mit den Sequenzen beim Frosch und verschiedenen Fischen viel

weniger Ähnlichkeit. In den wenigen auch hier übereinstimmenden Sequenzbereichen fanden sie schließlich den Schlüssel zu dem vermuteten Regelwerk. Bauten sie diese Basensequenzen in Mäusezellen jeweils vor ein klassisches genetisches Kontrollelement, das mit einem leicht zu überschauenden Gen verbunden war, kam es zur Verstärkung der Genaktivität. Versuche mit transgenen Mäusen zeigten, daß dieser Verstärkereffekt in jeweils ganz bestimmten Geweben zutage trat. In einigen Fällen spiegelt diese Gewebespezifität die des Dach-Gens in seiner natürlichen genetischen Umgebung.

Wie Abschnitte im Genom, mehr als tausend oder gar Millionen Basensequenzen entfernt, die Aktivität eines Gens zu steuern vermögen, ist noch weitgehend rätselhaft. Die Forscher vermuten, daß unter anderem im linearen DNS-Molekül

weit voneinander entfernt liegende Basensequenzen sich durch das Auffalten des Erbmoleküls zum sogenannten Chromatin der Chromosomen räumlich nahe kommen. Mit der Unterstützung von Proteinen fällt es ihnen dann leicht, miteinander kommunizieren, so daß die Aktivität von Genen subtil reguliert werden kann. BARBARA HOBOM

21.11.2003 - Genforschung

Wie sich Junk-DNA durchsetzt

Forscher: Bei kleinen Populationsgrößen bestimmt der Zufall, bei größeren der Konkurrenzdruck die Genomgröße

Amerikanische Wissenschaftler haben eine neue Hypothese aufgestellt, warum Menschen, Tiere, Pflanzen und Pilze im Gegensatz zu Bakterien zum Teil anscheinend überflüssige DNA-Abschnitte, so genannte Junk-DNA, besitzen: Die Genetiker fanden einen Zusammenhang zwischen der Populationsgröße einzelner Arten und der Größe des Genoms. Je kleiner die Population, desto mehr überflüssige DNA-Bereiche finden sich außerdem im Erbgut der Spezies. Diese Vermutung äußern die Forscher in der Fachzeitschrift [Science](#) (Ausgabe vom 21. November).

Das Genom von komplexeren Organismen, so genannten [Eukaryoten](#), besteht nicht nur aus Genen, welche Informationen für Eiweißmoleküle enthalten, sondern auch aus so genannten nicht-codierenden Bereichen. Die Funktion dieser DNA-Abschnitte ist häufig noch völlig unbekannt, und einige sind offenbar vollständig überflüssig. Dagegen gibt es im Genom von Bakterien ausschließlich DNA, die unverzichtbare Informationen enthält. [Michael Lynch](#) von der Universität von Indiana in Bloomington und [John Conery](#) von der Universität von Oregon in Eugene erklären diesen Unterschied mit den verschiedenen Populationsgrößen der Spezies.

Während bei Organismen mit großen Populationsgrößen die Konkurrenz zwischen den einzelnen Individuen sehr groß ist und damit ein starker Selektionsdruck herrscht, wird bei komplexeren Organismen, die kleinere Populationsgrößen haben, der Lauf der Evolution hauptsächlich durch zufällige genetische Veränderungen bestimmt. Daher können sich bei diesen Spezies zum Teil überflüssige DNA-Bereiche ansammeln, schreiben die Wissenschaftler.

Um diese These zu überprüfen, bestimmten die Wissenschaftler an den Genomen von dreißig Eukaryoten die Anzahl der Gene und den Anteil der unnötigen DNA-Sequenzen. Im Vergleich mit den genetischen Daten von Bakterien ergab sich ein Muster: **Je kleiner die Bevölkerung, desto größer ist die Zahl der Gene und auch die Zahl der überflüssigen DNA-Stränge.**

ddp/bdw - Sandra Saladin

© wissenschaft.de, Konradin Relations GmbH 2005

GENFORSCHUNG

Schrottgesteuerter Anfang

Junk-DNA-Elemente regulieren früheste Embryonalentwicklung mit
[www.wissenschaft-online.de/abo/ticker/761702]

Nicht gerade ein Muster an Effektivität, unsere DNA: zwei Prozent sinnreiche Sätze Bauanleitung, überschwemmt von 98 Prozent unverständlicher Buchstabensuppe. Oder scheint das nur so, weil wir nicht richtig lesen können?

Es ist schon zeitsparend und clever, alles für überflüssig zu erklären, was man nicht auf Anhieb versteht. In der Biologie allerdings bringt einen das selten sehr weit: Irgendwie scheint alles Nutz- und Zwecklose in der Natur doch Sinn zu haben - und irgendwann, umso deutlicher, je intensiver man es untersucht, eine bislang unentdeckte Funktion zu offenbaren.

Zum Beispiel die Junk-DNA.

Sie erhielt ihren wenig schmeichelhaften Namen (*junk* = "Schrott") in Zeiten, in denen Genom-Forscher etwas voreilig zu meinen begannen, man wisse bereits ziemlich viel über die innere Organisation des Erbgutes von Maus, Mensch, Mais und Co. Als Junk-DNA bezeichneten sie also jene Abschnitte des Erbgutes der höheren, eukaryotischen Organismen, die keine für Proteine kodierenden oder für Ablese- und Kopiervorgänge notwendige Anleitungen enthalten. Alles andere schien überflüssiger, auf unbekannte Weise im Laufe der Evolution in den Genen angesammelter Schrott.

Was dann allerdings ziemliche Berge an Schrott auftürmte: Das Genom des Menschen enthält, wie man heute nach mühevoller Entzifferung weiß, 98 Prozent vermeintlichen Ausschuss - nur zwei Prozent, vielleicht 30 000 Gene, sind dieser Rechnung zufolge zur Eiweißproduktion von Nutzen. Klingt etwas zu verschwenderisch, um wahr zu sein.

Und wirklich, nach und nach entdeckte man bei genauem Hinsehen mehr und mehr interessante Bewegung im vorgeblichen Müll: Rund ein Drittel davon machen bei Mensch und Maus etwa Transposons aus, mobile genetische

Elemente, die innerhalb der DNA umziehen können, indem sie sich selbst aus dem Erbgut-Molekül herausschneiden, kopieren und andernorts wieder integrieren. Bei Maus und Mensch sind fast alle dieser Elemente so genannte Retrotransposons, ein Zehntel - die so genannten LTR-(*long terminal repeat*)-Retrotransposons - degenerierten offenbar, wie ihre Sequenzen nahe legen, in grauer Vorzeit einmal aus einem gemeinsamen Retrovirus-Stammvater.

Eben dieser wohl retroviralen Nachkommen widmeten sich nun Anne Peaston und ihre Kollegen vom Jackson-Laboratorium sowie Davor Solter vom Freiburger Max-Planck-Institut für Immunobiologie. Auf der Suche nach einer bislang unerkannten Funktion der scheinbar sinnlosen Erbgut-Herumtreiber richteten die Forscher ihren Blick zum Anfang allen Lebens: auf die gerade erst ausgereifte sowie die ganz frisch befruchtete Eizelle.

Tatsächlich sind einige Retrotransposons des LTR-Typus, wie die Wissenschaftler bei ihrer Analyse des Genexpressionsmusters der Oozyte fanden, besonders aktiv in reifen Maus-Eizellen, die auf ein befruchtendes Spermium warten; andere wiederum begannen sich erst nach wenigen Zellteilungen intensiv zu rühren. Insgesamt scheint in der ganz frühen Entwicklung jedenfalls die Retrotransposon-Unruhe deutlich erhöht. Nur - warum?

Wie die Wissenschaftler erkannten, fügen sich sich die herumhüpfenden Transposons offenbar durchaus nicht wahllos in die verfügbare DNA-Masse. Vielmehr springen sie offenbar an Orte, in denen sie als Alternative zu regulatorischen Promotoren - den genetischen Konstruktions-Startbefehlen - von im Zuge der frühen Entwicklung wichtigen Eiweißen fungieren können. Auf diese Weise regulieren die Transposons also durchaus gezielt die Proteinherstellung in Eizelle und frühem Embryo, wie die Forscher zu ihrer Überraschung feststellten. Das erste Beispiel für eine synchronisierte, entwicklungsregulierte Expression mehrerer Gene durch transposable Elemente - möglicherweise trügen die Retrotransposons ein Gutteil dazu bei, das embryonale Genom der Säugetiere für die folgenden Entwicklungsschritte zu reprogrammieren, so die Wissenschaftler.

Damit könnte nun eine wichtige Aufgabe für gleich ein Drittel der Junk-DNA-Massen aufgedeckt sein. Bis zu ein Drittel, eben die Retrotransposons, könnte für ein Mehr an Regulations-Alternativen zuständig sein. Und so kann das weiter gehen, denn irgendwie muss ja eigentlich auch diese Retrotranspon-Alternativ-Regulation reguliert werden. So gesehen gut, das noch ein genügend großer Batzen angeblicher Sequenz-Sinnlosigkeit in der DNA für weitere zu entdeckende Aufgaben frei bleibt.

Jan Osterkamp

QUELLEN:

Developmental Cell 7: 597-606 (2004)

© spektrumdirekt

9. Februar 2005, Neue Zürcher Zeitung

Müll, der keiner mehr ist

Eine Neubewertung der sogenannten «junk-DNA»

Bis vor kurzem galt DNA, die keine Gene oder Kontrollregionen enthält, als unnützer Müll. Doch allein ihr Umfang - im menschlichen Genom immerhin 98 Prozent - sprach gegen diese Theorie. Neuere Arbeiten zeigen nun, dass gerade in diesem Müll DNA-Abschnitte mit wichtigen Aufgaben für die Regulierung der Gen-Aktivität liegen.

slz. Die Entzifferung unseres Erbguts (Genoms) schien es an den Tag zu bringen: Unser Genom gleicht einem Müllhaufen. Nach neuesten Schätzungen enthalten nur ungefähr 2 Prozent unserer DNA Abschnitte, die für Proteine codieren, also Bauanleitungen für Eiweisse darstellen. Nicht mehr als 30 000 Gene hat der Mensch, so die gegenwärtigen Schätzungen. Und der grosse Rest unserer rund 3 Milliarden DNA-Bausteine? «Unnütz und damit Müll», befanden die Experten. Entsprechend wurde dieser «Rest» in der Wissenschaftssprache zu «junk-DNA». Doch eine Vielzahl von Studien hat im letzten Jahr diese Theorie gehörig ins Wanken gebracht - und zwar gleich derart, dass das renommierte Wissenschaftsmagazin «Science» die **Neubewertung des DNA-Mülls zu einem der zehn wichtigsten Forschungsgebiete des Jahres 2004** ernannt hat.

Hochkonservierte Abschnitte

Dass zumindest ein Teil der junk-DNA kein Müll sein kann, hat vergangenes Jahr eine kalifornisch-australische Arbeitsgruppe berechnet.[1] Die Forscher hatten das Genom des Menschen mit jenem der Maus, des Huhns sowie des Kugelfisches Fugu verglichen. Dabei stellten sie fest, dass ungefähr 480 Regionen des angeblichen Mülls bei allen genannten Spezies vorkommen und sogar identisch sind. Sie müssen also in dieser Form im Wirbeltiergenom vorhanden sein, seit sich die Entwicklungswege dieser Arten vor mindestens 400 Millionen Jahren trennten. Wenn jedoch eine DNA-Region keinerlei Bedeutung für einen Organismus hat, so wird sie höchstwahrscheinlich im Laufe der Zeit durch Mutationen verändert. Die Wahrscheinlichkeit, dass eine unbedeutende und ungenutzte DNA-Region über 400 Millionen Jahre unverändert bestehen bleibt, geht stark gegen Null - die Forscher haben den sehr kleinen Wert von $10^{[2][2]}$ berechnet. Und dass ein solch unwahrscheinliches Ereignis gleich in allen 480 untersuchten Regionen eintritt, halten die Experten für nahezu ausgeschlossen. Demnach müsse es eine Funktion für derart konservierte DNA-Abschnitte geben, folgern sie.

