RESEARCH HIGHLIGHTS

⮕ **POPULATION GENETICS**

Size doesn't matter

Mitochondrial DNA (mtDNA) diversity is widely used as a marker for species abundance. However, Galtier and colleagues have now shown that, unlike nuclear DNA, mtDNA diversity is not proportional to population size and that this is probably due to positive selection.

The number of neutral polymorphisms in a species should be proportional to the effective population size. As the overwhelming majority of mutations are assumed to be at least almost neutral, the overall sequence diversity of a species should also be governed by the population size. However, other factors also affect diversity, such as population history, life history and population structure, making it difficult to test the effect of population size on diversity in any one species.

Therefore, Galtier and colleagues carried out a meta-analysis of over 1,000 species from a range of taxa. They looked at diversity at three levels — allozyme, nuclear DNA and mtDNA —

expecting all to show the same patterns as a result of population size: invertebrates should have more polymorphism than vertebrates; marine species more than terrestrial or freshwater species; and large animals more than small animals. These patterns were observed for allozymes and nuclear DNA, but mtDNA diversity did not vary with population size.

The authors considered possible explanations for the failure of mtDNA to conform to the neutral theory of molecular evolution. They dismissed three explanations as improbable: if variation in mutation rate was the explanation, the mutation rate would have to vary inversely with population size across taxa, which seems unlikely; demographic effects that might have lowered diversity should have the same effects on nuclear DNA; and negative, purifying selection decreases diversity but still leaves larger populations more diverse.

Instead, the authors analysed the effects of positive selection. The fixation of advantageous mutations leads to a loss of diversity at linked sites — the alleles at linked loci that happen to occur in the advantageous haplotype will also be fixed. If this process is frequent, overall diversity will be reduced. However, this process will
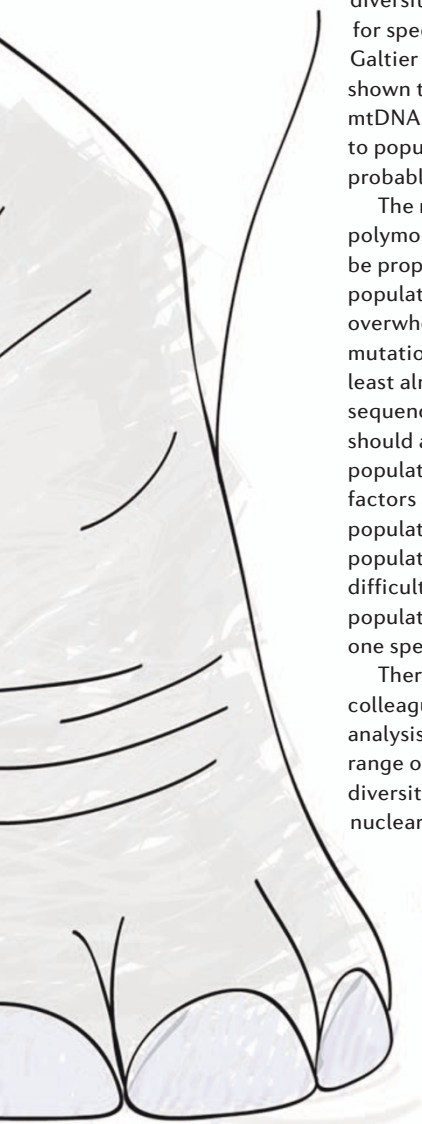
depend on population size: the larger the population, the more advantageous mutations will arise, and the greater the decrease in diversity at linked sites. This should exactly compensate for the increase in the rate at which diversity arises in large populations because of drift. The overall effect will be that diversity is invariant.

To test whether this was happening in mtDNA, the authors measured the relative rates of non-synonymous and synonymous mutations across taxa. There was a higher rate of non-synonymous change in species with larger populations, but no increase in diversity, which is consistent with positive selection.

It is not known why positive selection is stronger for mtDNA than nuclear DNA. It might be because the mitochondrial genome is particularly gene-rich. Whatever the reason, these results could change the way population diversity is measured and how conservation biologists identify important lineages.

