SYSTEMS BIOLOGY

Timing is everything

Even a well-crafted comedic performance will leave an audience cold unless the timing is right. It now emerges that, like comedic expression on stage, gene expression in metabolic pathways also needs to be well-timed to be effective.

Alon Zaslaver and colleagues studied the timing and strength of the expression of genes in a well-characterized type of pathway: amino-acid biosynthesis (AAB) in *Escherichia coli*. They developed an ingenious system for monitoring gene expression that involved cloning the promoter regions of ~50% of all known *AAB* genes into plasmids, upstream of reporter genes (*Lux* or *GFP*). By using an automated multiwell fluorimeter, the authors could monitor promoter activity in these 52 reporter strains every few minutes to an accuracy of 10%.

Zaslever *et al.* were able to address sophisticated questions about the dynamics of gene expression in the AAB pathway. First, they showed that, as might be expected, excess amounts of an individual amino acid in the medium led the genes in that amino acid's biosynthetic pathway to be downregulated. However, the really interesting results came when they examined reactivation of such pathways when the corresponding amino acids were removed.

In the three amino-acid systems that were studied in detail — arginine, methionine and serine — the expression of genes in unbranched pathways was reactivated in the same order as they were positioned in the pathway: that is, the first gene in the pathway was also the first to be reactivated, and so on. Zaslever and colleagues also showed another striking correlation between functional hierarchies and gene-expression patterns: in all three cases, the closer a gene was to the beginning of the pathway, the higher was its maximal level of expression.

To explain these remarkable patterns, the authors went on to mathematically model enzyme production in such unbranched 'production pipeline'-style pathways. Their model clearly showed that the hierarchies in the timing and magnitude of gene expression in these pathways minimized the time that was required and the metabolic cost of producing the final enzyme. Higher and earlier expression of the genes at the top of the pathway provides a boost in the intial enzymes required, allowing a faster response to withdrawal of the amino acid.

The results of Zaslever *et al.* are fascinating just for the insight that they provide into the delicate regulation of amino-acid synthesis in bacteria. Experiments such as these might be as important to the development of systems biology as similarly elegant experiments in *E. coli* were to that of molecular biology and genetics.

Nick Campbell

References and links

ORIGINAL RESEARCH PAPER Zaslaver, A. *et al.* Just-in-time transcription program in metabolic pathways. *Nature Genet.* **36**, 486–491 (2004)

FURTHER READING Barabasi, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004)

WEB SITE
Uri Alon's laboratory:
http://www.weizmann.ac.il/mcb/UriAlon/

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?

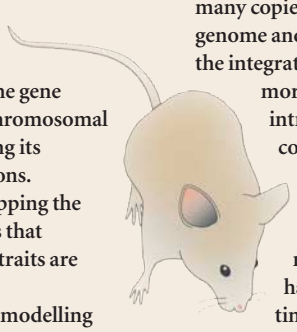
To answer that, the authors compared AEx sequences with those of their intronic ancestors. They focused on the 5′-splice site: >98% of human introns begin with GT and very few with GC, although these are the ones that are supposed to be mainly involved in alternative splicing. Strikingly, the most significant 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)

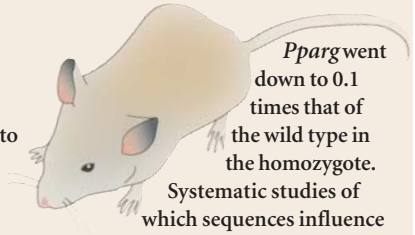
FURTHER READING Lev-Maor, G., Sorek, R. Shomrom, N. & Ast, G. The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. *Science* **300**, 1288–1291 (2003)

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).

ETHICS WATCH

Ethics guidelines for population genetics research

Large-population biobanks and their related databases have multiplied in recent years. The United Kingdom, Estonia and Iceland have led the way, and others are expected to follow soon. Even an international consortium, the HapMap Project, is currently regrouping partners from six countries and comparing haplotype variations in various populations around the world.