Erste handfeste Hinweise darauf, dass im DNA-Müll wahre Schätze versteckt sein könnten, lieferte unter anderem ein amerikanisches Forscherteam aus Kalifornien und Boston.[2] Es fand bei einer detaillierten Analyse der menschlichen Chromosomen 21 und 22 heraus, dass anhand dieser Chromosomen viel mehr RNA-Moleküle produziert werden, als aufgrund der Genanzahl eigentlich zu erwarten wäre - RNA-Moleküle sind Kopien von DNA-Abschnitten, die beim Ablesen des Erbguts hergestellt werden und Informationen aus dem Zellkern ins Zellplasma transportieren. Als man nun diese RNA-Moleküle genauer unter die Lupe nahm, zeigte sich, dass viele von ihnen keine Bauanleitung für ein Protein enthielten. Es handelt sich also nicht um Boten- RNA, sondern um sogenannte nichtcodierende RNA. Doch was ist deren Funktion?

Ein Vergleich mit Bakterien half weiter. Dort existieren sogenannte antisense-RNA-Moleküle, die

nicht eine Kopie des eigentlichen Gens, sondern des zum Gen komplementären Strangs in der DNA-Doppelhelix sind - eben des sogenannten antisense-Strangs. Bei Bakterien kontrollieren solche antisense-RNA-Moleküle die Genaktivität, indem sie sich an das Gen selbst (oder an eine für das Gen zuständige Kontrollregion) anlagern und dieses so blockieren. Auch in Zellen des Wurms *C. elegans* sowie bei Pflanzen gibt es solche antisense-RNA zur Kontrolle der Genaktivität. Und tatsächlich stiessen die amerikanischen Forscher bei einer entsprechenden Analyse der nichtcodierenden RNA-Moleküle menschlicher Zellen nun ebenfalls auf antisense-Moleküle. Diese könnten an der Kontrolle der Genaktivität beteiligt sein, wird vermutet. Ob dies aber zutrifft und wie genau diese antisense-RNA-Moleküle menschliche Gene dann kontrollieren würden, ist noch offen.

Räumliche Blockade

Eine weitere Form der Genkontrolle könnte eine zweite Gruppe von nichtcodierenden RNA-Molekülen ausüben. Diese werden von einem junk-DNA-Abschnitt abgelesen, der direkt neben einer Kontrollregion für ein Gen liegt oder sogar mit dieser überlappt. Es wäre also möglich, so spekulierten die Entdecker dieser RNA, dass bei der Herstellung einer solchen nicht-codierenden RNA die dazu nötige Produktionsmaschinerie den Kontrollpunkt für das eigentliche Gen räumlich blockiere. Auf jeden Fall gingen die Forscher davon aus, dass die nichtcodierende RNA nicht ohne einen Zweck gebildet wird - das wäre für die Zelle nämlich viel zu energieaufwendig.

Erste konkretere Hinweise auf eine solche räumliche Blockade fanden Forscher um den Genetiker Fred Winston aus Boston schliesslich im vergangenen Sommer.[3] In Hefezellen wird nämlich eine nichtcodierende RNA hergestellt, der man den Namen SRG1 gab. Wenn diese RNA anwesend ist, kann das Gen namens SER3 nicht abgelesen werden. Durch eine Vielzahl von genetischen Experimenten und Tricks habe man herausgefunden, wie diese Blockade molekular ablaufe, erläuterte Winston im Gespräch. Es sei dabei tatsächlich nicht die SRG1-RNA entscheidend, sondern der Produktionsprozess derselben.

Denn um eine RNA herzustellen, wird ein grosser Komplex aus diversen Proteinen und RNA-Bausteinen gebildet, der sich dann sozusagen auf dem entsprechenden DNA-Abschnitt festklammert. Wenn nun dieser Komplex auf der SRG1- Stelle sitzt, wird das Ablesen des SER3-Gens räumlich behindert: Es kann sich kein weiterer Komplex auf der SER3-Stelle niederlassen. Nun sucht Winstons Team in Hefezellen nach weiteren RNA-Molekülen, die die Aktivierung von Genen durch eine räumliche Blockade verhindern. Laut Winston wäre es aber auch vorstellbar, dass ein solcher Mechanismus nicht nur in Hefe-, sondern auch in Tier- oder menschlichen Zellen existiert.

Darüber hinaus seien natürlich eine Vielzahl anderer Kontrollmechanismen denkbar, die von RNA-Molekülen vermittelt würden, sagte der Genetiker. So könne sich etwa eine nichtcodierende RNA direkt auf einem Startsignal für die Genaktivierung festsetzen oder bereits gebildete Boten-RNA abfangen - prinzipiell könne jeder in einer Zelle ablaufende Prozess, bei dem DNA oder RNA eine Rolle spielt, durch nichtcodierende RNA mitgesteuert werden. Denkbar ist also nicht nur eine Kontrolle bei der Genaktivierung, sondern auch bei der Proteinproduktion, der Chromosomenaufteilung während der Zellteilung oder dem Ausbessern von DNA-Fehlern.

Springende DNA-Stücke als Kontrolleure

Zusätzlich zu den DNA-Abschnitten, die als Basis für eine nichtcodierende RNA dienen, können junk-DNA-Regionen aber auch noch auf eine andere Art nützlich sein. Es existieren im DNA- Müll nämlich bewegliche Abschnitte, die innerhalb des Genoms herumspringen. Diese sogenannten Transposons besitzen die Fähigkeit, sich selber aus der DNA herauszuschneiden und an einem anderen Ort wieder einzubauen. Bei Maus und Mensch sind rund ein Drittel des DNA- Mülls solche Transposons. Vermutlich stammen diese Elemente von Retroviren, die sich in grauer Vorzeit einmal in das Wirbeltiergenom eingeschlichen haben. Doch ihre Springerei ist offenbar nicht so ziellos, wie man lange angenommen hat.[4] Bei der Untersuchung von Maus-Eizellen stellte sich nämlich heraus, dass manche Transposons in reifen, noch unbefruchteten Eizellen aktiv sind, während andere erst

einige Zellteilungen nach der Befruchtung aktiv werden. Diese Transposon springen dann offenbar an Orte, an denen sie regulatorisch wirksam sind und so für die frühe Entwicklung wichtige Proteine aktivieren.

Möglicherweise sind diese Transposons also unerlässlich, um das Säugetier-Genom von dem einer ruhenden Eizelle auf die embryonale Entwicklung umzuprogrammieren. Noch liegen allerdings zu wenig Erkenntnisse vor allem von menschlichen Eizellen vor, um genau sagen zu können, was welches Transposon zu welcher Zeit tatsächlich macht. Doch man hat bereits gefunden, dass Transposons auch andere Aufgaben, zum Beispiel bei der Nervenfunktion, erfüllen.

Noch sind alle Daten über Aufgaben des angeblichen DNA-Mülls nur erste Anzeichen für eventuelle Funktionen. Besonders über die junk-DNA im menschlichen Genom kann momentan nur spekuliert werden, da fast alle Experimente mit Zellen von anderen Spezies durchgeführt wurden. Doch eins ist klar: Der DNA-Müll hat seinen Namen zu Unrecht. Zumindest ein Teil dieses angeblich nutzlosen Ballasts im Erbgut hat wichtige Kontrollfunktionen. Allerdings lassen die Daten einer weiteren Veröffentlichung aus dem vergangenen Jahr vermuten, dass nicht der ganze angebliche DNA-Müll für einen Organismus unerlässlich ist.[5] Denn als Forscher bei Mäusen zwei grössere Abschnitte der junk-DNA entfernten, wurden derart veränderte Embryonen ohne vorgeburtliche Komplikationen geboren und lebten anschliessend beschwerdefrei. Auch ein ausführlicher Gesundheitscheck ergab keine Defekte bei den Tieren.

Um endlich mehr Klarheit in die Rolle der junk-DNA zu bringen, arbeiten immer mehr Forschergruppen weltweit daran, mögliche Funktionen dieser DNA aufzuklären. Zudem hat das amerikanische National Human Genom Research Institute in Bethesda, Maryland, ein grosses Forschungsprogramm namens **Encode** gestartet, um alle funktionalen Elemente im Erbgut aufzuspüren. Momentan will man in der Pilotphase an einem vorher festgelegten Prozent des Genoms Methoden dafür etablieren. Und man muss wohl kein Prophet sein, um vorhersagen zu können, dass im sogenannten Müll unseres Erbguts noch viele Schätze auf ihre Entdeckung warten.

Quellen: [1] Science 304, 1321-1325 (2004); [2] Cell 116, 499-509 (2004); [3] Nature 429, 571-574 (2004); [4] Developmental Cell 7, 597-606 (2004); [5] Nature 431, 988-993 (2004).

Diesen Artikel finden Sie auf NZZ Online unter: <http://www.nzz.ch/2005/02/09/ft/articleCIVR8.html>

Genius of Junk (DNA)

10 July 2003

From July 6 - 11 the world's leading geneticists gather in Melbourne for the 50th anniversary of Watson and Crick's discovery of the structure of DNA. Right in the midst of this event, Genetic Congress 2003, Catalyst reveals the extraordinary mistake made by the vast majority of the genetics community - the failure to recognise the vital importance of so-called Junk DNA.

Producer: Sonya Pemberton
Researcher: Carina Dennis and Rhian Skirving

[Transcript](#)
[Related Info](#)

Dr Malcolm Simons is an internationally recognised immunologist. A New Zealand born Australian, he has spent 30 years of his life hunting for new and better ways to diagnose disease. Along the way he has gained a reputation as somewhat of an eccentric. He has been married – and divorced – five times. He has fathered three pairs of children. He's been bankrupt, a classical piano player, and played for Australia in international squash.

Genius of Junk is the story of how Malcolm Simons turned Junk into gold, enflaming one of the greatest controversies of our time - the control and ownership of our genetic material.

It is a story of triumph and tragedy. The triumph of a man flying in the face of conventional scientific thought, facing ridicule for his ideas and living to see those ideas vindicated. The tragedy of seeing his dreams come to fruition as he faces death. For he himself has cancer, Multiple Myeloma. A fatal and incurable cancer, formed in the very Junk DNA he spent 16 years exploring.

This is also a story of genius and character. Malcolm Simons had the genius to realise that the non-coding part of our DNA wasn't in fact the junk DNA that many scientists had labelled it, but vital to the processes of life. And he has a character that fits the cliché of the eccentric scientist - brilliant at his work but hopeless with everyday life.

As he faces the greatest battle of all, Malcolm Simons takes us on a unique personal and scientific journey, to the heart of the things that matter to the very core of life - and death.

Background

Fifty years after Crick and Watson's historic discovery of DNA, the scientific community has come a long way towards unravelling the mysteries of the key to life. The Human Genome Project has mapped our entire genetic code; we are genetically modifying plants and animals, finding genetic cures for diseases. There is no doubt that our investigation of what was called the coding region of DNA has revolutionized science and the world. But the coding part of DNA makes up less than 5% of our entire genome. Because the rest of our DNA didn't seem to have any known function it was dubbed non-coding, or Junk DNA.

Malcolm Simons couldn't believe that evolution would be so wasteful. In 1987, despite having no formal training in genetics, he had a moment of remarkable insight that convinced him that Junk DNA was serving a vital function; it provided markers that indicated susceptibility to disease. At a Workshop in the United States, he saw patterns emerge from the non-coding DNA. He realised that whatever was going on in non-coding DNA was not random. Malcolm Simons, "There was order in the 95%. If there was order there was likely to be function. Maybe this was a way to also contribute to understanding the function of genes and therefore their malfunction in disease and in so doing help diagnosis - make earlier diagnosis - help save lives." When he posed his radical theory that this junk might actually have a critical role in diagnosis, his peers announced, "Malcolm, you're off your friggin' head."

Undeterred, Malcolm set out to prove that this was no junk. The majority of geneticists were focused on coding DNA, the non-coding region was left field and probably irrelevant - but for Malcolm it became the main game. Fortunately one man



believed in his genius - entrepreneur Mervyn Jacobson. In a real-life Odd Couple pairing (Mervyn is as organised and efficient as Malcolm is eccentric and chaotic) they formed Genetic Technologies in 1989. They believed that this non-coding DNA could prove valuable in diagnosing disease, perhaps in developing therapies, even cures.

There were others around the world exploring similar ideas, but Malcolm Simons took the crucial, unprecedented step - in the mid 1990s he patented the use of the so called Junk. It was an act of extraordinary and provocative foresight.