Patrick Goymer

ORIGINAL RESEARCH PAPER Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572 (2006)
FURTHER READING Chen, X. J. & Butow, R. A. The organization and inheritance of the mitochondrial genome. *Nature Rev. Genet.* **6**, 815–825 (2005)

⊐ MOUSE GENETICS

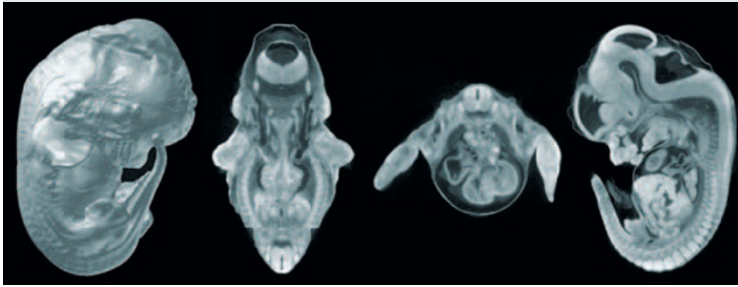# No Mickey-Mouse phenomics

Image courtesy of Charles Keller, University of Texas Health Sciences Center at San Antonio, and John T. Johnson, Scientific Computing and Imaging Institute, University of Utah

The current emphasis on large-scale biology means that we require contemporary technologies to be high throughput or at least scalable. For genomics and genotyping, this goal is largely being achieved, but phenotyping analysis has notably lagged behind. So the new method for rapid, high-resolution phenotypic assessment of mouse embryos developed by John Johnson and Mark Hansen and colleagues should meet with much interest.

Laborious, histology-based phenotypic analysis has been improved by the introduction of magnetic resonance microscopy (MRM), but this approach requires highly specialized equipment and is therefore prohibitively expensive to most researchers. Johnson and Hansen *et al.* now report a new method that is based on X-ray microscopic computed tomography (microCT). In this approach, a virtual histological representation of an embryo, which had been stained in 1% osmium tetroxide solution, is obtained at up to 8-μm resolution.

The resolution of microCT exceeds that of routine MRM scans. Importantly, scanning takes less time, is less expensive and several embryos can be scanned at once. To maximize the high-throughput aspect of this approach, the authors advocate scans at 25–27 μm (which is comparable to most MRM methods). With a small animal CT scanner, these only take 2 hours for up to 120 embryos. A higher resolution scan can always be done on a subset of the sample.

The authors tested their approach by characterizing *Pax3: Fkhr* transgenic mouse embryos, which are known to have complex neural tube defects. To this end they used a scanner that is similar to commercially available small animal *in vivo* scanners, which are frequently operated in regional core facilities. The new approach provided resolution that was as good as conventional histology of paraffin-embedded samples; moreover, it provided a clearer overall picture of the defects in three dimensions.

Among the drawbacks of this technique is its dependence on osmium tetroxide staining — because permeability of embryos decreases as the epidermis develops, the approach is most suited to embryos in mid-gestation. Embryos that are older than embryonic day 13.5 need to be treated with protease or have the epidermis removed manually.

As the authors point out, high-throughput phenotyping will also require automated or semi-automated computer-based analysis of the data. Given the recent advances in microimaging and high-resolution microCT instrumentation, and the use of existing and new electron-dense stains, who knows, true phenomics might be coming soon to a laboratory near you.

Magdalena Skipper

ORIGINAL RESEARCH PAPER Johnson, J. T & Hansen, M. S. *et al.* Virtual histology of transgenic mouse embryos for high-throughput phenotyping. *PLoS Genet.* **2**, e61 (2006)

IN BRIEF

⊐ EVOLUTION

Parallel evolution of the genetic code in arthropod mitochondrial genomes.

Abascal, F. *et al. PLoS Biol.* **4**, e127 (2006)

To investigate the extent to which the genetic code varies, the authors compared the codon usage of mitochondria in 626 animal species using automated computational methods. Arthropods stood out as they have evolved non-standard codons: AGG is translated as either Lys or Ser (instead of Ser only). The frequent reoccurrence of this codon reassignment, as well as the pattern of genetic change, indicates that codons are reassigned more easily than was previously believed, and suggests that codon usage in other lineages might be similarly labile.

⊐ TRANSCRIPTION

The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model.

Luo, W. *et al. Genes Dev.* **20**, 954–965 (2006)

Two models have been proposed for how RNA polymerase II terminates transcription. The allosteric model posits that the polymerase falls off the template owing to conformational changes that are triggered by the poly(A) sequence. According to the 'torpedo' model exonucleases digest the growing RNA strand until they hit the elongation complex. This paper shows that an integrated model is more plausible: the enzymatic function of the Rat1 and Xrn1 exonucleases is not sufficient to displace the polymerase, which requires additional factors, including an allosteric effector, that are recruited by Rat1.