CARTaGENE is a population genomics project that plans to map genetic variation in a large reference population of Quebec. At the request of its leader, Claude Laberge, a multidisciplinary research team from the Genetic and Society Project (University of Montreal) has been studying the ethical and legal challenges that are raised by population genetics research. The team has found that despite the existence of many guidelines on genetics research, none was tailored to the specific challenges raised by these types of project.

The research team therefore drafted a *Statement of Principles on the Ethical Conduct of Human Genetic Research Involving Population*[1]. The statement, which was developed by a multidisciplinary team composed of M. Deschênes, G. Cardinal, B. Knoppers, T. Hudson, D. Labuda, G. Bouchard, É. Racine, C. Fecteau, S. Truong and C. Laberge, includes a vision of ethical conduct in population genetics research, based on ten fundamental principles that should be upheld in undertaking these types of project: individuality, diversity, complexity, reciprocity, solidarity, security, accountability, equity, citizenry and universality. The principles are articulated in practical recommendations and procedures to guide researchers in setting up these types of project. The statement covers topics such as consultation, recruitment, consent, confidentiality, governance, communication of research results, commercialization and contribution to the welfare of the population and of humanity.

Let us consider public consultation. In a population genetics research project, the entire population must be recognized as an important partner. Early international experiences have shown that public consultation is key to the overall success of such projects. Indeed, although the whole population (including participants and non-participants) will eventually benefit from such initiatives, it is also collectively taking potential risks. In addition, an important amount of public and private research resources must be devoted to a population genetics research project. These types of initiative need consent not only from the participants, but also meaningful public support and engagement. Open and continued dialogue, throughout the project, will help to foster the trust of the population; it can even serve as a preparatory stage for individual informed consent.

According to the statement: "Respecting the principles of reciprocity, diversity and accountability requires that research on a given population be based upon an open dialogue between the population and the research team. A guiding mechanism for population genetic research is prior and ongoing public consultation." In suggesting how to implement these recommendations, the statement emphasizes the importance of a proper information and public-engagement process that will enable the public to take part in the debate and voice their concerns or support for such a project.

Setting up a population genetics research project poses a scientific and logistic challenge, but also a legal and ethical challenge. Statements such as the one based on the CARTaGENE project promise to provide useful guidance to the research community and their partners, including participating populations.

Mylène Deschênes, B.C.L., LL.B, LL.M.
e-mail: MDeschenes@cihr-irsc.gc.ca

[1]Knoppers, B. M. (Ed.) in *Population and Genetics Legal and Socio Ethical Perspectives* 641 (Martinus Nijhoff, The Netherlands, 2003); *RMGA* [online], <http://www.rmga.qc.ca/en/index.htm> (2003)

GENE REGULATION

microRNA, control yourself!

From plants to fish to humans, microRNAs — arguably the most sought-after molecules in genetics — work by interfering with the expression of complementary mRNAs. Now, Hervé Vaucheret and colleagues show that microRNAs (miRNAs) are not averse to biting the hand that feeds them: the *Arabidopsis thaliana* ARGONAUTE PROTEIN 1 (AGO1) is needed for the proper function of miRNAs and is itself regulated by a miRNA. Proving the existence of this negative feedback involved engineering artificial miRNAs, potentially spurring a new approach to targeted gene silencing.

Circumstantial evidence indicated that the AGO1 protein worked in the miRNA pathway, specifically as a component of the RNA-induced silencing complex (RISC), the ribonucleoprotein complex in which miRNAs act. Animal AGO homologues are found in the RISC, but it was the combination of genetic analysis of *ago1* mutants in *A. thaliana* and some clever sequence manipulations that provided the proof.

Vaucheret and colleagues found that mRNAs that are normally targeted for cleavage by miRNAs accumulate in *ago1* null mutants; this indicates that AGO1 could be needed for proper miRNA function and that this is the only AGO family member — out of ten — to be predominantly associated with miRNAs in the plant RISC.