Today, \$20 million of investment later, that foresight is paying off literally. Researchers the world over are confirming that non-coding DNA holds critical clues to a vast range of diseases; breast cancer, HIV, Crohn's disease, Alzheimer's, heart disease, ovarian and skin cancer... the list is growing daily. A leading figure in world genetics, Prof. John Mattick, recently claimed that, "the failure to recognise the implications of the non-coding DNA will go down as the biggest mistake in the history of molecular biology". In the last year Genetic Technologies has signed a series of licensing deals allowing companies to use their Junk DNA patents. This is bringing in millions of dollars for the company and the profits look like they will continue to roll in. The junkyard it seems, is a goldfield.

And suddenly, the implications of the patents are staggering. Genetic Technologies controls access to 95% of DNA of every creature on earth. And they can charge license fees to anyone, anywhere in the world, working in the non-coding regions. The critics of DNA patenting are outraged. But for Malcolm Simons the vindication and controversy is somewhat meaningless. He resigned from Genetic Technologies in 2000. He no longer has any shares in the company he once co-owned, he's broke and he's grappling with the realities of final stage cancer.

Transcript

Narration: DNA. Within its exquisite structure lie the clues to our destiny... how we are formed... how we will live... and how we may die. Yet vast tracts of DNA code were dismissed by science as meaningless 'JUNK'.



But an Australian scientist saw order in this sea of chaos. He saw ways to use 'Junk' to diagnose disease and save lives. Controversially, he laid claim to the 'Junk', patenting the use of 95% percent of the DNA of all species on earth.

Yet just as his billion dollar vision is being realised Malcolm Simons's life hangs in the balance.

Ann Abrahmsen: Living with Malcolm was like living on the edge of a whirlwind...

Mervyn Jacobson: He doesn't fit the normal mould. He is definitely not a grey man in a grey suit.

Dr Malcolm Simons: I see things in black and white. I'm correctly regarded, correctly described as a maverick, pigheaded, obstinate. My experience is that everything that I've been told is absolute... written in law and certain - just isn't so. Under certain circumstances, light doesn't even travel in straight lines

Narration: Malcolm Simons is an immunologist. He has spent 30 years of his life hunting for new and better ways to diagnose disease. Along the way he has gained a reputation as somewhat of an eccentric. He has been married - and divorced - 5 times. He has fathered 3 pairs of children. He's been bankrupt, a classical piano player, and an Australian squash champion.

After years of walking an unconventional path, today Malcolm's professional and personal worlds are colliding. Malcolm the immunologist, has cancer. Multiple Myeloma - a disease of the immune system.

Today, he is undergoing a series of tests that will help him decide how best to proceed... His options are bone marrow transplant and heavy dose chemotherapy - or letting the disease take its course.

Dr Malcolm Simons: It's a fatal disease, there's no known cure. So it's about how long you don't die. My prognosis is that

the disease will kill me unless I stay alive long enough for there to be better therapies, up to and including a cure.

Narration: The hope for such therapies - even cures - could well lie in the research that Malcolm himself pioneered research into the very core of life - DNA. For woven into these delicate strands are the genes that control everything from physique to intelligence, depression to longevity. These complex threads of information also contain the defects, the mutations, that can trigger disease.

Science of genetics is the quest to understand this code of life and death. Fifty years after Watson and Cricks' historic discovery of the structure of DNA, scientists have mapped our entire genetic code. They discovered an extraordinary matrix of four molecules - represented by the letters A, T, C and G - repeated more than three billion times.

Within this matrix they explored the genes, the powerhouses of the genome... And within each gene they identified the specific sections, which they called the "coding regions" - packed with instructions for building proteins the building blocks of life.

Prof John Mattick: The primary output of the genome is to create protein. Proteins of course are the primary components of our system. They form the structural components of our hair and skin, our oxygen carrying molecules in our blood, the hormones - enzymes that digest our food and our energy metabolism. So they're critically important

Narration: The "coding regions" became the major focus of genetic research... Even though they account for less than 5% of our entire DNA. All the rest - the other 95% - was assumed to be genetic gibberish with no known function. So they called it 'non-coding' or 'Junk' DNA.

Dr Mervyn Jacobson: The word junk was applied and it stuck and people who came along thereafter saw that it was junk and took that as a message that there was no point looking in that area. So it became almost a convenience that instead of looking at 100% of a DNA you only need to worry about looking at 5%. But even that was daunting.

Prof John Mattick: What people should've done was take stock at that point. Instead they simply swept the observation under the intellectual carpet.

Narration: For decades this thinking dominated mainstream genetics. But Malcolm Simons couldn't believe that evolution would be so wasteful. He believed that non-coding DNA must serve some sort of function.

Dr Malcolm Simons: Under Darwinistic notions you would think that junk would drop off under the theory of natural selection just like species drop off if they hit ecological niches which is incompatible with survival. If they can adapt to those niches, then those that can survive and those that can't die. There's the notion. If you apply that to the DNA sequence, then the coding region genes which survived have a function and by the way the non coding sequences have survived as well. So the proposition would have to be that if they're there, they've got a function.

Narration: It was 1987 when Malcolm first glimpsed the potential of this 'Junk' DNA - it was a moment of insight that would change the course of his life. Malcolm was researching genes of the immune system called HLA; Human Leukocyte Antigens. He was attending a Workshop in the United States, where over 70 laboratories compiled genetic data from hundreds of individuals.

For a brief moment in history there was an extraordinary amount of information on both coding and non-coding regions. Malcolm was not a geneticist, but he sensed this data was important. Fearing this wealth of information might be buried, lost forever he felt driven to make sense of it.

Dr Malcolm Simons: This was likely to be the only time in the world's history that this amount of data, which was providing potentially so much first time information about DNA genetics and HLA Region, was ever going to be seen.

Narration: So began many months of obsessive research. He flew back and forth across the States, nagged those with knowledge for assistance, learnt as much as he could as fast as he could.

Dr Malcolm Simons: I slept on the floor, I was there all the time taking up 10s to 100s of hours basically moving columns and rows and when you move columns and rows the jigsaw puzzle started to develop systematic blocks of information. At that stage I didn't fully understand the significance of it.

Narration: Then came the moment of insight it dawned on him - there were patterns in amongst the chaos of code and they were being created in the non-coding region. In the so-called 'Junk'.

Dr Malcolm Simons: Junk bespoke chaos – junk indicated or implied that whatever anybody found out there in the other 95% - when they got around to looking at it - it would be chaotic, that is the sequence variations would be random. So the significance of the observation was that could not possibly be the case... that notion could be rejected and if that notion could be rejected then the question is - what is it telling you?

Narration: Our understanding of coding DNA - a mere 5% of the genome - had already transformed medicine, allowing scientists to predict, diagnose, even cure disease. But what was the potential of the other 95%? Malcolm believed this so called 'Junk' provided markers that indicated if genetic abnormalities were present. It could provide an important, new way of diagnosing disease.

Dr Malcolm Simons: What I showed was that there was order in the 95%. If there was order there was likely to be function. Maybe this was a way to also contribute to understanding the function of genes and therefore their malfunction in disease and in so doing help diagnosis – make earlier diagnosis – help save lives.

Narration: Malcolm was defying decades of scientific dogma. But – as an ex-champion squash player – he was used to a bit of opposition.

Dr Malcolm Simons: When I showed the professional geneticists the data, which indicated to me that the 95% non-coding region wasn't junk, and was ordered...The reaction was smiling disbelief at best – you're off your friggin' head and if you're any good at squash – stick to your day job.

Narration: This thinking outside the square was the hallmark of Malcolm's life in more ways than one.

Ann Abrahmsen: Mal's got an amazing free ranging brain that just loves everything, but the rest of us are not living there. It's a wonderful place to visit. I used to love to romp with him to the fields of the fields of possibility. It was great fun, but you have to live somewhere else and he wasn't concentrating on the people who were here or we weren't in focus.

Narration: Anne is the most recent of Malcolm's five ex-wives, and the mother of the youngest of his six children. He's never been short of love, but success in his personal life has been as hard to achieve as success in his intellectual one.

Dr Malcolm Simons: In the past what some people call BC (before cancer) this person was of limitless energy – zoom mode and my mother used to describe it as like trying to catch hold of the tail end of a comet.

Dr Mervyn Jacobson: Malcolm is irrepressible, challenging and that's how he is, you have to take the good with the bad.
(smiles)

Narration: One man was prepared to take Malcolm on – entrepreneur Dr Mervyn Jacobson. In 1989 they set up a company, Genetic Technologies, to test Malcolm's theory that junk wasn't junk.

Dr Mervyn Jacobson: I was intrigued. I saw it had obviously commercial potential. It was going to be a hard grind – multi year process, very expensive, but if we were successful this would be truly revolutionary - I wanted to be part of it.

Narration: They set up a diagnostic laboratory ... with what was then a novel strategy... Using non-coding DNA they believed they could invent new approaches to diagnostic testing.

Dr Mervyn Jacobson: Some people look in the coding - in other words they utilize 5% of what's there – we utilize the 5% and the other 95%. Originally it was thought all those abnormalities had to be in the coding region. It's now known that many of those abnormalities are in the non-coding region. So you have to look in both regions and in many cases, the coding region is intact – perfect – pristine and the only abnormality is in the non-coding region.

Narration: There were others around the world exploring similar ideas. But Malcolm took the crucial, unprecedented step - he turned 'Junk' into gold. In the mid 1990's - in an extraordinary act of foresight - he successfully patented the use of non-coding DNA in **every creature on earth**. For the

company it would prove a profitable – and provocative – step.

Dr Mervyn Jacobson: The patent process is more than 400 years old. It's implemented by governments. It's government law. They set the rules and it's the government who issue the patents.

Dr Malcolm Simons: I so to speak captured the 95 percent that wasn't coding. So that these inventions cover that 95 percent and according to those for whom patentable positions is not palatable, particularly the Europeans who have more difficulty accepting these processes than the North American's, we have it for all DNA in all species.

Dr Mervyn Jacobson: We took the risk. We could have in fact failed. It could have been that Malcolm's original ideas were wrong. It could have been they were right and non patentable. It could have been they were right and patentable, but someone else beat us to it.

Narration: Today, 13 years and almost 20 million dollars later, that risk is paying off – big time. Researchers the world over are confirming that non-coding DNA holds critical clues to a vast range of diseases. And suddenly, the implications of the patents are staggering - if you want to use non-coding DNA to test for diseases like breast cancer or AIDS, Genetic Technologies could demand a license fee.

It is bringing in millions of dollars for the company. Some suggest it's worth billions. The Junkyard, it seems, is a goldmine.

Dr Mervyn Jacobson: It's not true that we own the non coding DNA. We don't own any DNA. Everybody owns their own DNA – every human and every animal and every plant if you like. What we did is we applied human intelligence to a process that enabled scientists to do things they could not do before and that intelligent approach we developed is what's patentable.

Narration: Over the past year Genetic Technologies has targeted international biotech and pharmaceutical companies, perceived to be infringing their patents. And now, universities and research labs may have to pay to use Junk DNA.

Dr Graeme Suthers: These patents essentially have a potential to freeze research at a particular point because they put the control of those genes out of the public domain and keep it private. It means then that the patent holder can control who does what research on their particular gene

Narration: Mervyn Jacobsen argues that patents encourage scientific research... without a financial incentive investors would stay away in droves.

Dr Mervyn Jacobson: And patents have a very limited lifespan and then it's gone. It's a brief moment in time when you're allowed to commercialise your invention and then it's gone. It reverts to public ownership. Of course the invention is there forever for the benefit of mankind,

Prof John Mattick: Now that Genetic Technologies has started to attempt to enforce the patent – aggressively – across a range of organizations, I think that people are starting to realize the impact – potential impact of this and in a sense are starting to band together to try to find the resources that would be necessary to challenge the patent in court.

Narration: With billions of dollars potentially at stake, the battle over who can – and who cannot – have access to the non-coding regions is intensifying. Some scientists like Prof John Mattick, have a vested interest. He's researching the actual function of non-coding DNA, taking Malcolm's conviction that junk isn't junk to new levels.

John Maddick: The failure to recognize the implications of the non-coding DNA will go down I think as the biggest mistake in the history of molecular biology.