⊐ SYNTHETIC BIOLOGY

Emergent properties of reduced-genome *Escherichia coli*.

Pósfai, G. *et al. Science* 27 April 2006 (doi:10.1126/science.1126439)

The authors used a recombination-based approach to delete up to 15% of the *Escherichia coli* K-12 genome. Removal of mainly large islands, insertion-sequence-containing islands and individual genes containing insertion-sequence elements led to some unexpected properties. Without compromising growth dynamics, the authors created a strain that showed increased electroporation efficiency and enhanced propagation and stability of plasmids, which are used as delivery vehicles in vaccine and gene therapy research.

⊐ PARASITE GENETICS

Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region.

Riehle, M. M. & Markianos, K. *et al. Science* **312**, 577–579, 2006

A survey of genetic variation in a West African population of the mosquito *Anopheles gambiae* has identified a cluster of strong resistance loci on chromosome 2L, which forms a *Plasmodium*-resistance island. Among the candidate genes within this 15-Mb region, the authors identified *Anopheles Plasmodium*-responsive leucine-rich repeat 1 (*APL1*), the homologues of which are involved in natural resistance mechanisms in plants and mammals. The role of *APL1* in *Plasmodium* resistance was confirmed by RNAi.

➡ **STEM CELLS**

Poised for action

The pluripotent nature of stem cells relies on two seemingly opposing requirements — to keep differentiation-specific genes switched off while retaining the ability to switch them on quickly when needed. Four new papers reveal how unique chromatin properties of mammalian stem cells allow this delicate balance to be achieved.

Azuara and colleagues showed that many genes in mouse embryonic stem (ES) cells — including several that are involved in differentiation — are characterized by an unusual combination of epigenetic modifications. This includes acetylation at histone H3 lysine 9 (H3K9) and methylation at H3K4, which are marks of active chromatin, and trimethylation at H3K27, which is typical of silent chromatin. In most non-stem cells genes have either active or repressive marks, but not both.

An unusual epigenetic profile was also identified by Bernstein and colleagues for mouse ES cells. They identified a bivalent chromatin structure — containing both methylated H3K4 and trimethylated H3K27 — at genes that encode developmentally important transcription factors.

Both groups showed that loss of these distinctive combinations of epigenetic marks is correlated with differentiation. The authors propose that the presence of both active and repressive marks allows differentiation-specific genes in ES cells to be repressed but also to be primed for activation when the right signals are received.

How are the unique epigenetic profiles of stem cells specified? The involvement of Polycomb group (PcG) proteins was suggested by Azuara and colleagues. They showed that ES cells that are deficient for embryonic ectoderm deficient (EED), a PcG-complex

> “
> …embryonic stem cells … are characterized by an unusual combination of epigenetic modifications.
> ”

component that is required for H3K27 methylation, express differentiation-specific genes that are otherwise repressed in these cells. This finding is supported in two other ES-cell studies from Boyer and colleagues and Lee and colleagues. Both studies identified large numbers of PcG binding sites across the ES-cell genome that correspond to genes with functions in development and differentiation. These genes were also found to be marked with trimethylated H3K27 and to be transcriptionally silent in ES cells.

Similar to Azuara and colleagues, Boyer et al. showed that EED deficiency results in the activation of these otherwise repressed genes, and both Lee et al. and Boyer et al. showed that PcG target genes are activated during mouse ES-cell differentiation. The authors suggest a dynamic role for PcG proteins in which they maintain a repressive chromatin state before differentiation-inducing signals are received, but are displaced to allow gene expression at the appropriate time. Similar conclusions were drawn from four other recent

---

➡ **DEVELOPMENT**

A developing role for Polycomb proteins
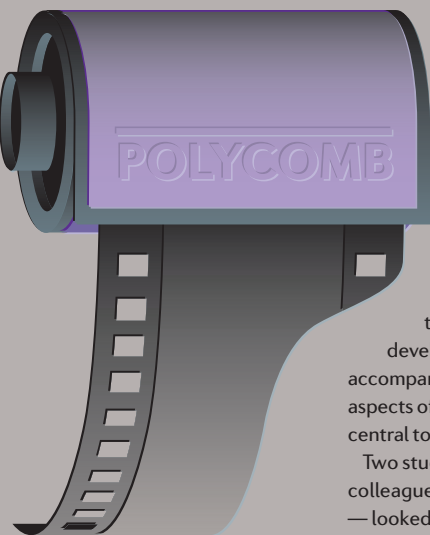
Polycomb group (PcG) proteins have essential roles in differentiation, allowing cells to memorize the transcriptional responses of genes to transient developmental signals. A slew of recent papers now reveal many new targets for PcG proteins and add to the emerging evidence that they have dynamic roles during development. As described in an accompanying Research Highlight, these aspects of PcG function also turn out to be central to the properties of stem cells.