Among the mRNAs that accumulate in the mutants is the *AGO1* transcript itself, prompting the idea that *AGO1* mRNA is negatively regulated by a miRNA. A miRNA that is complementary to *AGO1* (miR168) suggested itself — correctly, as it turns out — as the negative-feedback regulator. An otherwise wild-type *AGO1* gene was engineered to reduce its complementarity to miR168; this increased the levels of *AGO1* mRNA and caused developmental defects that resemble those of *dcl1*, *hen1* or *hyl1* mutants that are impaired at other steps in the miRNA pathway. To be certain that *AGO1* mRNA was regulated by miR168, the authors reversed the defects by generating compensatory mutations in the miR168, so that it now bound to the mutant *ago1* mRNA and presumably allowed its normal degradation.

As the authors note, further experiments will be needed to prove that AGO1 is in the RISC, and, if it is, that it is the only AGO family member to be involved in miRNA function.

Tanita Casci

References and links

ORIGINAL RESEARCH PAPER Vaucheret, H. *et al*. The action of *ARGONAUTE1* in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* **18**, 1187–1197 (2004)
FURTHER READING Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004)
WEB SITES
David Bartel's laboratory: http://web.wi.mit.edu/bartel/pub
Hervé Vaucheret's laboratory: http://www-ijpb.versailles.inra.fr/bc/equipes/Epigenetics/index.html

MOUSE GENETICS

What a screen!

Most yeast, fly and worm geneticists know how useful modifier genetic screens can be; screening for mutants that enhance or suppress a certain phenotype can be used to build up the signalling cascades that underlie that trait. Marina Carpinelli, Doug Hilton and colleagues now show that suppressor screens are equally applicable to vertebrate models in general and to mice in particular.

The authors focussed on a mouse model of thrombocytopenia — a disease that is caused by a lack of blood platelets and that is seen in $Mpl^{-/-}$ animals. They treated 300 male, *Mpl* homozygous knockout mice with the mutagen *N*-ethyl-*N*-nitrosourea (ENU). These mutagenized mice were then crossed with isogenic knockout females. A total of 5 of the 1,575 F_1 offspring from these crosses suppressed the mutant phenotype: their platelet counts far exceeded those of their thrombocytopenic parents. Subsequent crosses with untreated $Mpl^{-/-}$ mice showed that two of the suppressed mice (*Plt3* and *Plt4*) harboured dominant ENU-induced mutations.

So, without too much trouble, Carpinelli *et al.* were able to use a large-scale suppressor screen to isolate new mutant mice that are relevant to the study of the platelet-production pathway. However, it is their subsequent work that shows just how useful mutants isolated from suppressor screens can be. First, they intercrossed mice that were heterozygous for the new mutations to obtain homozygotes in an $Mpl^{-/-}$ background. These mice had platelet levels that were above normal, in contrast to the below wild-type levels in the heterozygotes, which clearly indicated that *Plt3* and *Plt4* are semi-dominant mutations.

Further crosses of *Plt3*/*Plt4* compound heterozygotes to $Mpl^{-/-}$ mice showed that the two mutations are tightly linked and possibly allelic. Using standard segregation analysis of 148 microsatellite markers, the authors localized the mutations to a region on chromosome 10 that contains *cMyb*, a gene that was previously linked to elevated platelet levels. They then identified candidate mutations in *cMyb*, in both *Plt3* and *Plt4* mice, that caused single amino-acid substitutions in functionally significant regions of the cMyb protein. Follow-up protein transactivation assays confirmed that these mutations reduce the activity of cMyb. Coupled with the finding that *cMyb* heterozygous knockout mice have higher platelet levels than wild-type homozygotes, these studies provide compelling evidence that the *cMyb* mutations underlie the *Plt3* and *Plt4* gain-of-function phenotypes.

