

because it seems to be genetically indistinguishable from other variants embedded in this haplotype block. The advent of whole-genome association studies using the haplotype map¹¹ will probably implicate a substantial number of new disease associations with indirect markers that are in linkage disequilibrium with rSNPs.

Associations in search of a cause

Providing that Pol II (or another DNA-binding protein with similar properties) is equally useful for testing any human gene, haploChIP has the potential to allow the testing of each gene contained in a risk haplotype block for differential transcriptional activity. HaploChIP requires that suitable cells or tissues that express the gene are available and that the regulatory variant affecting the test gene is heterozygous. Another surrogate approach to *in vivo* transcriptional activity is allele-specific RT-PCR^{12,13}. This is probably easier to carry out but has an extra requirement that the RT-PCR product contain polymorphisms, which is not possible for genes (such as *TNF*) that lack exonic SNPs. By

comparing both gene transcripts in the same sample, both methods decrease the effects of environmental and other confounding factors. They also have advantages over *in vitro* techniques, such as transient transfection assays with allele-specific promoter constructs, as these studies are executed outside of their normal chromosome environment.

The haploChIP assay is useful as a surrogate marker of allelic variation in gene transcription, although it does not directly identify the *cis*-acting polymorphism or mechanism that is responsible for this variation. This is not surprising, as there are many SNPs on the haplotype associated with higher *LTA* transcriptional activity. The task is also difficult because of the extensive distances at which transcriptional control elements can be found¹⁴. Should human geneticists care about finding the true regulatory SNP when the risk gene is identified and there already are many highly correlated SNPs that tag the disease haplotype? Yes! The full knowledge of the risk variant is important, as it helps refine the knowledge of gene transcription and points

to a target for therapy. Thus, the genetic community continues to require more tools to efficiently pinpoint risk variants that affect gene expression. There is no doubt that many more inventive approaches will be developed for this. Ultimately, the manipulation of regulatory mutations that affect expression levels should be easier than repairing or modulating the effects of an abnormal protein. □

1. Cargill, M. *et al. Nat. Genet.* **22**, 231–238 (1999).
2. Sachidanandam, R. *et al. Nature* **409**, 928–933 (2001).
3. Brem, R.B., Yvert, G., Clinton, R. & Kruglyak, L. *Science* **296**, 752–755 (2002).
4. Cheung, V. *et al. Nat. Genet.* **33**, 422–425 (2003).
5. Wasserman, W.W., Palumbo, M., Thompson, W., Fickett, J.W. & Lawrence, C.E. *Nat. Genet.* **26**, 225–228 (2000).
6. Rockman, M.V. & Wray, G. *Mol. Biol. Evol.* **19**, 1991–2004 (2002).
7. Knight, J.C., Keating, B.J., Rockett, K.A. & Kwiatkowski, D.P. *Nat. Genet.* **33**, 469–475 (2003).
8. Ozaki, K. *et al. Nat. Genet.* **32**, 650–654 (2002).
9. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. *Nat. Genet.* **29**, 229–232 (2001).
10. Rioux, J.D. *et al. Nat. Genet.* **29**, 223–228 (2001).
11. Gabriel, S.B. *et al. Science* **296**, 2225–2229 (2002).
12. Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B. & Kinzler, K.W. *Science* **297**, 1143 (2002).
13. Cowles, C.R., Hirschhorn, J.N., Altshuler, D. & Lander, E.S. *Nat. Genet.* **32**, 432–437 (2002).
14. Hardison, R.C., Oeltjen, J. & Miller, W. *Genome Res.* **7**, 959–966 (1997).

A clinician's plea

Judith G. Hall

Department of Pediatrics and Medical Genetics, University of British Columbia, Vancouver, British Columbia V6H 3V4, Canada. e-mail: jhall@cw.bc.ca

For the detection of human gene mutations to have practical application, clear clinical descriptions of the affected individuals (as well as those clinically affected in whom mutations are not found) should be part of the publication.

Over the last ten years, there has been an exponential increase in the number of papers describing mutations in genes responsible for human disease. Clinicians are thrilled that this new information may translate into meaningful changes in health care for affected individuals. We carefully read such papers for clinical details to help determine which individuals and which families might benefit from molecular evaluation or testing. Sometimes these details are there, but more often they are not. Thus, clinical geneticists are frustrated by an inability to determine to which condition or conditions the reported gene mutation actually refers.

The project to sequence the human genome, nearing completion this year, has been a remarkable effort. It was powered

in part by the promise that insights into human health and disease would result. The large number of excellent articles describing genes mutated in human diseases in this issue of *Nature Genetics*^{1–8} reflects the fruits (and hard work) that are to be celebrated. At the same time, however, they show how little we know about these genes, how much more work there is to be done and how urgently accurate and clear clinical descriptions are needed.

The clinical approach

A clinician can't help but wonder what was different about the 11 individuals with Ondine's curse (congenital central hypoventilation syndrome) in whom no mutation in *PHOX2B* was found, as described by Jeanne Amiel and colleagues¹

on page 459, or what additional features might have been present in the 31 individuals with anophthalmia who had no mutations in *SOX2*, as outlined by Judy Fantes and colleagues³ on page 461. Did they have other structural or functional abnormalities? What was different about the pregnancies of those individuals? Has unusual behavior developed as they age? Was their mother's health or nutrition unusual before conception or during the pregnancy?