Narration: John believes that far from being 'Junk' the non-coding regions form an intricate, multilayered, operating system co-ordinating the function of all the components of the genome. The non-coding regions may in fact DRIVE the coding regions telling them what proteins to create. The importance indeed the 'genius' of so-called 'junk' may be its ability to organise and arrange the very building blocks of life.

Prof John Mattick: I think the future of molecular biology is going to be non-coding DNA. In the next 5-10 years I predict that we will have to find what all the components do – first approximation and then of course the much bigger question is how you put it together. How do you build a human? How do

you build a tree? Why are you and I different from each other but the same components? So the main game is going to be non-coding DNA – not because it's non-coding but because it actually is coding. It's coding for the information that puts you and me together.

Narration: Sixteen years after Malcolm's original insight, 'junk' DNA is now revolutionizing our understanding of genetics. All the years of hard work, and defying conventional thought, should be paying off. But for Malcolm, his life is heading in another direction. In 2000 he fell out with Mervyn and resigned from Genetic Technologies. He no longer has any shares in the company he once co-owned. Instead of getting rich he's broke. Acknowledgment it seems has come too late.

Dr Malcolm Simons: I sort of feel vindicated, but vindication isn't a comfort. I mean in a sense I've moved on. I've not only left the company, I've pretty much left that process. I guess I'm so exhausted by the entire process that recognition doesn't really matter anymore and certainly once you develop cancer it doesn't matter if – it matters even less. (laughs)

Narration: For Malcolm, his journey has taken him beyond the struggles of commerce and controversy, to a battle on a more profound front. His life has new focus as he grapples with the realities of being a patient.

Ann Abrahmsen: I remember he turned to me and he said maybe I will have to get one of these diseases in order to focus... so when he rang me and said he had cancer, I wasn't surprised. It seemed to me that it was his way of bringing at governance and an order to himself and giving himself permission to stop that frantic, frenetic, over-developed part of himself from careering from wherever it was careering to - which wasn't a good place .

Dr Malcolm Simons: As a doctor it opens up a whole different other side of the world which you have no comprehension of unless you're in the bed as a patient. No amount of standing around the bed gives anybody the slightest idea of what it means to be in bed with a cancer, living in places with people who are coping with their diagnosis of cancer. It's a transformation.

Narration: And in a final twist, Malcolm has just discovered that his own cancer – Multiple Myeloma – originated in non-coding DNA. His very disease is a strangely fitting affirmation of his own research. But this understanding comes too late to be of use to Malcolm. It's for the next generation to find a cure.

Narration: For Malcolm, his cancer has provoked an unexpected, positive outcome. It has brought his family together.

Dr Malcolm Simons: It has promoted to reappraise my life in its totality – including what I do with whatever time I've got. How I deal with life from day to day – re-prioritizing I guess back to children and families. It's a tear provoking thought each time I think about it... what are the chances that I'll see them into their teenage years are statistically not good but the flipside of that is that I'm motivated with a determination which I've never lacked to prove the numbers wrong.

TEXT: Malcolm proceeded with bone marrow transplantation. His prognosis is 6 – 18 months.

He has submitted a new idea to his lawyers. The patents are being filed... July 2003

*"We used to think our fate was in our stars.
Now we know, in large measure, our fate is in our genes."*
James D. Watson

Story Contacts

Dr Malcolm Simons

Email
P O Box 5201
North Geelong Victoria 3215
Australia
Ph: 03 5222 2999

Prof John Mattick

Email
Molecular Geneticist
University of Queensland
The University of Queensland
St Lucia, QLD 4072 Australia

Why Do Humans Have So Few Genes?

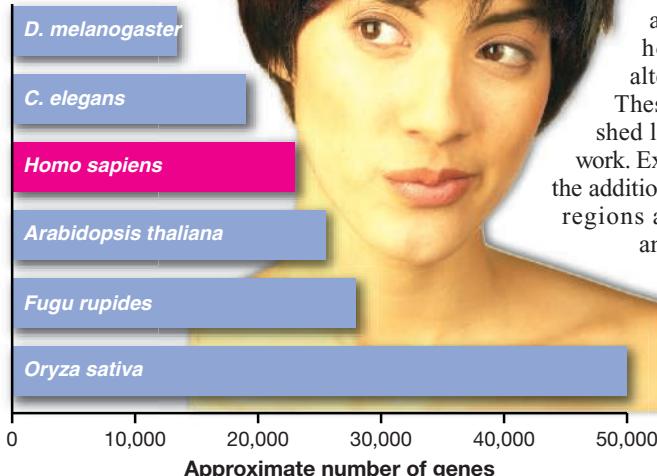
When leading biologists were unraveling the sequence of the human genome in the late 1990s, they ran a pool on the number of genes contained in the 3 billion base pairs that make up our DNA. Few bets came close. The conventional wisdom a decade or so ago was that we need about 100,000 genes to carry out the myriad cellular processes that keep us functioning. But it turns out that we have only about 25,000 genes—about the same number as a tiny flowering plant called *Arabidopsis* and barely more than the worm *Caenorhabditis elegans*.

That big surprise reinforced a growing realization among geneticists: Our genomes and those of other mammals are far more flexible and complicated than they once seemed. The old notion of one gene/one protein has gone by the board: It is now clear that many genes can make more than one protein. Regulatory proteins, RNA, noncoding bits of DNA, even chemical and structural alterations of the genome itself control how, where, and when genes are expressed. Figuring out how all these elements work together to choreograph gene expression is one of the central challenges facing biologists.

In the past few years, it has become clear that a phenomenon called alternative splicing is one reason human genomes can produce such complexity with so few genes. Human genes contain both coding DNA—exons—and **noncoding DNA**. In some genes, different combinations of exons can become active at different times, and each combination yields a different protein. Alternative splicing was long considered a rare hiccup during transcription, but researchers have concluded that it may occur in half—some say close to all—of our genes. That finding goes a long way toward explaining how so few genes can produce hundreds of thousands of different

proteins. But how the transcription machinery decides which parts of a gene to read at any particular time is still largely a mystery.

The same could be said for the mechanisms that determine which genes or suites of genes are turned on or off at particular times and places. Researchers are discovering that each gene needs a supporting cast of hundreds to get its job done. They include proteins that shut down or activate a gene, for example by adding acetyl or methyl groups to the DNA. Other proteins, called transcription factors, interact with the genes more directly: They bind to landing sites situated near the gene under their control. As with alternative splicing, activation of different combinations of landing sites makes possible exquisite control of gene expression, but researchers have yet to figure out exactly how all these regulatory elements really work or how they fit in with alternative splicing.



Why is there more matter than antimatter?

To a particle physicist, matter and antimatter are almost the same. Some subtle difference must explain why matter is common and antimatter rare.



JUPITER IMAGES

Does the proton decay?

In a theory of everything, quarks (which make up protons) should somehow be convertible to leptons (such as electrons)—so catching a proton decaying into something else might reveal new laws of particle physics.

What is the nature of gravity?

It clashes with quantum theory. It doesn't fit in the Standard Model. Nobody has spotted the particle that is responsible for it. Newton's apple contained a whole can of worms.



JUPITER IMAGES

Why is time different from other dimensions?

It took millennia for scientists to realize that time is a dimension, like the three spatial dimensions, and that time and space are inextricably linked. The equations make sense, but they don't satisfy those who ask why we perceive a "now" or why time seems to flow the way it does.

In the past decade or so, researchers have also come to appreciate the key roles played by chromatin proteins and RNA in regulating gene expression. Chromatin proteins are essentially the packaging for DNA, holding chromosomes in well-defined spirals. By slightly changing shape, chromatin may expose different genes to the transcription machinery.

Genes also dance to the tune of RNA. Small RNA molecules, many less than 30 bases, now share the limelight with other gene regulators. Many researchers who once focused on messenger RNA and other relatively large RNA molecules have in the past 5 years turned their attention to these smaller cousins, including microRNA and small nuclear RNA. Surprisingly, RNAs in these various guises shut down and otherwise alter gene expression. They also are key to cell differentiation in developing organisms, but the mechanisms are not fully understood.

Researchers have made enormous strides in pinpointing these various mechanisms.

By matching up genomes from organisms on different branches on the evolutionary tree, genomicists are locating regulatory regions and gaining insights into how mechanisms such as alternative splicing evolved.

These studies, in turn, should shed light on how these regions work. Experiments in mice, such as the addition or deletion of regulatory regions and manipulating RNA, and computer models should also help. But the central question is likely to remain unsolved for a long time: How do all these features meld together to make us whole?

—ELIZABETH PENNISI

CREDIT: JUPITER IMAGES



Visualising 'junk' DNA through bioinformatics

Nancy L Elwess, Sandra M Latourelle, Olivia Cauthorn,
Plattsburgh State University, USA

One of the hottest areas of science today is the field in which biology, information technology, and computer science are merged into a single discipline called bioinformatics. This field enables the discovery and analysis of biological data, including nucleotide and amino acid sequences that are easily accessed through the use of computers. As educators it is important to provide new challenges for students that incorporate the advancements in the areas of molecular biology, the human genome project, and computer technology. This learning experience, using bioinformatics, targets advanced high school and college-level students. The activity allows students an opportunity to visualise the amount of non-coding regions existing within a DNA sequence.

Key words: Bioinformatics; Introns; Exons; Intragenic; Intergenic.

Introduction

Molecular biologists are currently sorting through 'junk': 'junk' DNA that is. For our purposes we will define this 'junk' as stretches of DNA that do not code for genes. One could almost liken the biologists' sorting process to cleaning out a great aunt's attic. What do we keep, what do we throw away? What do we tag for estate sale buyers? Why do people refer to ~98% of human DNA as 'junk'? The most compelling answer is that molecular biologists are struggling to ascribe function to it.

Where is this junk located? The majority of the human genome comprises intronic regions, stretches of repeat sequences, and other assorted gibberish (Maher 2003). In Figure 1 we see the regions that comprise junk. Within a gene, such as gene 1, are intervening sequence areas termed introns. They are intragenic stretches of non-coding DNA. Between the genes resides intergenic DNA, repeat sequences and other unknown entities. Both areas presently have no fully supported function.

It is a real mystery. Since 1977 when Phil Sharp at the Massachusetts Institute of Technology and Rick Roberts at Cold Spring Harbor reported their findings that viral genes contained nucleotide sequences that did not code for protein, the mystery has slowly been unwinding (Sharp & Roberts 1977).

The term intron, *in gene*, was coined to express the portion of a gene that is a spacer and is lost from mature messenger RNA (Gilbert 1978). In addition to introns, eukaryotic genomes are replete with repetitive DNA elements, intergenic or *out of gene* nucleotides. Roy Britten, the California Institute of Technology biologist has stated, "Trash you throw away. Junk you keep in case it may be useful" (Bornstein 2000). Nature has kept the 'junk'. Man is trying to figure out why.

The intent of this article is to provide a learning experience allowing high school and college students to visualise the impact of 'junk' DNA on selected DNA sequences. The DNA

sequences were selected to cover a range of organisms while providing a mixture of results. Feedback from high school faculty teaching advanced placement biology and college faculty teaching general biology and general genetics courses indicates bioinformatics was a helpful tool. They felt their students got a sense of just how large a space the mysterious 'junk' encumbers. This field enables the discovery and analysis of biological data including nucleotide and amino acid sequences that are easily accessed through the use of computers. The National Science Standards approved by the Governing Board of the National Research Council in the United States strongly support the use of technology, modelling systems and inquiry based learning (National Research Council, 1996). This learning experience lends itself to all three.

Before beginning the learning experience, it is wise to consider items already known about DNA. Any new classroom experience always benefits by assessing students' prior knowledge. The Advanced Placement Biology (The College Board, 2004) and the United Kingdom higher grade benchmark standards (Souter, 2003) include the eukaryotic DNA concepts bulleted below:

- DNA is double stranded
- a DNA sequence is defined as the succession of its constituent nucleotides listed from the 5' phosphoral terminus to the 3' hydroxyl end
- both complementary DNA strands carry genes that code for proteins
- eukaryotic organisms have non-coding DNA within and between their genes
- the very simplest organisms have little or no non-coding DNA
- genes have control regions which assist in regulating gene expression
- following the control region is the section of gene transcribed to RNA

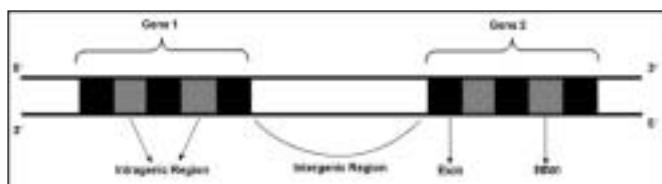


Figure 1. The schematic of DNA sequence shows two genes. Note the intragenic regions known as introns, the exons and the intergenic region between the genes.