Detailed haematological analyses of heterozygous and homozygous mutants showed that *Plt3* and *Plt4* increase the production of megakaryocytes and their progenitors, from which platelets are derived. So, we now have a more complete picture of the role of *cMyb* in haematopoiesis, but the real beauty of mutants that are identified from a supressor screen is that their epistatic interactions with the original mutant can be genetically analysed. In this case, Carpinelli *et al.* clearly showed that the effect of mutations in *cMyb* on platelet and megakaryocyte levels was independent of the *Mpl* genotype. So, the semi-dominance of mutations in *cMyb*, regardless of the genetic background, indicates that downregulation of *cMyb* is probably an important step in platelet production and that the gene is in the same signalling pathway as *Mpl*.

It seems to have taken vertebrate geneticists some time to catch on to the allure of suppressor screens. However, if the success of this initial application of the approach is anything to go on, it seems likely that we will soon see many more examples.

Nick Campbell

References and links

ORIGINAL RESEARCH PAPER Carpinelli, M. R. & Hilton, D. J. *et al.* Suppressor screen in $Mpl^{-/-}$ mice: *c-Myb* mutation causes supraphysiological production of platelets in the absence of thrombopoietin signaling. *Proc. Natl Acad. Sci. USA* 7 Apr 2004 (doi:10.1073/pnas.0401496)

WEB SITE

The art and design of genetic screens: http://www.nature.com/nrg/focus/screens

IN THE NEWS

Life without men

Dandelions do it all the time, aphids do it occasionally, and now mice can do it too, but only in the laboratory. Parthenogenesis — the reproduction of a gamete without fertilization — has been achieved in mice by Tomohiro Kono and colleagues, who announced the birth of live mice that were derived from two mothers but no father.

"This achievement, published in *Nature*, might be seen as being of comparable significance to the birth in 1996 of Dolly the sheep. And it may prove almost as controversial" (*Financial Times*).

The researchers knocked out a gene in an immature mouse egg to give it a 'male-like' imprinted character, and then combined the cell's genetic material with that of a normal mature egg. "From around 600 eggs, only two live mice were born" (*Financial Times*). "One mouse, named Kaguya, after a Japanese folk tale in which a princess is born from a bamboo stump, grew to adulthood and has become a mother herself — though by conventional means" (*Washington Post*).

"The practical implications are obscure, since the method is even more complex, inefficient and unsafe than cloning", said Ian Wilmut (*Daily Telegraph*). But could such a 'virgin birth' occur in humans too? "Researchers were quick to head off suggestions that the technique could be used to treat infertility" (*BBC News Online)* or that it "makes men obsolete" (*Daily Telegraph*). However, the use of the technique might "circumvent the political and ethical obstacles to using stem cells" (*BBC News Online)*.

Whatever its application, this work is "sure to stimulate conversation about the intrinsic importance of male–female pairing, at both the biological and the social level" (*Washington Post*).

Tanita Casci

GENE EXPRESSION

The ATGC of gene expression

The advent of whole-genome approaches, microarray technologies and improved computation has given us new insights into the regulation of gene expression, although relating the expression of regulatory genes to that of the genes under their control remains difficult. Beer and Tavazoie have recently reported on a systematic approach to the problem, based on the identification of 5′-upstream DNA sequences of the genes of interest.

The relationship between mRNA levels of a transcription factor and the gene it regulates might not be direct owing to, for example, post-transcriptional regulation. But correlating the abundance of the active transcription-factor protein to the mRNA level of the gene it regulates is technically challenging. Fortunately, the new method of Beer and Tavazoie sidesteps these difficulties by building sequence-to-gene networks, using short 5′-upstream DNA sequence elements as a surrogate for active transcription-factor protein levels.

