RESEARCH HIGHLIGHTS

⟩ EVOLUTION

Neanderthals help to tell our story

The availability of Neanderthal genomic sequence data is allowing us to date key events in our own evolution.

So far, only one gene has been convincingly associated with human language and speech — forkhead box P2 (*FOXP2*). Despite being a highly conserved gene, the human and chimpanzee versions differ at two positions in exon 7, and these substitutions have been implicated in our unique ability for speech.

Krause *et al.* found both substitutions in two Spanish Neanderthal samples. Further, they concluded that, for both substitutions, at least one of the individuals was homozygous for the derived allele that is predominant in modern populations.

Next, the authors analysed the Neanderthal samples for evidence of a selective sweep close to exon 7 of *FOXP2* that, from modern human genetic diversity data, was previously proposed to have occurred within the past 200,000 years. Seven polymorphic regions from the intronic region upstream of exon 7 were successfully amplified from the Neanderthal samples and, for six of these, all products from both Neanderthals represented the derived allele. These results suggest that the selective sweep in the *FOXP2* region began before the split from the human–Neanderthal common ancestor, which existed 300,000 to 400,000 years ago — much earlier than previously thought.

So, Neanderthal sequences can provide a useful tool to investigate our own evolution, making the extreme care that must be taken when retrieving genetic information from ancient samples worthwhile.

Carrie Patis

ORIGINAL RESEARCH PAPER Krause, J. *et al.* The derived *FOXP2* variant of modern humans was shared with Neandertals. *Curr. Biol.* **17**, 1908–1912 (2006)

⟩ GENOMICS

Decoding the regulatory genome

A new computational tool can predict gene regulatory elements from diverse taxonomic groups and types of expression data.

Existing methods for predicting regulatory elements are limited to specific species or to particular types of element and expression data — this is because they rely on underlying assumptions about the relationship between elements and expression patterns. Elemento and colleagues developed a method — FIRE (Finding Informative Regulatory Elements) — that dispenses with such assumptions, instead directly quantifying the relationship between gene expression levels and potential regulatory motifs using information theory

The authors tested their approach on gene clusters derived from existing *Saccharomyces cerevisiae* microarray data, using FIRE to identify short sequence motifs that are highly informative about gene expression. FIRE predicted 17 DNA motifs from promoter regions and 6 RNA motifs from 3′ UTRs; 14 of these matched known yeast regulatory elements, indicating that the approach works well. Further validation came from conservation of the predicted motifs in a closely related yeast species, and from the fact that the target genes of a particular motif generally have related functions.

Unlike existing methods, large numbers of false positives are not a problem for FIRE: rates were found to be zero or close to zero across a range of genomes and data types — an important advantage, especially for application to larger genomes.

Elemento and colleagues applied and validated their method for different types of expression data from other species including fruitfly, worm, mouse and human. Their predictions included sequence motifs from the malaria parasite *Plasmodium falciparum*, which is inaccessible using other methods

> "…this versatile approach promises to be an important resource for studying the sequence determinants of gene regulation across diverse species."

⟩ GENE THERAPY

Gene editors deliver

A major goal for gene therapy is to correct genetic defects in patient-derived stem cells, which could be put back into the patient to alleviate disease symptoms. However, the most widely used approach involves the genomic insertion of a transgene, which — owing to the non-specific nature of insertion — runs the risk of harmful side-effects. Advances using site-specific DNA-cutting enzymes are now showing promise for safer forms of gene therapy.

Zinc-finger nucleases (ZFNs) are engineered DNA editing enzymes that consist of a DNA binding zinc-finger domain and the nuclease domain of a restriction endonuclease. These enzymes produce double-stranded breaks, which can be targeted to a site of choice by altering the zinc-finger DNA binding specificity. ZFNs have been used previously to correct point mutations by inducing homologous recombination that brings about gene conversion from a donor DNA.

This approach avoids the insertion of new material into the genome. However, it requires the efficient delivery of three different constructs — two ZFNs (as the enzymes act as heterodimers) and the donor DNA — and current systems for delivery into clinically relevant target cells are highly inefficient. To overcome this problem, Lombardo and colleagues used an integrase-defective lentiviral vector (IDLV), which can infect most primary cells and deliver transgenes efficiently, but doesn't integrate into the host genome. For three human cell types, the authors showed that delivery of the necessary constructs using three IDLVs resulted in the desired gene editing outcome. For the different cell types, a variable but significant proportion of cells showed gene correction. In cases in which an endogenous gene was edited, there was normal expression of the gene product.

BRANDX

owing to its AT-rich genome and poorly understood gene regulatory mechanisms.

This approach also provided functional insights, for example, the importance of coordination between transcriptional and post-transcriptional regulation in several species: in yeast, worm and human, regulatory modules were identified with both DNA and RNA elements. However, RNA elements were not predicted in *P. falciparum*, suggesting that such coordination is absent or less important in this parasite.