- an RNA transcript begins with a 5' untranslated region followed by exons and introns and ends with a 3' untranslated region
- introns are removed to make exons contiguous
- contiguous exons code for eukaryotic proteins
- proteins are made by ribosomes translating the RNA message found in the exons
- proteins are made of amino acids
- an amino acid is coded for with three nucleotides
- there are four different nucleotides in DNA: Adenine, Cytosine, Guanine and Thymine
- the four nucleotides can be mixed and matched to code for 20 different amino acids

Procedures

For this activity the students must have a computer with Internet access and be prepared to use all three of the sites shown in Table 2 in sequence (see page 80). Students can work individually or in pairs and should have a basic understanding of how to navigate the Internet.

1. Look at Table 1 and choose a gene of interest and write down its Accession number.
2. Access the National Center for Biotechnology Information (NCBI) (Table 2).
3. Once at the NCBI site, use the drop-down menu to the right of the word *Search*, and highlight the word nucleotide. This will allow the user to search this site for DNA sequences.
4. Type in the Accession number from Table 1 and hit *GO*. The accession number is the identification number for a particular DNA sequence.
5. This will lead to a link. Click on this link, which will provide a wealth of information, including the desired DNA sequence at the bottom of the page.

Table 1. Accession numbers needed for this activity and recovered information retrieved from the NCBI database. *The percent coding DNA is to be calculated by the students.

Organism/Gene	Accession Number	DNA length in base pairs	No. of amino acids	No. of necessary bases for amino acids	% coding DNA*
Xenopus Cad2 gene	NM20485	3214	262	786	24.45%
Chicken H1 histone gene	M17019	1587	225	675	42.53%
Zea mays alcohol dehydrogenase	M32984	6167	379	1137	18.43%
Norway Rat neurogranin/RC3 gene	U22062	12728	255	705	6.0%
Arabidopsis thaliana DNA chromosome 4 fragment	Z97343	207674	22,199	66,597	32.06%
Homo sapiens myosin heavy chain gene	AJ310932	288,888	2567	7701	2.66%

6. Copy the DNA sequence at the bottom of the page by highlighting the entire sequence then clicking on the copy option under the edit function on the tool bar.
7. Access the Sequence Manipulation Suite (Table 2).
8. Click on the *Filter DNA* link found in the upper right hand corner.
9. When the site comes up, hit the *Clear* button to remove any sequence that might be in the text box. Paste your DNA sequence into this site and click on the *Submit* button. This will remove the number found at the beginning of each line of DNA sequences. The databases will not properly run until the numbers are removed.
10. Once the numbers are removed, copy this sequence.
11. Open the GENSCAN program (Table 2).
12. Paste the DNA sequence into the text box. Use the drop down menu to select the type of organism the sequence came from; the selection is limited to Vertebrate, Arabidopsis, and Maize. Then click on the *Run GENSCAN* button.
13. The results provide a large amount of information including the predicted gene(s), exons, promoter region, poly A tail, as well as the predicted amino acid sequence. Click on the *Click here to view PDF image of the predicted gene(s)* to see an image of the predicted gene(s).

An alternative

Another approach can also be used to visualise the coding versus non-coding regions for human DNA sequences. This can be done once again by accessing the NCBI site. This time look under the *Hot Spots* section of the site, scroll down, select and click the link for human genome resources. At this site, one can select one of the human chromosomes (found in the upper right corner). Once a chromosome has been selected, a map of that particular chromosome will appear with a list of the genes found on that chromosome. Click on any one of these genes; a wealth of information will appear about this gene including a map of the gene with the coding and untranslated regions of the gene. This map however is not as detailed as the map generated in the GENSCAN program.

Results

For the purpose of this paper we are only presenting the DNA sequences from *Arabidopsis thaliana* and the Norway Rat (Table 1) for analysis. When the *Arabidopsis thaliana* DNA sequence was



a Gene	Ex Type	S	Begin	End
2.00	Prom	+	1618	1657
2.01	Init	+	4021	4025
2.02	Intr	+	4366	4489
2.03	Intr	+	4755	4835
2.04	Intr	+	5167	5298
2.05	Intr	+	5556	5693
2.06	Intr	+	5719	5832
2.07	Intr	+	5989	6117
2.08	Term	+	6200	6418
2.09	PlyA	+	6421	6426
3.10	PlyA	-	6728	6723
3.09	Term	-	6893	6750
3.08	Intr	-	7287	7083
3.07	Intr	-	8129	7955
3.06	Intr	-	8366	8266
3.05	Intr	-	9839	9691
3.04	Intr	-	10132	9928
3.03	Intr	-	10433	10351
3.02	Intr	-	10748	10609
3.01	Init	-	10990	10945
3.00	Prom	-	11056	11017

b
GENSCAN predicted peptide 2 313_aa MSLLYGAEDPSIAGMVLDSAFSNLFDLMMELVDVYKIRLPKFTVKVAVQYMRRIIQKKAKFNFNMDLNCKVSYYDSVLVFFYNVLRPPPISSYSSKLESYYSGLGVNSATGLDESFLYEIISGLRSACIDVA SSSAPPAPLTTKPTNELLSEAMPMDTNFLPVAWSIPFDNLNSIMKQDTVLVEDNDHNVDDPENFEGKRIDQFEEGCSFTSSNRESWGRCSSLGGTEDESLTAGEGDQVEKTADVTERKPRDSSREEEEEDSKEKKIKNGGETDAKKPRHEKLERLEAFSKRLRHYILKRVNHRRHRSP
GENSCAN predicted peptide 3 415_aa MSYAYLFKYIIIGDTVGVKSCLLLQFTDKRFQPVHDLTIGVEFGARMITDNKPKIQLQIWDTAGQESFRSITRSYYRGAAGALLVYDITRRETNFNLASWLEDARQHANANMTIMLIGNKCDAHRRAVSTEEEQFAKEHGLIFMEASAKTAQNVEEAFKTAATIYKKI QDGVFDVSNESYGIKVGYGGIPGPSGGRDGSTSQGGGRRD SFLSEPTDSDSTSMSYAYRFKYIIIGDTVGVKSCLLLKFTDKRFQAVHDLTIGVEFGAKTTIDNPKIQLQIWDTVFSNLLLDLYTR ETFNHLASWLEEARQHASENMTTMLIGNKCDLEDKRTVSTE EGEQFAREHGLIFMEASAKTAHNVEEAFVETAATIYKRIQDG VVDEANEPEGITPGPFGGKDASSQQRRGCCG

Figure 2. Figure 2a provides just a fraction of the information from the GENSCAN program. The Gn column in Figure 2a has the numbers 2 and 3 listed, this represents gene 2 and gene 3 within the sequence. The next two columns (Ex and Type) represent the number of predicted exons (Ex) within the gene and the type of exon. These columns also indicate whether the exon is an initial, internal, or terminal exon or if the region is a promoter region (indicator for the direction of transcription) or predicted poly A tail region (post-transcriptional string of adenine nucleotides located at the 3' end of mRNA). The Begin and End columns refer to the base pair location within the DNA sequence and designate where each exon sequence begins and ends. Finally the S column in Figure 2a refers to the DNA strand that is the coding region for each gene. Since DNA is a double helix, either strand can potentially be used as the coding strand. Figure 2b gives the predicted peptide sequences for genes 2 and 3 respectively. Notice it also states how many amino acids are present in the protein sequence, 313 amino acids for gene 2 and 415 amino acids for gene 3.

entered into GENSCAN, the program predicted 44 genes to be present. Figure 2A represents just a fraction of the amount of information that the GENSCAN program provided. Included in this information is the number of genes the program predicts to be present, the number of predicted exons within the gene, and the location of each exon. Finally the program also predicts which DNA strand (of the double helix) is the coding region for each gene. Figure 2b gives the predicted peptide sequences for genes 2 and 3 respectively.

Once the students have analysed the initial information provided by GENSCAN, they can look at the predicted map of the *Arabidopsis thaliana* DNA sequence (Figure 3). The black regions on this map are the coding regions for this particular DNA sequence, with the arrowhead and tail representing the termination and initiation regions respectively. The white regions within and between the genes represent the non-coding, intervening regions of the DNA sequence. Figure 3 was a great visual for our students in that it provided a means by which they could see the direction the gene was being transcribed. Notice that Genes 2 and 3 code in opposite directions.

Figure 4 represents GENSCAN's predicted map for the coding region of the Norway Rat DNA sequence. It clearly illustrated the amount of 'junk' DNA between the exons. In addition to having our students learn the above programs we had them calculate the percentage of 'junk' DNA for various sequences. For example, when the Norway Rat DNA sequence (12,728 bp) was entered into the GENSCAN program the result was a predicted protein sequence of 255 amino acids. Since three nucleotides are needed for each amino acid, the students multiplied 255 x 3 to find that 765 nucleotides were needed for the coding region. The students then divided 765 by 12,728 to find that 6.0% of the gene was coding DNA with 94% of the DNA sequence being 'junk'.

Figure 5 illustrates chromosome 22 from the alternative approach that was mentioned earlier. For the purpose of this exercise the MYO18B gene (arrow on Figure 5) was selected. Once the MYO18B gene was chosen, another page appeared giving background information on this gene including a proposed map of the gene (Figure 6) with its coding and nontranslated regions. Using this predicted map, the students then have a good idea of how much of the gene is 'junk'.

Conclusions

So you've seen the space the mysterious junk occupies, you've experienced a bioinformatics probe into the reality of percentages of intron/exon space and intergenic regions. You have inquired, with the use of technology, into a model system. Hopefully you, too, have questions about the unravelling of this DNA mystery. What do we think we know about 'junk'? What is the significance of this DNA that is cluttering up the genome?

One thing we do know is that the 'junk' DNA variation is so great that it is used to produce DNA fingerprints that differentiate between individual humans and individuals in many other species (Moxon & Wills 1999, Higgins 1999, Baker et al 1993, Jeffreys, Wilson & Thein 1985). We also know that the amount of 'junk' DNA varies in eukaryotic organisms.

There is a strong suggestion that DNA should be examined for functions other than protein coding (Standish 2002). Standish also states, '...noncoding DNA represents the broken remains of old genes that no longer function, mixed in with the

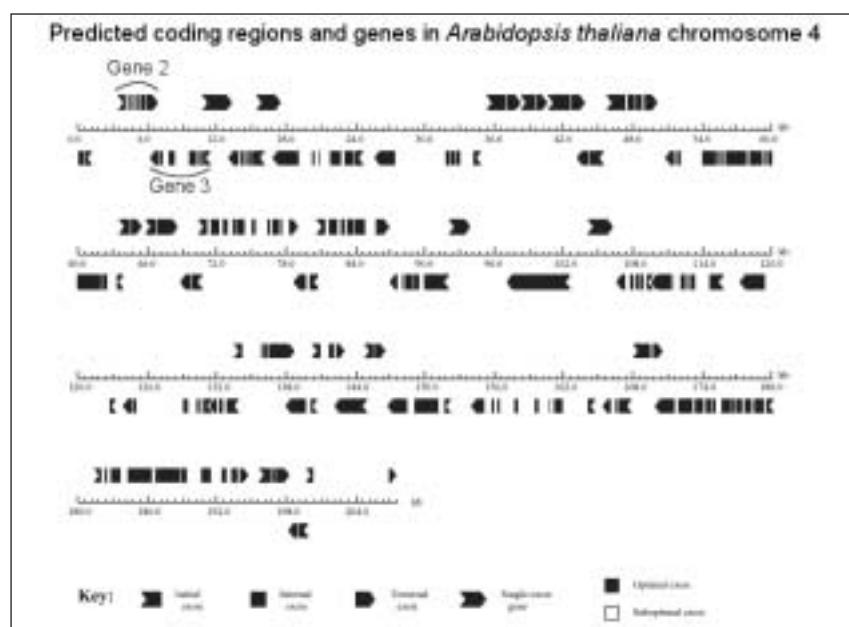


Figure 3. This represents the predicted coding regions found on Chromosome 4 in *Arabidopsis thaliana*. The GENSCAN program predicted 44 different genes to be present. Notice that genes 2 and 3 code in opposite directions. More detailed information on genes 2 and 3 can be found in Figure 2.

dust-like repetitive sequences that have blown in and multiplied. Thus, noncoding DNA, as long as it lacks function, may be mined for evidence of life's distant past...' (Standish 2002). Perhaps we can think of these as genetic 'hangovers' of vestigial organs, or even viral inserts occurring over time.