Working in yeast, the authors began by looking for groups of genes that were coexpressed under a range of experimental conditions (for example, heat shock or diamide treatment). In all, they assigned 2,587 genes to 49 'expression patterns'. Next, they identified overrepresented sequence motifs within 800 bp upstream of the genes in each pattern. The rationale was that such sequence elements were likely to be involved in the regulation of the corresponding genes. Indeed, many of the motifs that were pinpointed closely matched known regulatory elements. The authors used a Bayesian approach to apply further constraints to the motifs, such as orientation and distance to ATG, so that regulatory 'rules' could be inferred. They also took into account combinations of motifs. For example, the two elements PAC and RRPE were both found upstream of a high proportion of genes in a particular expression pattern, indicating that they coregulate genes in this group. It emerged that the order and distance between the two elements also strongly affected the degree of correlation between genes.

So, having identifed the upstream sequence elements that are involved in transcriptional regulation, along with the positional and combinatorial constraints that govern their role, the authors tested the predictive power of their approach. Based on promoter sequences alone, they attempted to predict the expression patterns of 'test' sets of genes, not used while learning the rules. Impressively, their predictions were accurate in 73% of the genes, and fine-tuning of the system is likely to improve on this.

As high-quality mRNA expression data becomes readily available in different organisms, the process reported here will be invaluable to our understanding of the regulation of genes, and of cellular behaviour more generally. The authors have already begun to apply their new approach to multicellular organisms, starting with *Caenorhabditis elegans*. In a preliminary study using expression data collected over a time course from oocyte to adult, they were able to predict the expression patterns of half the genes. As these studies are expanded to take into account further complexities, such as downstream or intronic regulatory elements, it should become possible to unravel the transcriptional regulatory mechanisms behind diverse spatiotemporal processes.

Ruth Kirby, Nature Publishing Group

References and links

ORIGINAL RESEARCH PAPER Beer, M. & Tavazoie, S. Predicting gene expression from sequence. *Cell* **117**, 185–198 (2004)

WEB SITE

Tavazoie's laboratory:
http://genomics.princeton.edu/tavazoie/index.htm

HUMAN GENETICS

Narrowing down the candidates for asthma

In 2001, a genome-wide scan in asthma identified chromosome 7p as 1 of 6 possible asthma-susceptibility loci. Kere and colleagues have now narrowed the candidate region from a 20-cM to a 133-kb region, which contains 2 genes — *GPRA* and *AAA1*.

The authors genotyped 874 subjects from the Finnish Kainuu subpopulation by interspersing successive rounds of genotyping that increased the density of markers (SNPs and microsatellites) with analysis using the haplotype pattern mining (HPM) algorithm, which searches large sets of unrelated haplotypes for allele patterns that are shared between several haplotypes. Having identified a strong association of a conserved 47-kb haplotype pattern in this way, Kere and colleagues sequenced 133 kb that encompassed the 47-kb region from a homozygous asthmatic patient. Comparison with the public sequence identified 80 new polymorphisms.

Asthmatics from North-eastern Quebec and individuals with high serum immunoglobin E levels from North Karelia, Finland, also had a 133-kb haplotype pattern with the same limits, for which most SNPs were conserved. For the 3 populations combined, 7 alternative haplotypes were formed by 13 SNPs across the most-conserved region of 77 kb. Sequencing of the 133-kb region from 6 individuals who were homozygous for the remaining 6 haplotypes confirmed that they had different SNP compositions. Phylogenetic analysis showed these risk haplotypes to be related and distinct from non-risk haplotypes. By SNP-tagging the risk haplotypes, the authors confirmed that they confer risk in all three populations, which is consistent with the common-disease/common-variant hypothesis.

With the 133-kb region confirmed as a susceptibility locus, the authors looked for open reading frames. They found 2: exons 3 to 5 of a gene that they named *GPRA* (for G-protein-coupled receptor for asthma susceptibility) and, on the opposite strand, exons 3 to 10 of a gene that they called *AAA1* (for asthma-associated alternatively spliced gene 1). Both *GPRA* and *AAA1* are probably alternatively spliced, but *AAA1* might not be a protein-coding gene (for example, *in vitro* translation of its longest open reading frame, which encodes only 74 amino acids, did not produce a stable polypeptide, and no recombinant protein was produced by transiently transfected cells).