Two studies — from Tolhuis and colleagues and Bracken and colleagues — looked at the role of PcG proteins in cell lines (*Drosophila melanogaster* and human, respectively) that are no longer pluripotent, but are not terminally differentiated. A third study, from

Nègre et al., examined PcG functions during fly development. All three used genomic profiling methods to generate unbiased pictures of how PcG protein binding is distributed, either across the whole genome or in large chromosomal regions.

In each case, a wealth of new target genes for PcG proteins was revealed. As might be expected, a large proportion of these encode transcriptional regulators or signalling proteins with developmental functions, in line with known PcG targets. However, intriguing new roles of PcG proteins were also highlighted. For example, Tolhuis and colleagues identified several genes with roles in steroid hormone biosynthesis, suggesting a potential function for PcGs in fly metamorphosis.

PcG proteins maintain transcriptional silence in a process that is thought to involve trimethylation of histone H3 at

Lys27 (H3-K27). Consistent with this, Tolhuis et al. and Bracken et al. found that PcG binding generally correlated with this epigenetic mark, and with low levels or absence of transcription.

However, PcG proteins seem to have a more dynamic role than simply binding and remaining at their target sites. Nègre et al. found that the profile of PcG binding changes as *Drosophila* development progresses, and that although a particular gene might be a PcG target, this is not necessarily the case in all tissues or developmental stages. In line with this, Bracken et al. showed that the displacement of PcG proteins is important for differentiation: the increased expression of target genes in response to a differentiation-inducing stimulus corresponded with a decrease in PcG binding.

Two further studies, which are discussed in the accompanying Research Highlight, reveal that this dynamic regulation is also important in embryonic stem cells. They describe how PcG proteins keep differentiation-specific genes in a state that is

papers — which are discussed in a related Research Highlight — that examined the roles of PcG proteins in other non-pluripotent cell types, indicating that PcG complexes function in a similar way at various developmental stages.

Excitingly, Lee and colleagues also found that the transcription factors OCT4, SOX2 and NANOG — which are already known to have important roles in pluripotency and self-renewal — are also present at many PcG binding sites in ES cells. Working out how these proteins contribute to the unique chromatin properties of ES cells will be an important step in tying together key aspects of stem-cell biology.

Louisa Flintoft

ORIGINAL RESEARCH PAPERS Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* March 29 2006 (doi: 10.1038/ncb1403) | Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006) | Lee, T. I. *et al.* Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006) | Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* April 19 2006 (doi:10.1038/nature04733)

transcriptionally silent and yet poised for activation when expression is needed — seemingly a common theme in the role of PcG proteins throughout development.

These studies hint at many roles of PcG proteins in development and differentiation that have yet to be investigated in detail. Exploring how these proteins interact with other chromatin components, and how they are regulated by developmental cues, will be key to our understanding of how transcriptional states are memorized and modified as cells progress towards their ultimate fates.

Louisa Flintoft

ORIGINAL RESEARCH PAPERS
Tolhuis, B. *et al.* Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster. Nature Genet.* 20 April 2006 (doi:10.1038/ng1792) | Nègre, N. *et al.* Chromosomal distribution of PcG proteins during *Drosophila* development. *PLoS Biol.* **4**, e170 (2006) | Bracken, A. P. *et al.* Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 17 April 2006 (doi:10.1101/gad.381706)

⮕ EVOLUTION

A tinkerer's tales

The general framework in which evolution acts — mutation followed by selection — is understood and accepted by (almost) everyone, but what concerns practising evolutionary biologists most is what goes on at the nitty-gritty level. How does a trait get from A to B? How many paths can evolution take? And are they all equally probable? Universal answers to these questions do not exist but three papers have each used a different approach to address one of these fundamental problems.

The three articles examine the routes that evolution takes in creating an adaptive function. One study, by Prud'homme, Gompel and colleagues, concentrated on the black spot that exists on the male wing of some *Drosophila* species. The wing spot has evolved independently in two lineages, where it is used to woo females, and in both cases the evolution of the pattern involves the *yellow* pigmentation gene. What is most interesting is that the *cis*-regulatory elements that have been used in the two cases are distinct, indicating that evolution can use different mechanisms to reach a convergent, adaptive endpoint.