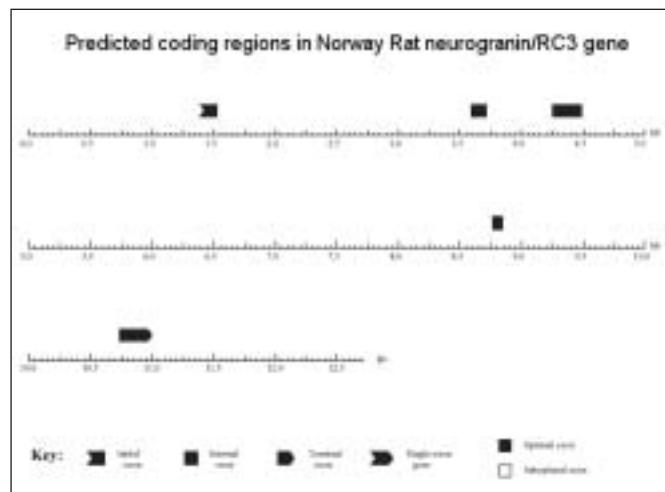


Figure 4. The Norway Rat neurogranin/RC3 gene sequence was entered in the GENSCAN program, the schematic above represents the coding regions from the 12,728 nucleotides that were entered into the program.

Cryptoanalysts (those who break secret codes), linguists (those who study language) and physicists have worked together and found peculiar indications of a concealed *language* in the intervening sequences of DNA (Flam 1994). Scientists involved in this investigation suggest that there may be messages in the 'junk'. This is another tantalising strand in the DNA mystery.

There have been reports that sequences of 'junk' are inherited and some of the repetitive patterns are associated with increased cancer risk. The DNA has been found to rapidly mutate in response to cancer. Perhaps some of the intervening DNA is

involved with cellular processes (Suurkula 2001).

Several pieces of research have led investigators to suggest that intervening DNA plays an essential role in regulating gene expression during development (Ting 1995). At least 700 studies have confirmed that non-coding DNA acts as enhancers for transcription of proximal genes (Suurkula 2001). Simon Shepherd, a cryptography lecturer and computer security analyst in the UK, looks at 'junk' DNA as another secret code. He feels that introns are some sort of error correction code and they are in place to repair the occasional mistakes made during DNA replication (Kruszelnicki 2003).

Some individuals argue that introns were an essential feature of the earliest organisms. They believe that introns are a recent arrival in eukaryotic ancestry and are there to help spawn the diversity of regulatory mechanisms needed for gene expression in complex organisms (Mulligan 2002). Still others believe that some introns are exons of other genes.

The sorting of junk is complex, it is tedious and frustrating and just as in any estate sale there are those 'relatives' that want to keep an item and those who do not. Only time will tell what needs to be tagged for the future buyers of this generation.

As more and more functions are established, the quantity of

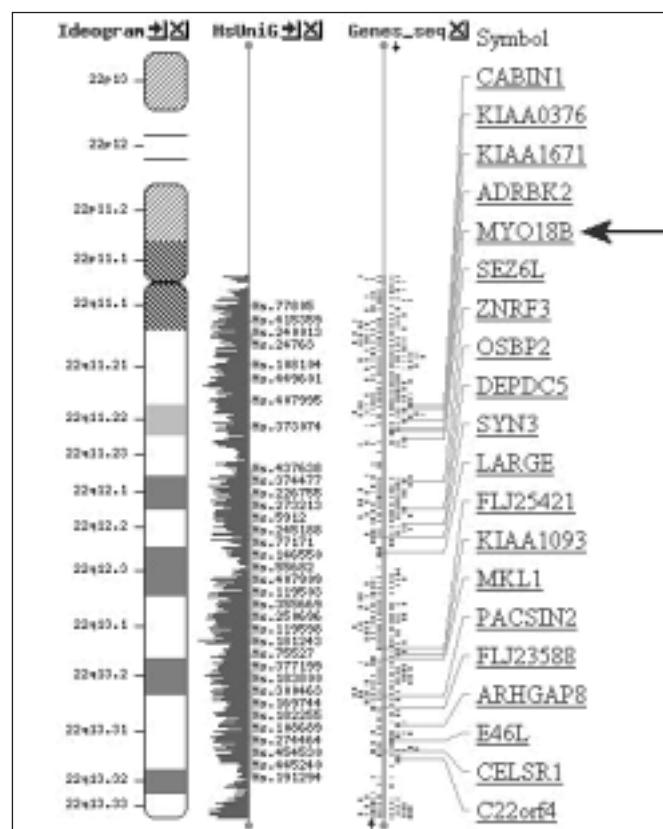


Figure 5. Predicted map of human chromosome 22 using the National Center for Biotechnology Technology (NCBI) site. The known genes located on chromosome 22 are listed to the far right. The arrow is next to the gene (MYO18B) that was selected in Figure 6.

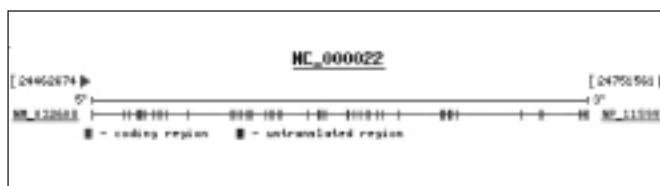


Figure 6. Once the MYO18B gene was selected on human chromosome 22 (Figure 5), this map appears. It demonstrates the coding and untranslated regions of this gene.

DNA we now call junk will likely decrease and questions will be answered and new mysteries will emerge. Isn't that the nature of science?

The learning experience outlined in this paper has been used in Teacher-to-Teacher workshops for high school teachers of biology as well as in an introductory college bioinformatics course. This simple activity done in the classroom or as a homework assignment made our students aware of the vast wasteland of DNA that lies within higher organisms' genomes. This experience *engaged* students through focusing on unanswered questions. What is this 'junk'? What is the reason for so much DNA with no apparent purpose?

Our students *explored*. They developed an awareness of bioinformatics and practised their skills by navigating through massive databases. They collected data.

Students *explained*. Their knowledge base for manipulating bioinformatics information to analyse nucleotide and amino acid sequences increased enough to allow for explanatory conversation between instructor-student and student-student.

Students *elaborated*. They used their skills in analysis through database manipulation for other learning experiences such as the mitochondrial DNA sequence and *Alu* insertion laboratory activities carried out over the course of the semester.

Lastly, students *evaluated*. They provided feedback to their instructors indicative of their understanding of the nature of science. In the future we will incorporate this learning experience into our general genetics curriculum.

Table 2. Databanks used for this activity

National Center for Biotechnology Information (NCBI)
www.ncbi.nlm.nih.gov

The Sequence Manipulation Suite
www.ualberta.ca/~stothard/javascript/

GENSCAN
<http://genes.mit.edu/GENSCAN.html>

References

- Baker C, Gilbert D, Weinrich M, Lambersten R, Calambokidis J, McArdle B, Chambers G, O'Brien S (1993) Population characteristics of DNA fingerprints in humpback whales (*Megaptera novaengliae*). *Journal of Heredity* **84**: 281-290.
- Bornstein S (2000) *Discovery of junk DNA*. Retrieved June 29, 2003, from www.cephasmistry.com/news_dna_junk.html
- Claverie J and Notredame C (2003) *Bioinformatics for dummies*. New York: Wiley.
- Flam F (1994) Hints of a language in junk DNA. *Science* **266**: 1320
- Gilbert W (1978) Why genes in pieces? *Nature* **271**: 501.
- Higgins M (1991) Acid test. *ABA Journal* **85** (October): 64-67.
- Jeffreys A, Wilson V, Newmann R, Keyte J (1988) Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. *Nucleic Acids Research* **16**: 10953-10971.
- Kruszelnicki K (2003) *Great moments in science: Language in junk DNA*. Retrieved 17 August 2003, from: www.abc.net.au/science/k2/moments/s133634.htm
- Maher B (2003) The dark side of the genome. *The Scientist* **31**.
- Moxon E and Wills C (1999) DNA microsatellites: agents of evolution? *Scientific American* **280**: 94-99.
- Mulligan M (2002) *Introns & evolution*. Retrieved 17 August 2003, from www.mun.ca/biochem/courses/3107/Lectures/Topics/Introns.html.
- National Research Council (1996) *National science education standards* (Q183.3.AIN364). Washington DC: National Academy Press.
- Sharp P and Roberts R (1977) Adenovirus mazes at Cold Spring Harbor. *Nature* **268**: 101-104.
- Smith J, Newton C, Alves A, Anwar R, Jenner D, Markham A (1990) Highly polymorphic minisatellite DNA probes. Further evaluation for individual identification and paternity testing. *Journal of the Forensic Science Society* **30**: 3-18.
- Souter N (2003) DNA in the school curriculum - a poorly exploited asset? *School Science Review* **84** (308) 57-63.
- Standish T (2002) Rushing to judgment: Functionality in non-coding or 'junk' DNA. *Origins* **53**: 7-27.
- Suurkula J (2001) *Junk DNA: Over 95 percent of DNA has largely unknown function*. Retrieved 16 August 2003 from www.psra.org/junkdna.htm.
- The College Board (2004) *The College Board Advanced Placement Program for Biology Course Description*. Retrieved 10 August 2004 from College Board AP Central® for Educators www.apcentral.collegeboard.com.
- Ting S (1995) A binary model of repetitive DNA sequence in *Caenorhabditis elegans*. *DNA Cell Biology* **14**: 83-85.
- Venter J C, Adams M D, Meyers E W, Li P W, Mural R J, and 269 other authors. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.

Nancy L Elwess is Associate Professor, Department of Biological Sciences, Plattsburgh State University, USA.

Tel: + 1 518 564-5169. Fax: + 1 518 564-3175.

Email: nancy.elwess@plattsburgh.edu.

Sandra M Latourelle is Instructor, Department of Biological Sciences, Plattsburgh State University. Tel: + 1 518 564-5163.

Email: latoursm@together.net. Olivia Cauthorn, BS Biology, Plattsburgh State University.

MÉDIMENTO

A découvrir



Les gestes qui sauvent résumés en 64 pages : c'est la gageure réalisée par Frédéric Lador et Roch Ogier, fondateurs et formateurs d'une école de premiers secours dont le siège est à Genève en publiant, aux éditions Médecine & Hygiène, sous le titre *Premiers secours*, un ouvrage simple et précis, facilement utilisable. Un manuel indispensable préfacé par le Dr Bernard Vermeulen, médecin responsable du centre d'accueil et d'urgences des HUG.

A écouter



Sous la direction d'Eric Bauer, plusieurs solistes de l'Ensemble Instrumental Romand interpréteront, le dimanche 13 février, à 15h, une version pour quintette à vent signée Gordon Davis des *Jeux d'enfants*, opus 22, de Georges Bizet. Répétitions le samedi 12 de 14h à 16h et le dimanche 13 de 13h30 à 14h45, à la salle Opéra, sur le site Cluse-Roseraie. Pour info, tél. 022 305 41 44.

A lire



Dans la collection *Traits de plume*, les éditions Clé de sel publient *Les mots de Marie*, un roman signé Colette Bonnet-Seigne. Après la blessure d'un premier enfant mort-né, Marie ressent une envie d'écrire ses révoltes. Les mots se choquent, s'entrechoquent. Elle les écrit en vrac, pour ne pas les perdre. Tout dire, il faut tout dire, ne rien oublier pour dégorger la plume lourde et l'âme débordante. Ce n'est qu'ainsi que Marie peut poursuivre sa vie de femme, sa vie de mère.