GPRA's two main transcripts, A and B, encode proteins of 371 and 377 amino acids, respectively. Whereas the A isoform is mainly expressed by smooth-muscle cells, the B isoform is mainly found in epithelial cells; however, in asthma patients, the B isoform is strongly expressed by smooth-muscle cells.

So, Kere and colleagues suggest that the balance between the A and B isoforms might be altered by polymorphisms in the risk haplotypes. In support of a role for GPRA in asthma pathogenesis, *Gpra* mRNA is significantly upregulated in a mouse model of ovabulmin-induced lung inflammation.

Although these are not the first candidate asthma-susceptibility genes, this study identifies *GPRA* as the strongest candidate so far. And, most encouragingly, the properties of GPRA as a G-protein-coupled receptor, which might be used by a novel ligand as part of a new signalling pathway in asthma, mean that it also has potential as a possible drug target.

Natalie Wilson

References and links

ORIGINAL RESEARCH PAPER Laitinen, T. *et al*. Characterization of a common susceptibility locus for asthma-related traits. *Science* **304**, 300–304 (2004)

WEB SITES

Juha Kere's laboratory:
http://www.cbt.ki.se/groups/jke;
http://www.ktl.fi/diseasegenetics/HUIPPU_MULTIFACTORIAL.HTM

TECHNOLOGY

A new approach to microarrays

Reporting in *Genome Research*, Mark Chee and colleagues introduce an approach that looks set to realize the potential of randomly assembled microarrays.

Unlike conventional microarrays, the location of probes on randomly assembled microarrays is unknown. However, random assembly allows more effective miniaturization and higher packing densities to be achieved, which improves cost-effectiveness and statistical power. Until now, though, identifying the probes after assembly required complex dye chemistries and labelling processes, which limited attempts to develop randomly assembled microarrays.

Chee and colleagues have developed a simple and efficient algorithm that can accurately decode many thousands of DNA-linked objects, in this case, DNA-coated beads, using only a few labels and several sequential rounds of hybridization.

To demonstrate this approach, the authors prepared 1,520 different oligonucleotide-linked bead types (each bead type represents a unique 22–24-base oligonucleotide sequence that acts as both a probe and as a decodable sequence). These bead types were pooled and assembled into the wells of 96 optical-fibre bundles — each bundle has ~50,000 wells and each well holds one bead — to create 96 arrays at a time. The ~30-fold redundancy per bead type per array increases both the precision and the robustness of this approach, which also has the advantage of using a standard 96-well microtitre plate layout.

For the decode process, three different sets of oligonucleotides that are complementary to the bead sequences were used — two sets labelled with different fluorescent dyes and the other set unlabelled. Oligonucleotides from these sets were used to create eight decoder pools, each with a different combination of oligonucleotides. Although seven successive hybridizations can decode the bead types, the authors included an extra hybridization step that allows for error checking and provides the first functional test for the quality of every probe on a microarray. The fluorescent signals imaged from all eight hybridization stages are decoded by an algorithm that uses the fluorescence intensity profiles of each bead to provide a combinatorial code that can identify them. Encouragingly, decoding randomly assembled microarrays using this approach has a very low error rate, with fewer than five misclassified beads per array.

Not only is this high-performance and cost-effective approach general, and could therefore be applied to any DNA-linked fixed objects or molecules, but it can also be scaled up and is proven in large-scale genomics applications, such as SNP genotyping for the International HapMap Project. In the future, this technology could also help to speed the use of microarray-based assays for large-scale clinical applications.

Natalie Wilson

References and links

ORIGINAL RESEARCH PAPER
Gunderson, K. L. *et al*. Decoding randomly ordered DNA arrays. *Genome Res*. **14**, 870–877 (2004)