The repeated use of the *yellow* gene in different lineages raises another theme in evolution — constraint. That constraint exists was the strong conclusion to emerge from the work of Weinreich and colleagues when they addressed a related issue: how many paths can a protein take towards a fitter state? Their choice was the evolution of bacterial β-lactamase, which can evolve a 100,000-fold increase in antibiotic resistance by acquiring just five point mutations. Of the 120 hypothetical trajectories to drug-resistant alleles, 102 are inaccessible to evolution. This prediction was based on the probability of fixation of mutant combinations; in fact, the situation is more extreme than this, because of the 18 plausible combinations as few as 2 are probable, indicating that the path to adaptive protein evolution is largely predictable.

Gradual changes and stepwise optimization are fine in principle, but how, in practice, do you end up with the level of precision that is seen in complex molecular interactions? Bridgham and colleagues address this question with respect to the distinct specificity of the mineral corticoid hormone receptor for its ligand, aldosterone. This and similar relationships present a conundrum: the receptor and ligand match each other perfectly, so how could the evolution of one partner be explained unless the other is already present? Modern gene sequences say little about how the system evolved so the authors used phylogenetic information to reconstruct ('resurrect') the sequence of the ancestral receptor. Surprisingly, the receptor already had some affinity for aldesterone — not because aldesterone was present but because of the existence of a structurally related molecule. This is a case of an old molecule being exploited to acquire a new function — a switch that, the authors discovered, was only two amino-acid mutations away. Whereas Weinreich's study looked at all the evolutionary paths that could have occurred, this work examined those mutations that actually took place. The theory of evolution is therefore safe for now, and has been tested in precisely the way Darwin had suggested.

These three studies allow us to inch a little closer to discerning the subtler mechanics of evolution by suggesting that evolution does not generate complexity by creating new parts, but by tinkering with existing components (be they *cis*-regulatory sequences or proteins) and increasing the number of ways in which they interact.

Tanita Casci

ORIGINAL RESEARCH PAPERS Bridgham, J. T. *et al.* Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006) | Weinreich, D. M. *et al.* Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006) | Prud'homme, B. & Gompel, N. *et al.* Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006)
FURTHER READING Thornton, J. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Rev. Genet.* **5**, 366–375 (2004) | DePristo, M. A. *et al.* Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Rev. Genet.* **6**, 678–687 (2005)

RESEARCH HIGHLIGHTS
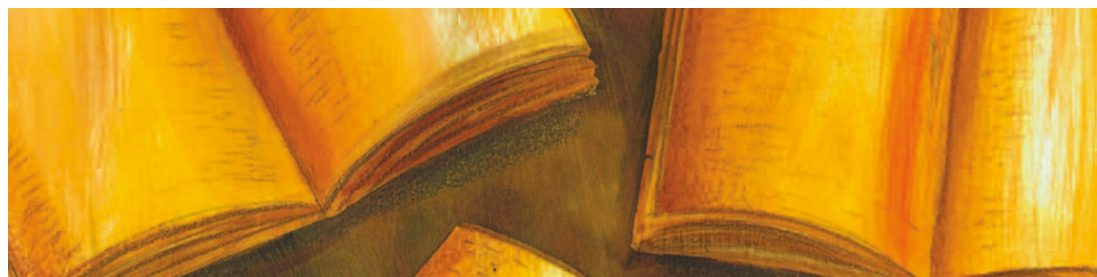
IN BRIEF

DRUG DISCOVERY

A small-molecule screen in *C. elegans* yields a new calcium channel antagonist.

Kwok, T. C. Y. & Ricker, N. *et al. Nature* **441**, 91–95 (2006)

Caenorhabditis elegans can be used to rapidly identify new small-molecule inhibitors and their targets, both of which are powerful tools for biological analysis and drug discovery. In a screen of 14,100 small molecules, 308 compounds induced a range of phenotypes. Nemadipine A, which is similar to a class of anti-hypertension drugs that antagonize a particular type of calcium channel, caused abnormal morphology and egg-laying defects. A suppressor screen identified *egl-19*, which encodes the correct type of calcium channel, as a nemadipine A target. Nemadipine A was then used to reveal calcium-channel redundancy in the egg-laying circuitry.

EPIGENETICS

Circadian regulator CLOCK is a histone acetyltransferase.