TÊTE-À-TÈTE

L'ère post-génomique de la médecine

Seul participant suisse au décryptage du génome humain, le Pr Stylianos Antonarakis contribue à un nouveau projet international. La recherche sur l'ADN ouvre des perspectives pour la médecine.

Comment vous êtes-vous retrouvé à bord du projet de décodage du génome humain ?

Très simplement : j'avais un grand intérêt pour la trisomie 21 ou syndrome de Down. On savait que ce syndrome vient de la présence de trois chromosomes 21 au lieu de deux. A l'époque, la question était de savoir quels étaient les gènes impliqués dans ce syndrome et la seule façon de répondre à cette question était de connaître toutes les séquences ADN de ce chromosome. C'est pourquoi j'étais intéressé par le séquençage du chromosome 21. Avec notre groupe de recherche, nous avons établi une carte génétique d'abord, puis une carte physique ensuite et nous avons commencé d'identifier des gènes. Nous avons participé à la lecture de toute la séquence du chromosome, pour trouver l'étiologie des différents symptômes de la trisomie 21. Le but était médical.

Quels ont été les résultats les plus importants de vos travaux ?

On a contribué à clarifier le mécanisme de la genèse de la trisomie. On a établi les cartes génomique et physique du chromosome. Notre laboratoire a identifié quelque 150 gènes, soit plus de la moitié des 250 gènes impliqués dans le chromo-

some 21. Enfin – et cette découverte qui est issue principalement du travail d'un des collaborateurs de notre laboratoire, le Dr Dermitzakis – on a trouvé qu'à côté des gènes, d'autres séquences d'ADN étaient aussi importantes que les gènes, puisqu'elles ont été conservées tout au long du processus de l'évolution. La fonction de cet ADN poubelle (ou CNG, « séquence non génique conservée » en français) est l'un des objectifs du projet ENCODE (lire ci-dessous). Nous sommes intéressés à trouver comment ces CNGs donnent le phénotype du syndrome de Down, ce qui permettra de développer des thérapies ciblées, existantes pour l'instant.

Séquencer l'ADN est un travail effectué par des machines. Quelle est la part de l'humain dans ces recherches ?

L'homme permet d'identifier des séquences qui ont une fonction. Il existe 3 billions de lettres ou nucléotides (A, C, G et T) dans notre génome, dont 5% sont importantes. Ces 5% ont été découverts par comparaison avec le génome d'autres espèces, comme la souris.

A quoi ce décodage sert-il ?

Le décryptage du génome humain fournit une infrastructure extraordinaire pour

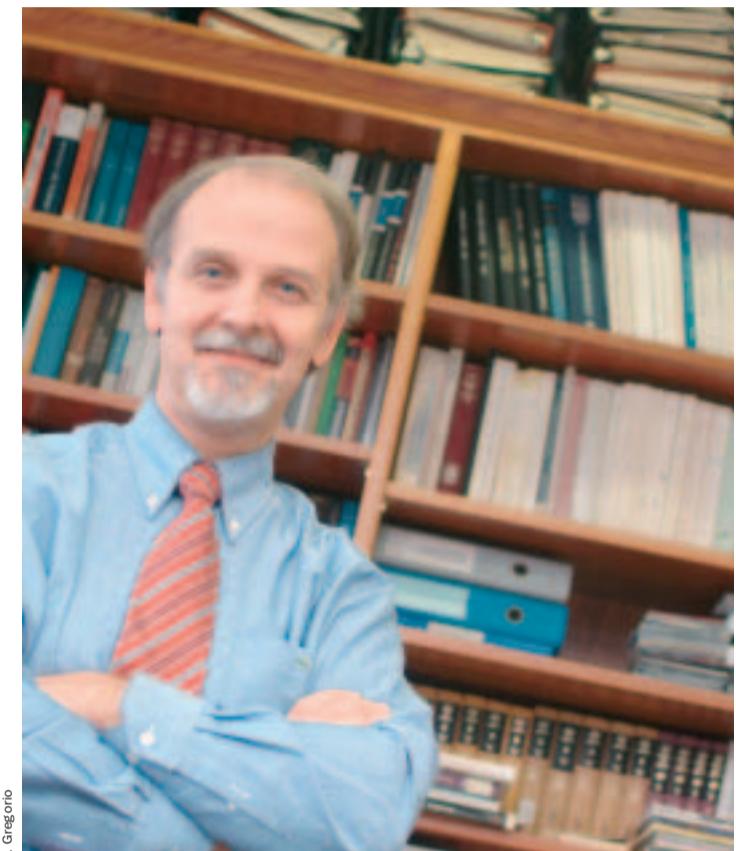
trouver les causes de nos maladies. Le séquençage d'autres espèces nous donne deux possibilités : utiliser mieux qu'avant les modèles animaux et comparer les séquences animales avec les nôtres pour trouver celles qui sont importantes dans notre génome. L'hypothèse de base est que ce qui est conservé par l'évolution l'est parce qu'il a une fonction. Sinon, il n'est pas conservé.

Que réserve encore la recherche sur l'ADN ?

Beaucoup de surprises ! On ne sait pas tout, par exemple sur les régions fonctionnelles des séquences qui se trouvent dans cet ADN poubelle. En fait, on sait très peu sur ce génome et on va l'étudier encore pendant tout ce siècle. Nous sommes entrés dans une ère post-génomique de médecine : non seulement nous savons lire notre génome mais aussi celui d'autres mammifères. On peut comparer cette époque à celle qui a suivi la description de l'anatomie humaine, publiée en 1543 par Andreas Vesalius, à Bâle.

Pour les médecins, et les patients, qu'est-ce que cela signifie ?

L'ère post-génomique sera celle d'une meilleure compréhension des maladies génétiques. Pas seulement les maladies rares mais aussi



Pour le Pr Antonarakis, l'ère post-génomique sera celle d'une meilleure compréhension des maladies génétiques.

celles dites polygéniques, multifactorielles complexes ayant une composante génétique forte. Des maladies bien de notre époque comme l'athérosclérose, l'hypertension, le diabète, l'obésité, la sclérose en plaques, Alzheimer, les maladies mania-codépressives, la schizophrénie, l'asthme, le vieillissement précoce, etc. Une autre chose montrée par la lecture du génome est la variation entre individus : une lettre sur mille est différente entre

deux génomes, ce qui fait 3 millions de lettres différentes par individu. Cette différence est très importante pour la susceptibilité aux maladies.

Ce qui veut dire des traitements individualisés ?

On pourra surtout faire une médecine prédictive meilleure, pratiquement au niveau individuel.

Propos recueillis par Véronique Preti

FICHE PRATIQUE

Nom de code : ENCODE

L'objectif de cette recherche est de comprendre le rôle de l'ADN poubelle.

L'ADN, notre patrimoine génétique, continue de livrer ses secrets. Seulement 5% d'ADN a une fonction utile : 2% sont les gènes, qui codent pour des protéines, et 3% ces séquences non géniques conservées, dites CNG, sur lesquelles on ne sait rien, si ce n'est que l'évolution les a gardées. « On pense qu'une partie de ces CNGs pourrait réguler l'expression des gènes, mais... tout est à découvrir ! », indique le Pr Antonarakis.

Trouver tous les éléments fonctionnels de l'ADN est l'objectif d'ENCODE, financé par le NIH (National Institute of Health) américain. Un projet international qui regroupe 150 chercheurs de 6 pays différents, dont, pour la Suisse, les équipes du Pr Antonarakis et d'Alex Raymond, professeur associé à l'Université de Lausanne.



Le Pr Antonarakis a reçu le prix Lejeune pour l'ensemble de sa carrière.

Ministre français de la santé, Jean-François Mattei, le premier Prix Lejeune, doté de 30'000 euros. Remis par la Fondation Lejeune à l'occasion des dix ans de la mort de Jérôme Lejeune, le prix récompense l'engagement d'une carrière dans la recherche sur le syndrome de Down. Surtout, le Pr Antonarakis a été ému de la présence de Bruno, un homme important dans l'histoire de la médecine : « En 1959, le pro-

fesseur Lejeune, de l'Hôpital Necker, a été le premier à observer le troisième chromosome 21, or il l'a fait sur le sang de Bruno, né en 1958. »

Hormis ses recherches sur le chromosome 21, le service de génétique médicale que dirige Stylianos Antonarakis, collaborateur des HUG et de l'Université de Genève, voit chaque année un millier de patients atteints de maladies génétiques.

V.P.

Pulsations

Le mensuel des HUG

Hôpitaux universitaires de Genève
Service de la communication
Rue Michel-du-Crest 24
CH-1211 Genève 14
Tél. (+41 22) 372 60 07
Fax (+41 22) 372 60 76
<http://www.hug-ge.ch>

Editeur responsable
Bernard Gruson
Responsable des publications
Séverine Hutin Dub

Rédactrice en chef
Agnès Reffet
pulsations-hug@hug-ge.ch

Conception / réalisation
csm sa

Régie publicitaire
Imédia SA
Tél. (+41 22) 307 88 95
Fax (+41 22) 307 88 90

Impression
ATAR Roto Presse SA
Tirage
28'000 exemplaires

Les manuscrits ou propositions d'articles sont à adresser à l'éditeur. La reproduction totale ou partielle des articles contenus dans "Pulsations" est autorisée, libre de droits, avec mention obligatoire de la source.

Ein Test aus "Abfall"

Der "genetische Fingerabdruck" ist schon mit nur einer Zelle nachweisbar

Von Sonja Kastilan

Unvorstellbar lang wäre die Schlange der wartenden Männer, wenn sie tatsächlich alle zum Gentest antreten müssten. Die Gesundheitsämter oder Polizeistationen hätten über Wochen und Monate zu tun. Und anders als bei der Schluckimpfung in Kindertagen bekäme man kein Zuckerstück auf die Zunge gelegt. Stattdessen würden Helfer mit Wattestäbchen in den Mündern nach Spucke fischen, um Proben für das genetische "Fingerprinting" zu erhalten.

Für den genetischen Fingerabdruck sind die eigentlichen Gene einer Person völlig unerheblich. Für solche Erbinformationen haben Kriminalbeamte wenig Sinn. Interessant sind vielmehr so genannte "nicht codierende" Bereiche zwischen den Genabschnitten. Sie besitzen keinerlei Informationsgehalt für die Kontrolle oder Herstellung von Proteinen, machen aber dennoch einen Großteil des menschlichen Erbguts aus. Jedoch zeichnen sich die Abschnitte dieser "Junk-DNS" durch ganz bestimmte Längen und Wiederholungen in der Basensequenz aus: einzigartig und individuell verschieden.

Auf diesen feinen Unterschieden beruht ein genetischer Fingerabdruck, den jetzt empörte Stimmen für jeden deutschen Mann fordern. Molekularbiologen erstellen ihn mit Hilfe des PCR-Verfahrens (Polymerase Chain Reaction), ausgehend von Speichel-, Blut- oder Spermaproben. Mit der PCR lässt sich ein DNS-Molekül in wenigen Stunden millionenfach vervielfältigen, eine einzige Körperzelle mit Erbmaterial genügt. Theoretisch. "In der Praxis sind es meist zehn bis 20 Zellen", erklärt Cornelia Schmitt vom Institut für Rechtsmedizin der Universität Köln. Aber auch das ist nicht wirklich viel: Eine weggeworfene Zigarettenkippe, Hautzellen unter dem Fingernagel eines Opfers oder ein ausgerissenes Haar würden ausreichen.

Zur Analyse werden die Proben zuvor meist biochemisch von störenden Eiweißen befreit. Das Enzym Proteinase K erledigt das schnell und ohne größeren Aufwand. Danach kann die Aufarbeitung im Labor mittels PCR beginnen. Um einen genetischen Fingerabdruck zu erhalten, wird allerdings nicht das gesamte Erbmaterial vervielfältigt: So genannte Primer lagern sich gezielt an die vorhandenen DNS-Moleküle an, damit nur der dazwischenliegende Bereich kopiert und im weiteren Verlauf vielfach angefertigt wird. Die Produktion der neuen DNS-Ketten übernimmt das Enzym Polymerase, das in den achtziger Jahren bei wärmeliebenden Bakterien entdeckt wurde.

