## IN THE NEWS

**Doctor in a cell**

"The sci-fi vision of a molecular medical team that can be injected into a patient, coursing through his bloodstream to diagnose a disease and treat it, has taken a step nearer to reality" (*AFP Discovery Channel*).

On 28th April 2004, Ehud Shapiro and colleagues, from the Weizmann Institute, reported online in *Nature* the creation of the first molecular computer that could have medical use. This computer exploits the base-pairing properties of DNA to detect mRNAs that are diagnostic for disease and then destroys them by releasing antisense DNA molecules. The computer is "so small that about a trillion can fit in a drop of water" (*The Telegraph*) and "is listed in the 2004 Guinness Book of World Records as the world's smallest biological computing device" (*The Guardian*).

"The computer has two states, 'yes' and 'no', and changes from one to the other on the basis of a single variable, like the presence or absence of the RNA it is looking for. If at the end of a series of steps it is in the 'yes' state, the diagnosis is positive" (*The New York Times*).

What do the experts think? '"I think it's very elegant — almost like a beautiful mathematical proof," said George Church. "But it's not working in human cells yet"' (*The New York Times*).

The molecular computer proved its worth in the optimal conditions of the laboratory: "To actually track down and disable cancer cells in a human body, it would have to survive the hurly-burly of proteins, lipids, polysaccharides and nucleic acids, any of which could block or disable it" (*The Guardian*). But, Professor Shapiro is upbeat: '"Only two years ago we predicted that it would take another 10 years to reach the point we have reached today"' (*The Guardian*).

*Tanita Casci*

GENOME EVOLUTION

# *Alu* elements — a complex human affair

Formerly described as junk and parasitic, who would have expected that transposable elements would help us out from a potential evolutionary embarassment — the low gene number in the human genome? Ast and colleagues have shown that inclusion of *Alu* elements in exons promotes alternative splicing and therefore genome diversity. Using bioinformatic and experimental approaches, they now identify the mutational steps that create 5′-splice sites in alternatively spliced *Alu* elements.

The surprisingly low gene number and much higher protein number made alternative splicing an obvious process that could account for our biological complexity. Mouse and human sequence comparisons have indicated that alternative splicing is often associated with recent exon creation and/or loss. The authors previously showed that ~5% of human alternatively spliced exons are derived from *Alu* elements — short primate-specific retrotransposons, of which humans have ~1.4 million copies. These *Alu* exons (AExs) have evolved from intronic *Alu* elements. But how can a 'free' intronic *Alu* element turn into an exon that is alternatively spliced?

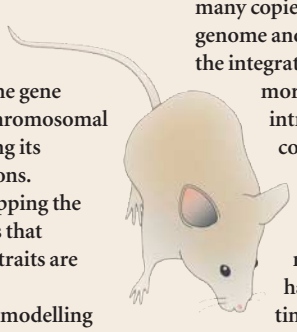
To answer that, the authors compared AEx sequences with those of their intronic ancestors. They focused on the 5′-splice site: >98% of human introns begin with GT and very few with GC, although these are the ones that are supposed to be mainly involved in alternative splicing. Strikingly, the most significant change in AExs was at position 2 of the intron, where a C→T transition creates a canonical GT 5′-splice site. Comparing over 300,000 sequences showed that positions 2 and 5 in the intron are those that matter.

But how is the alternative splicing of AExs regulated? Using site-directed mutagenesis, the authors found that alternative splicing of AEx is possible because C at position 2 of the 5′-splice site unpairs from the U1 small nuclear RNA (snRNA); its interaction with the 5′-splice site is crucial for constitutive splicing. Moreover, it seems that positions 3 and 4 of the intron control the level of exon inclusion, whereas G at position 5 ensures that the 5′-splice site is selected.

TECHNOLOGY

# Three prime mice

**Modelling quantitative traits in mice has just got easier thanks to a new technique, reported by Masao Kakoki and colleagues, that allows the expression of a gene to be subtly modified by tweaking the sequences at its 3′ end. Gene expression can be varied over a 100-fold range, without moving the gene from its normal chromosomal position, or altering its promoters or introns.**



**Methods for mapping the genetic alterations that underlie complex traits are getting ever more sophisticated, but modelling these often subtle changes is usually left to rather crude manipulations. In mice, for example, the expression of transgenes that are microinjected into a one-cell embryo is unpredictable as it depends on how many copies are integrated into the genome and where in the genome the integration occurs. To obtain a more predictable and less intrusive means of controlling gene expression, Kakoki and colleagues turned to the 3′-untranslated region (UTR), which has been shown many times before to influence the stability of a gene's mRNA. Their aim was to find defined changes in the 3′ UTR that would allow the expression of a gene to be altered *in situ* in a predictable way.**

**To test their idea, they created mouse embryonic stem (ES) cells in which a *GFP* gene was inserted into an endogenous locus. Altering the 3′ sequences of the transgene allowed them to assess the effect of various 3′ regions on *GFP* expression by monitoring the level of fluorescence emitted by the cells. Protein expression (which correlated with mRNA levels) varied over a 100-fold range according to whether the 3′ sequences were derived from say, the *Fos* gene, which has a very**

Both the *in silico* and *ex vivo* approaches point to the same conclusion — that the decay of CG dinucleotides in the human genome, as a result of hypermethylation, drives *Alu* exonization by creating new splice sites. Might this be an unexpected by-product of dampening down *Alu* transposition, given that mutating these CG dinucleotides renders the *Alu* retrotransposase inactive?