Doi, M. *et al. Cell* **125**, 497–508 (2006)

These authors show that CLOCK — a key component of the circadian pacemaker — has intrinsic histone acetyltransferase (HAT) activity. CLOCK proteins in which an acetyl-coenzyme A binding motif (which is similar at the sequence level to other well-characterized HAT proteins) has been mutated have reduced HAT activity. Their overexpression cannot rescue circadian gene rythmicity in cells in which the endogenous *Clock* gene has been mutated, demonstrating the importance of chromatin remodelling in circadian gene expression.

NEUROGENETICS

The molecular diversity of *Dscam* is functionally required for neuronal wiring specificity in *Drosophila*.

Chen, B. E. *et al. Cell* **125**, 607–620 (2006)

Dscam, the gene that encodes the Down syndrome cell-adhesion molecule, can potentially produce up to 38,016 different protein isoforms by alternative splicing. The authors showed that *Dscam* is essential for proper axonal branching in the fly, that alleles that could produce 22,176 isoforms did not fully rescue the phenotype, and that expression of individual isoforms has distinct partial-rescue phenotypic effects. This implies that much of the potential isoform diversity that is encoded by this gene is necessary for the differentiation of neurons.

EPIGENETICS

Intra- and inter-individual epigenetic variation in human germ cells.

Flanagan, J. *et al. Am. J. Hum. Genet.* 29 March 2006 [Epub ahead of print]

This study examines the amount of epigenetic variation in the human male germ line. By analysing cytosine methylation in specific genes, and a microarray analysis of CpG islands, the authors show that methylation patterns vary considerably between individuals and between sperm within an individual. Promoter CpG islands and pericentromeric satellites had the most variation, and some variation was age- and allele-dependent. The inter-individual epigenetic variation was much greater than the genetic variation, although it is unclear how much of it is inherited.

TECHNOLOGY

A targeted viewing

The killing part is easy; the real challenge — when it comes to dealing with cancer cells — is to find the cells to be removed, or indeed treated, and to be able to see how well you are doing. Researchers now have the tool for this purpose; a hybrid virus that combines the efficiency of adeno-associated virus (AAV) and the adaptability of a bacteriophage that can target specific cancer cells and deliver to them imaging or anti-cancer agents.

Eukaryotic viruses enter cells efficiently and specifically. This specificity is also a drawback, as their strong preference for particular molecules limits their versatility. Bacterial viruses, on the other hand, are not very accomplished at entering mammalian cells — but they are indifferent to cell type, which allows them to be engineered to target eukaryotic cell-surface markers. Renata Pasqualini, Wadih Arap and colleagues put two and two together and built a hybrid virus that delivers DNA as efficiently as a mammalian virus but, like a phage virus, can be targeted to a desired protein. The aim was to create a unique tool that would combine tumour targeting and molecular-genetic imaging.

The chimeric molecule consisted of the *cis*-regions of mammalian AAV and the ssDNA of the fd-tet phage, which is derived from M13. The phage portion of the construct was engineered to encode a peptide that would allow the hybrid DNA to bind to αv-integrins, which are cell-surface markers of tumour blood vessels, and to be internalized by such cells. Indeed, when the construct was injected into nude mice it was rapidly taken up specifically by the vasculature of tumour

DISEASE GENETICS

The benefits of networking

How do you distinguish a disease-susceptibility gene from the hundreds of other genes in a locus that has been identified by positional cloning? Cisca Wijmenga and colleagues have turned to network biology for the answer.

When several susceptibility loci have been identified for a particular human disease, the authors reasoned that the true disease genes from each locus are likely to be involved in the same molecular pathway. Therefore, they used publicly available interaction data to create a network for identifying such genes. There were three steps in this process: building a robust network; showing that genes for the same disease are indeed near neighbours in the network; and testing the predictive power of the network.

The authors constructed their network from databases of known interactions, and also Gene Ontology, microarray and yeast two-hybrid data. They tested their network on interaction data that had recently become available and found that it successfully predicted the interactions. Next, they showed that the genes from any one disease were closer to each other in the network than would be expected by chance.

To assess the predictive power of their network, the authors took about a hundred genes that are located around several disease genes and tested whether it could successfully pick out the correct disease genes from their chromosomal neighbours. The network significantly improved the chances of finding the correct genes, although the results were better for diseases with more loci, and also for Mendelian rather than complex diseases.

The gene network and the software for using it are freely available

xenografts, as expected, and hybrid viruses that carried *GFP* also stained vascular cells specifically and stably. A qualitatively similar effect was seen in genetic mouse models of cancer. Similarly, delivery of the 'suicide' gene *HSVtk* using this tool reduced tumour size considerably, showing that anti-cancer agents could be delivered to therapeutic effect *in vivo*.