"In Deutschland besteht ein genetischer Fingerabdruck aus fünf Kernsystemen", so Schmitt. Was kompliziert klingt, bedeutet schlicht, dass fünf nicht codierende DNS-Abschnitte analysiert werden. Im Einsatz sind dafür insgesamt fünf verschiedene Primerpaare, die Bereiche der Chromosomen vier, sechs, elf, zwölf und 21 vervielfältigen. So entstehen pro Primerpaar zwei Kopien der DNS-Probe - je eine für das väterliche und das mütterliche

Chromosom im Erbgut. Viele Labore setzen allerdings noch mehr Primer ein, immer auf der Suche nach besser geeigneten Chromosomenabschnitten.

Zur vergleichenden Analyse werden die PCR-Ergebnisse dann elektrophoretisch in einem Gel aufgetrennt. Auf diese Weise lassen sich die unterschiedlichen Längen der vervielfältigten "Kernsysteme" erkennen. Der berühmte "Streifencode" einer Person entsteht dabei nicht zwangsläufig: Einige Labore färben inzwischen keine Gele mehr, sondern nutzen moderne Sequenziermaschinen für die Längenbestimmung mittels Laser. "Und die liefern Kurven mit Peaks statt einem Strichcode", so Schmitt aus dem Kölner DNS-Labor. Wenn es schnell gehen soll, auch schon innerhalb von einem Tag.

Bei den fünf deutschen "Kernsystemen" wird es wohl nicht bleiben. "Es ist geplant, einheitliche Bestimmungen für den genetischen Fingerabdruck festzulegen", erklärt Gerhard Schlemmer vom Bundeskriminalamt in Wiesbaden. In Europa sollen dann zukünftig acht Chromosomenabschnitte den Fingerabdruck festlegen.

reviews

BOOKS • CD ROMS • ART • WEBSITES • MEDIA • PERSONAL VIEWS • SOUNDINGS



DNA 50 Four Plus: Writing DNA

A Wellcome Trust exhibition at the TwoTen Gallery and the Wellcome Building, London NW1, until 29 August 2003.
Admission free
www.wellcome.ac.uk

Rating: ★★★★

This exhibition celebrating the 50th anniversary of the discovery of the structure of DNA takes us beyond science into a vivid representation of the personalities involved. Ten artists use visual, literary, and digital media to present fresh perspectives on the discovery and to show science as social history, science as passion.

The exhibition is structured like a quest. On entering, visitors get a map and have to guide themselves through two buildings to find various art works at different locations. Along the way the personalities of the four scientists who sought to solve the puzzle of the DNA molecule emerge.



In this World

Directed by Michael Winterbottom
UK release date: 28 March 2003
US release scheduled for autumn 2003

Rating: ★★★★

If *Dirty Pretty Things* (BMJ 2002;325: 1427) is about what happens to asylum seekers in the United Kingdom, then *In This World* is about how they get here.

Sixteen-year-old Jamal (Jamal Uddin Torabi) and his cousin Enayat (Enayatullah) leave a refugee camp in northern Pakistan for London. We're told that they are from Afghanistan, but not how they got to the camp or why they want to leave for London. The film follows their journey and the danger along the way.

*Items reviewed are rated on a 4 star scale
(4=excellent)*

Two of the scientists, Rosalind Franklin and Maurice Wilkins, used x ray images to probe DNA's three dimensional structure. Wilkins shared Franklin's results with the other two, Francis Crick and James Watson, who applied them to their own studies and unfolded the mystery. But when, in 1953, their discovery was published in the journal *Nature* with Franklin's and Wilkins' x ray pictures, Crick and Watson did not acknowledge Franklin's contribution. Crick, Watson, and Wilkins won the Nobel prize, while Franklin died before she could be considered for it.

Artist Kevin Clarke has incorporated photography into DNA sequences obtained from blood samples—thereby combining art and science—to create an original portrait of James Watson.

Inspired by Rosalind Franklin, Jessica Curry and Dan Pinchbeck have created a multisensory experience of DNA. They use digital sound, video projection with x ray photographs, and animations with DNA sequences interposing Franklin's image. Their work is an analogy of the complex structures beneath the surface and Franklin's role within science, both of which have finally emerged. "We went past the anger into a celebration of who she was," they say.

This is a film about the loss of native language and native clothing, the loss of money, the loss of dignity, and the loss of identity. Jamal and Enayat are passed along a network of people smugglers via Iran and Turkey and have a number to call in every city that they reach. They are advised to don Western outfits and are prepped with phrase-book style remarks. But they are also forced to work in sweatshops and to pay the smugglers more cash than they thought they had agreed.

Picked up by the Iranian authorities on the bus to Tehran, they are returned to the border with Pakistan and left to walk back to the nearest city. Effectively starting again, they have to draw on more of the secret dollar stash that Enayat keeps in his shoes to return to Iran.

Jamal and Enayat go to extraordinary lengths to reach every destination. By the time the story gets to Sangatte, the refugee camp in northern France made famous by a thousand xenophobic newspaper headlines, lives have been lost. After I had watched this film, the headlines seemed more shameful and disgusting than ever.

In this World may be a work of fiction, but it felt more like a documentary. It was hard to believe that it was filmed on a set and that



The secret of life by Jessica Curry and Dan Pinchbeck

Penny McCarthy has produced pencil drawings to represent Crick and Watson's *Nature* paper. Neil Chapman's sound installation consists of manipulated pieces of text about DNA that create a set of apparently accidental phrases, but where "sense spreads itself." The exhibition also incorporates the first public display of the Crick archive.

I found in one of the exhibits a phrase that stays with me: "Life is the shape it is for a purpose."

Irina Haivas BMJ Clegg scholar
ihaivas@bmj.com



In this world: redefining courage

the characters were actors. Shaky camera work and subtitles can be irritating. But here they clearly add to the sense of realism. Cleverly, most of the dialogue that Jamal and Enayat cannot understand is not subtitled, forcing the viewer to share the cousins' inability to comprehend.

This film redefined courage and fear for me. Jamal and Enayat are extremely resourceful in the face of a full range of harrowing experiences. It's hard to believe that any of this really happens in this world.

Anna Ellis editor, studentBMJ
aellis@bmj.com

Wikström's experiments showing proton pumping by cytochrome oxidase, published early in 1977. But it was not until 1985 that Mitchell was prepared to accept the evidence presented by Wikström and others, and then only when he had thought up a mechanism (since shown to be incorrect) to accommodate it. In Mitchell's garden of the mind, ideas seem to have deeper roots than facts.

The second attack concerned the chemical mechanism of ATP synthesis. Mitchell's postulate (from which he never wavered) that protons are directly involved in the esterification of ADP by phosphate was challenged by Paul Boyer, Harvey Penefsky and my group, who believed that ATP synthesis is brought about by a conformational change in the ATPase such that ATP, already formed spontaneously on the enzyme, is released. This view led to the Nobel Prize in 1997 for Boyer and John Walker.

Have we reached the end of the story? Not quite. The postulate of a non-phosphorylated high-energy intermediate in the chemical hypothesis is correct; its nature is not. The postulate of the chemiosmotic hypothesis that this intermediate is an electrochemical proton gradient is correct; the way in which it was thought to synthesize ATP is not. The binding-change mechanism of ATP synthesis proposed by Boyer is probably correct; but will the rotary mechanism depicted in the textbooks survive challenges?

I recommend this biography as a scholarly account of the life of one of the giants of late-twentieth-century biochemistry who was also a fascinating, if enigmatic, human being. His amiable character, his eccentricities and his health and emotional problems are fully described. But I still do not think that I really understand the workings of his mind.

Like Tom Blundell, writing in the foreword to this book, I find it quite extraordinary that Mitchell refused to permit Bob Williams to publish their correspondence from before the publication of Mitchell's first paper on the chemiosmotic theory, concerning the possible role of protons in ATP synthesis. The reason given for Mitchell's 1977 reversal of the decision to cease biochemical research — namely that he believed "there was a strong movement among the antagonists of the general chemiosmotic theory... to take advantage of our withdrawal and press new, ill-founded research claims, the effect of which was to undermine the consensus of opinion which we had been working on to promote" — I find even more extraordinary.

But perhaps it is not possible for a mere biochemist to analyse the mind of a genius. ■

E. C. Slater is emeritus professor of physiological chemistry at the University of Amsterdam, the Netherlands. He is currently at 9 Oaklands, Lymington SO41 3TH, UK.

Science in culture

Art after DNA

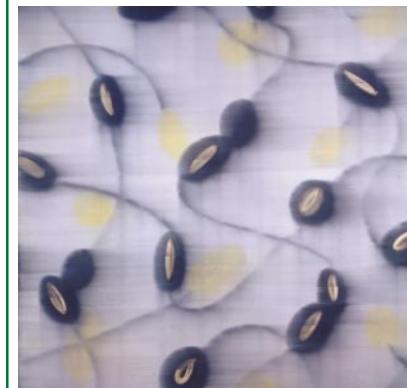
The double helix has inspired scientists and artists alike.

Lynn Gamwell

No sooner had James Watson and Francis Crick published their seminal paper on DNA in 1953 than artists began depicting the double helix as a cultural icon. Salvador Dalí painted swarms of spiralling DNA molecules in his paintings from the late 1950s, such as *Butterfly Landscape (The Great Masturbator in a Surrealist Landscape with DNA)*, 1957–58, and decades later sculptor Tom Otterness forged a bronze double helix joined by tiny stick figures (*DNA Chain*, 1986). Some artists responded to some of the troubling issues surrounding DNA, such as cloning and stem-cell research. Alexis Rockman, for example, painted a colourful landscape populated with genetically modified plants and animals to warn viewers that the future may contain mutated monsters (*The Farm*, 2000).

Although art about science icons and applied science is interesting, it does not focus on the pure science of DNA. That topic is addressed by artists who express wonder at the highly complex, ever-changing organic processes that are the physical mechanism for the force of life. One example is the British painter Mark Francis, whose *Source* (1992) is a wall-size depiction of sperm cells presented not with clinical accuracy but out of focus, as if seen through a veil. Inspired by microscopy, Francis uses a vocabulary of curved, flat shapes that he inherited from nineteenth-century art nouveau designers — who, in turn, first copied the shapes from stained, transparent slices of tissue prepared between glass plates for viewing with a light microscope. Francis uses a modern electron microscope for his source images, but his forms retain the flat, free-form appearance associated with a century of biomorphic abstract art. In a nod to recent methods of charting complex genomic data, Francis paints the sperm cells hovering above a grid.

In the 50 years since Watson and Crick cracked the structure of DNA, biologists and bioinformaticists, have painstakingly mapped and sequenced the genomes of various species: the fruitfly, the mouse



Sex cells: the view of sperm in *Source* by Mark Francis was inspired largely by microscopy.



Genetic screen: Benjamin Fry's *Valence* uses the BLAST algorithm to create moving patterns.



DNA 2 by Susan Rankaitis is both a stand-alone artwork and a backdrop to the dance of life.

and, in 2001, the human. The computing technology developed for this effort inspired a new generation of video artists, such as Benjamin Fry, a programmer at the Massachusetts Institute of Technology's Media Laboratory, to create visualizations of complex biological data. In *Valence* (2000), Fry has used the BLAST algorithm — a tool for searching through genomes — to produce ever-changing patterns that move rhythmically across a video screen.

The curved free-standing panel *DNA 2*, which measures over 5 metres across, is the result of a collaboration between visual artist Susan Rankaitis, molecular biologist Robert Sinsheimer and choreographer John Pennington. Inspired by the Human Genome Project and tutored by Sinsheimer, Rankaitis combined pictures of DNA with text by the biologist to create her chromosome-shaped collage. It can be viewed alone or used as a backdrop for Pennington, who, attired in coils of pulsing fibre-optic cable, spins across a stage. By projecting flickering light onto her giant chromosome, Rankaitis makes her layered images and text join the dance of life.

Charles Darwin was confident that there was a physical mechanism underlying natural selection, but never found it. Today's artists understand the central role of DNA in this mechanism. It inspires them by embodying Darwin's core idea: that nature is a web of dynamic forces with no predetermined purpose or meaning. Works by artists such as Francis, Fry and Rankaitis resonate with this concept of life, and with the complex, abstract processes that go on silently, systematically and invisibly within the double helix.

Lynn Gamwell is director of the Art Museum at the State University of New York, and author of *Exploring the Invisible: Art, Science and the Spiritual* (Princeton University Press, 2002).