FIRE has potential limitations — for example, it might work less well for highly degenerate motifs and those that are less informative about expression levels. However, this versatile approach promises to be an important resource for studying the sequence determinants of gene regulation across diverse species.

Louisa Flintoft

ORIGINAL RESEARCH PAPER Elemento, O., Slonim, N. & Tavazoie, S. A universal framework for regulatory element discovery across all genomes and data types. *Mol. Cell* **28**, 337–350 (2007)

Could the uses of ZFNs be extended by using them for site-specific gene addition? The authors showed that this can be achieved by using a construct in which the gene to be added is flanked by sequences that are homologous to the target site. This was achieved for the *IL2RG* gene (which is mutated in severe combined immunodeficiency), allowing a promoterless region of this gene to be inserted within the endogenous locus. The result was expression of the transgene under the control of the *IL2RG* promoter, suggesting that physiological expression could be achieved for gene therapy.

The methods described above have a limitation in that new ZFNs would need to be designed for different mutations and different target genes. In the case of gene addition, this could be overcome by finding a genomic location into which any transgene could be inserted, using the same ZFNs in each case. The authors picked the *CCR5* gene as a potential insertion site, as homozygous null mutations at this gene seem to have no ill effects in humans.

The authors successfully achieved specific and efficient gene addition at this site in a range of human cell lines, including both haematopoietic progenitors and embryonic stem cells. In the latter case, expression of an inserted GFP transgene was maintained after induction to a neural progenitor fate, suggesting that this approach does not affect the differentiation and self-renewal properties of stem cells.

Thorough safety assessments will be needed before such approaches can enter the clinic, and questions about possible off-target effects must be addressed. However, this study highlights an exciting range of possible applications for ZFNs in gene therapy.

Louisa Flintoft

ORIGINAL RESEARCH PAPER
Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger endonucleases and integrase-defective lentiviral vector delivery. *Nature Biotechnol.* 28 October 2007 (doi:10.1038/nbt1353)

IN BRIEF

➜ GENE REGULATION

Functional specificity among ribosomal proteins regulates gene expression.

Komili, S. *et al. Cell* **131**, 557–571 (2007)

Several eukaryotic species have more than one gene copy for various ribosomal proteins. This study examined the effect of knocking out individual ribosomal protein paralogues in the yeast *Saccharomyces cerevisiae* on phenotypes that ranged from the localization of an individual mRNA to gene expression profiles. The results indicate that ribosomal protein paralogues are not functionally redundant but regulate specific genes and cellular functions. The authors propose that, together with their post-translational modifications, the existence of more than one version of multiple ribosomal proteins might underlie a 'ribosome code' for gene expression.

➜ RNA WORLD

Riboswitch control of gene expression in plants by splicing and alternative 3′ end processing of mRNAs.

Wachter, A. *et al. Plant Cell* 9 November 2007 (doi:10.1105/tpc.107.053645)

This study highlights the importance of gene regulation by metabolite-sensing riboswitches in plants, and reveals a previously unknown control mechanism for these elements. A bioinformatic analysis revealed that a riboswitch that senses thiamin pyrophosphate (TPP) is present in the 3′ UTR of a thiamin-biosynthetic gene, *THIC*, in a wide range of plants. The authors showed that this riboswitch regulates *THIC* levels through the alternative processing of mRNA 3′ ends.

➜ EPIGENETICS

A novel Snf2 protein maintains *trans*-generational regulatory states established by paramutation in maize.

Hale, C. J. *et al. PLoS Biol.* **5**, e275 (2007)

These authors have identified a protein that is required for the heritable transmission of a paramutation-induced regulatory state in maize. *required to maintain repression1* (*rmr1*) encodes a novel member of the Snf2 family of small RNA metabolism proteins. RMR1 is shown to mediate DNA methylation and small RNA accumulation at a transposable element fragment in a paramutated allele. Rather than being required for its initiation, these functions are required for the maintenance of the altered regulatory state, probably by co-transcriptional repression.

➜ EVOLUTION

Comparative genomics search for losses of long-established genes on the human lineage.

Zhu, J. *et al. PLoS Comp. Biol.* 30 October 2007 (doi:10.1371/journal.pcbi.0030247.eor)

A new comparative genomics method has identified losses of well-established genes in the human lineage since the common ancestor of primates and rodents. This method excludes recent gene losses that have occurred following segmental duplication or retrotransposition. The authors identified 26 long-established genes, with a range of functions, that have been lost from the human genome, including both known and new examples. Among them is an ancient acyltransferase gene that is present in archaea, bacteria and many eukaryotes, but was lost from the human–chimpanzee ancestor.