The success of the hybrid virus highlights its potential usefulness — and that of eventual derivatives — in both a clinical and research setting. Tumours could be found, highlighted and then monitored non-invasively and cost-effectively while being targeted with specific genes. The *HSVtk* marker is picked up by a routine PET (positron emission tomography) body scanner and so, from a practical point of view, this tool is potentially ready for clinical use.

Tanita Casci

on the Prioritizer web site. This advance could make positional cloning faster, cheaper and more accurate, and the continuing growth of interaction data will further improve the predictive power of the network.

Patrick Goymer

TRANSCRIPTOMICS

Charting the functional landscape of the mammalian genome

An important goal of functional genomics is to understand how the information encoded in an organism's genome is retrieved. A range of approaches have been used to characterize functional elements in the genome that control transcription; for the human genome this is done perhaps most famously in an international collaborative project called ENCODE — Encyclopedia of DNA elements. An important study that contributes to this project has recently been published in which human and mouse transcription start sites (TSSs) are tagged and characterized on a genome-wide scale. The study provides new insights into basic promoter features, as well as the evolutionary conservation and dynamic regulation of mammalian promoters.

Wanting to refine the functional landscape of the mammalian genome, the authors turned to CAGE (cap analysis of gene expression), a technique for high-throughput analysis of TSSs and study of promoter usage. Approximately 650,000 and 730,000 TSSs were identified in the human and mouse genomes, respectively, by mapping onto unique genomic regions 20- or 21-nt long CAGE tags, which were derived from sequences (from hundreds of CAGE libraries derived from all major tissues) that lie in the proximity of the cap site.

TSSs that were defined by tag clusters of at least two tags were followed up in more detail. In fact, the distribution of tags within a cluster allowed the authors to categorize TSSs into three classes, which are conserved between mouse and human: a single peak (SP) class, indicative of a well-defined TSS; a broad shape (BR) class, indicative of multiple, weakly defined TSSs; and a bimodal/multiple (MU) class, indicative of several well-defined TSSs in one cluster. The SP class was mainly made up of TATA-box promoters, whereas the BR class of regions associated with CpG islands. The BR class predominates in mammalian genomes, whereas TATA-box promoters, mainly associated with tissue-specific and conserved gene expression, are in the minority.

The study identified TSSs in unexpected places. In some genes all of the exons showed promoter activity, whereas in other genes expressed at similar levels this activity was absent. The authors also identified a new class of promoters in 3′ UTRs. Although the function of these promoter types remains to be clarified, the authors speculate that they might have a role in enhancing RNA processing, including splicing and transcription itself.

The study provided valuable evolutionary insights. Notably, the initiator sequences — located at position –1 to +1 of TSSs — are subject to frequent changes in mammals. Moreover, pyrimidine–purine dinucleotides, which are overrepresented at the –1 to +1 sites, seem to contribute to the precise location of TSSs of the BR class. Overall, the CpG island-associated promoters seem to evolve more rapidly than the TATA-box promoters. The epigenetic control of CpG island-associated promoters and the fact that some contain bidirectional TSSs, which might regulate a locus's expression, might be important facilitators of adaptive evolution in mammals.

Having provided a wealth of information about the functional landscape of mouse and human genomes, the authors feel confident that meaningful models of transcriptional regulatory networks can be constructed on the basis of the information about core promoter sequences. And, because "technologies like CAGE are scaleable to whole organisms, these approaches pave the way for 'systematic' systems biology."

Magdalena Skipper

Ethics Watch

IMPLICATIONS OF COPY-NUMBER VARIATION IN THE HUMAN GENOME: A TIME FOR QUESTIONS

Just as bioethicists came to understand single nucleotide polymorphisms (SNPs) they are now faced with comprehending new discoveries of thousands of submicroscopic quantitative structural changes of chromosomes — including insertions, deletions and duplications — in the human genome, collectively called copy-number variants (CNVs) or polymorphisms (CNPs)[1]. Progress in completing genome-wide maps and databases of CNVs (for example, the Database of Genomic Variants) adds a new dimension to the study of genome variation. This, in turn, will affect our concept of 'normality' — it is now routinely observed that one of a pair of homologous chromosomes can be up to a million nucleotides and 20 genes shorter than the other[2] — with potentially profound ethical, legal and social implications (ELSI).