The authors provide more food for thought on genome evolution and organismal complexity by inviting us to consider the possibility that our genomes are littered with pre-exonic *Alu* elements, poised to be exonized. As such, they might serve as a reservoir for human-specific exons that might adapt to perform new functions, thereby promoting speciation of the human lineage.

*Magdalena Skipper*

**References and links**

**ORIGINAL RESEARCH PAPER** Sorek, R., Lev-Maor, G., Reznik, M. *et al.* Minimal conditions for exonization of intronic sequences: 5′ splice site formation in *Alu* exons. *Mol. Cell* **14**, 221–231 (2004)

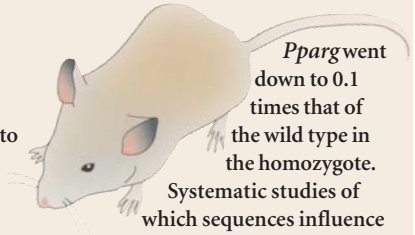
**FURTHER READING** Lev-Maor, G., Sorek, R. Shomrom, N. & Ast, G. The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. *Science* **300**, 1288–1291 (2003)

**WEB SITE**

**Author's laboratory:** http://www.tau.ac.il/~gilast

unstable message, or from the bovine growth hormone, which has a very stable message. These data could then be used to introduce systematic changes into a 3′ UTR, while being careful to retain features that are specific to that gene: for example, the inclusion of a GU/U-rich element downstream of the 3′ UTR increased expression levels by 2–3-fold.

This effect was not limited to ES cells — in fact, it persisted when the ES cells were differentiated into cardiomyoctes and trophoblastocytes. And what happens in cells is also true of whole animals. Two lines of mouse that were each derived from an ES cell that contained a different engineered gene showed altered expression in the predicted direction: expression of *Agtr1* went up by 1.8 times in the heterozygous animal and that of



*Pparg* went down to 0.1 times that of the wild type in the homozygote. Systematic studies of which sequences influence mRNA stability can therefore be as fruitful for controlling gene expression as those that focus on transcription or translation. Moreover, the panel of tested 3′ regions generated by this study provides a toolbox from which combinations of control elements can be selected to modify the expression of other genes.

*Tanita Casci*

**References and links**

**ORIGINAL RESEARCH PAPER** Kakoki, M. *et al.* Altering the expression in mice of genes by modifying their 3′ regions. *Dev. Cell* **6**, 597–606 (2004)

**WEB SITE**

**Nobuyo Maeda's and Oliver Smithies' laboratories:** http://www.unc.edu/~krfloyd

# IN BRIEF

PLANT GENETICS

## A $Ca^{2+}$/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning.

Mitra, R. M. *et al. Proc. Natl Acad. Sci. USA* **101**, 4701–4705 (2004)

Because of their often large and complex genomes, mapping mutations in crop plants by positional cloning can be a painstaking process. Raka Mitra and colleagues have successfully used transcript abundance to clone the *Medicago dmi3* gene, which is involved in symbiotic nodule development. Importantly, they show that the method is valid in barley, and by extension, could be useful in many crop plants.

MOUSE MODELS

## Increased airway epithelial $Na^+$ absorption produces cystic fibrosis-like lung disease in mice.

Mall, M. *et al. Nature Med.* **10**, 487–493 (2004)

Cystic fibrosis is characterized by altered $Na^+$ and $Cl^-$ transport in the lung, leading to infection and death. To investigate how ion transport causes disease, Mall and colleagues generated mice that overexpressed epithelial $Na^+$ channels in the lung. The animals developed spontaneous lung disease associated with low volumes of airway-surface liquid and poor bacterial clearance, indicating that accelerated $Na^+$ absorption alone causes cystic fibrosis.

EVOLUTION

## Detecting selection using single genome sequence of *M. tuberculosis* and *P. falciparum.*

Plotkin, J. B. *et al. Nature* **428**, 942–954 (2004)

Selection can now be detected by looking for signs of non-synonymous substitutions, without resorting to sequence comparisons. The 'volatility' of a codon — the proportion of codons that encode a different amino acid when changed at only one nucleotide — is used to calculate the probability that the most recent nucleotide substitution was non-synonymous. Tests on two pathogen genomes showed that the method requires fewer data than comparative approaches and can be used for any 'post-genomic' organism.

BIOINFORMATICS

## Ensembl Special.

*Genome Res.* **14**, (2004)

This series of articles begins with an overview of Ensembl, the bioinformatics project that was initiated in 1999 to help realize the potential of the human genome sequence. It describes the challenges of setting up Ensembl, as well as some of the motivation behind different aspects of the system. The nine other articles in this series focus more on the hidden details of Ensembl — the deliverables (annotation, integration of data and comparative genomics) and their technical implementation (storage, manipulation and computational requirements).