The discovery of CNVs as a universal feature of genomes coincides with growing interest in the influence of genomic variation on human characteristics, disease and evolution. Several ongoing or planned genotyping studies, such as the Human Genome Organisation Pacific Pan-Asian SNP Initiative and, of course, the International HapMap Project, hold interest for health-care providers, policy makers, bioethicists and the general public. Bioethicists are inextricably involved in examining social implications and developing ELSI guidelines. However, we will have to re-examine our previous conclusions and recommendations, because when earlier genomic mapping studies were designed, SNPs were the primary currency of genomic variation and little was known about CNVs and their implications.

Now, large-scale, genome-wide association or case–control studies, including those of the Wellcome Trust Case Control Consortium (15,000 samples), Broad Institute (12,000 samples), the Korean National Institute of Health (20,000 samples) and the public/private Genetic Association Information Network (GAIN), will undoubtedly incorporate CNV analyses into their designs. The Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER) project is perhaps one step ahead of others in its consideration of consent and database issues that are specific to CNVs, and might offer a precedent for future studies.

Generation of population-wide CNV catalogues means that databases that have relied on SNPs might have to be re-evaluated. Moreover, future investigations will benefit from reliable, complementary and harmonized technologies that account for several forms of genomic variation simultaneously. This means that ELSI experts should proceed with caution and perhaps wait for more complete scientific knowledge before they consider ELSI surrounding genome variation and its applications, particularly for such potentially controversial issues as forensics and race[3].

The description of CNVs has already enhanced our fundamental understanding of human evolution, disease predisposition and the body's response to drugs and other environmental influences[1]. But despite several news reports and scientific editorials, the topic has not yet captured the public imagination,

> "…how might new genomic data influence conventional classifications of 'common', 'normal', 'healthy' or 'diseased'?

even though accumulating data indicate that CNV content might exceed that of SNPs within humans[4], and between humans and chimpanzees.

There are several issues that could arise for ELSI experts. In a clinical research setting, how will the management and counselling of a patient and his/her family unfold with, and without, taking CNVs into account? Will ethical issues arising from analysis of CNVs simply mirror past issues encountered when using cytogenetic approaches, or will CNVs give rise to new and unforeseen complexities in genetic counselling? Should archived specimens be re-evaluated? Could patients sue if a past medical genetic diagnosis required revision in light of new knowledge afforded by CNVs?

We have argued that pharmacogenetics that is based on geographical ancestries might be harnessed to improve global health and energize the pharmaceutical industry in developing countries[5]. Can the latter benefit from harnessing knowledge of CNVs in pharmacogenomic testing and drug design? Furthermore, will industry's internalization and use of CNVs predict how such knowledge might be more widely applied?

As we move towards the potential for personalized genome fingerprints and sequences, how might new genomic data influence conventional classifications of 'common', 'normal', 'healthy' or 'diseased'? Indeed, given our partial understanding of genomic variation, how should scientists and ELSI experts advise governments on the roles of genomics in health care?

At present, ELSI studies on CNVs are in their infancy. So we should begin by developing a conceptual framework that places the new discoveries in the wider context of contemporary discourse on genomic variation studies and their biological, health and societal implications. We need to develop a taxonomy of the research issues to be addressed jointly by scientists, social scientists, clinicians, anthropologists, philosophers, historians, lawyers and other relevant disciplines.

This is the right time for scientists and ELSI experts to work together closely. Being able to address such questions sooner rather than later will better ensure realization of the potential of scientific discoveries related to genomic CNVs.

Abdallah S. Daar*, Stephen W. Scherer‡ and Robert A. Hegele§

*Departments of Public Health Sciences and Surgery, and McLaughlin Centre for Molecular Medicine, University of Toronto, Canada. ‡Program in Genetics and Genomic Biology, and The Centre for Applied Genomics, Hospital for Sick Children, Toronto; Department of Molecular and Medical Genetics, University of Toronto. §Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, and Schulich School of Medicine and Dentistry, University of Western Ontario, Canada.

REFERENCES [1]Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7**, 85–97 (2006) | [2]Check, E. Human genome: patchwork people. *Nature* **437**, 1084–1086 (2005) | [3]Royal, C. D. M. & Dunston, G. M. Changing the paradigm from 'race' to human genome variation. *Nature Genet.* **36**, S5–S7 (2004) | [4]Lee, C. Vive la difference! *Nature Genet.* **37**, 660–661 (2005) | [5]Daar, A. S. & Singer, P. A. Pharmacogenetics and geographical ancestry: implications for drug development and global health. *Nature Rev. Genet.* **6**, 241–246 (2005)