Often we read that the 'found' gene is actually one that was previously unknown. The gene structure may allow speculation about function, and there may be related genes that suggest clues as to pathway or pathogenesis, but the actual pathogenic mechanism must be determined through the hard work of defining time-specific, tis-

sue-specific and species-specific expression and through building animal models of disease. The isoforms, protein folding, control mechanisms and position along a biochemical pathway await elucidation.

Important clues will probably come from full clinical descriptions of which tissues are involved, when the onset of disease occurs and what complications develop over time. Recognizing the signs and symptoms that identify exceptions should give hints about alternative pathways. For instance, as indicated by Catherine Dodé and colleagues² on page 463, among the individuals with Kallmann syndrome who have mutations in *FGFR1* (but not among those with other mutations), marked craniofacial structural anomalies are often seen. Those individuals with Smith–Magenis syndrome who have heart and renal defects are likely to have chromosomal deletions that include *RAI1*, which is described by Rebecca Slager and colleagues⁸ on page 466.

Traditionally, clinicians approach their work quite differently from basic scientists. Medical school provides them with a broad background in anatomy, cell biology, physiology, pathology and therapeutics. When an affected individual or family presents with a problem, the clinician collects historical information, carries out a physical examination and gathers the results of laboratory and imaging studies. Then, because there may be hundreds of conditions to consider, the clinician usually consults the literature to arrive at a diagnosis that will allow appropriate counseling and therapy. By contrast, owing to training and background, the scientist selects a problem, formulates a hypothesis and then does experiments, collecting data to attempt to prove the hypothesis correct or incorrect. Although the two approaches have begun to overlap,

because the clinician is dealing with human beings, there are many constraints and diagnosis and therapy are the goals.

The clinician needs clinical details and clues from the affected individual, the family and a literature review to determine which tests are appropriate to make a diagnosis. Gender, ethnic background, age and natural history help in that process. For example, for individuals with small heads (microcephaly), when should Seckel syndrome (see the paper by Mark O’Driscoll and colleagues⁵ on page 497) be considered? Historically, individuals with Seckel syndrome have been described as having intrauterine growth retardation (small at birth for their gestational age) and a head size that is much smaller than would be expected. There are many conditions that share these two features⁹. Thus, the description of additional clinical features, such as beaked nose, large-appearing eyes and sociable personality, may help the clinician to determine who should be tested for mutations in *ATR*, which O’Driscoll *et al.*⁵ have found to be associated with Seckel syndrome. These ‘newfangled’ DNA-based tests are expensive (if you can even get them done). Insurance companies don’t like diagnostic ‘fishing expeditions’. The clinician tries to focus diagnostic testing and do only the most appropriate test(s) in a rational and justifiable manner. This requires knowing the indications and the clinical features that might implicate a mutation in a specific gene. If the clinical features reported have been unclear, the testing of appropriate individuals is unlikely to occur.

What information is needed?

“These are meant to be molecular papers,” you say. “Clinical information should be in a clinical journal.” “It takes too much space.” “We’re not interested in those fuzzy

details.” Without clear clinical descriptions, however, the relevance for human disease, diagnosis and therapy may be lost—and, after all, that is part of the purpose of such work. The clinician usually has a systematized way of collecting information (see box) and storing it for future reference (medical records). Some papers that report mutations in a human gene include tables giving phenotypic features^{2,3,7,8}. But too few papers include the clinical features of those individuals in whom no mutation was found. But what is or was different about those individuals? Occasionally, DNA samples are obtained from cell repositories that are notorious for lacking good, verifiable clinical information—yet another problem! As genomic research moves towards proteomics, and proteomics moves towards ‘proteotype–phenotype’ correlations, clarity concerning the clinical features will become even more important.

Unfortunately, clinical descriptions are rarely static. Just as the individual changes with age, so do the clinical features. The study of the natural history of a genetic disorder is reflected against the normal maturational processes (physiologic changes that occur with aging) of human beings. How the gene product fits into various biochemical pathways will, in the long run, lead to the signs and symptoms of the disease as normal processes go awry. Thus, recording information about the natural history requires one to cast a wide net and often results in unexpected correlations. But that is the real biology, the real secret of gene function and the future challenge.

Now or never

All research on human gene mutations requires collaboration with clinicians. The clinicians are responsible for making the diagnosis, getting permission, obtaining

Clinical information that should be available on individuals reported to have disease-related gene mutations.

Although the type of information to be collected obviously will vary from disorder to disorder, the following represents a minimum standard, ideally to be defined on any person in whom a mutation is found, as well as those in whom no mutation is found. It would also be useful to record this same information about individuals who are found not to have an identifiable mutation, so that clinical distinctions can be made and appropriate testing can be done.

- A full, three-generation family history with particular emphasis on the signs and symptoms of the disorder and features of pathogenetic relevance.
 - Natural history including other illnesses, pregnancy and birth history and a system-by-system review, including behavior, with changes over time in the areas of interest, noting the ages at which signs or symptoms develop.
 - Information on height, weight, head circumference, blood pressure, pulse and respirations. If the syndrome includes abnormalities of the craniofacial area, measurements of ears, eyes, head and face should be included¹⁰. If the syndrome includes disproportion, span, lower segment and measurements related to the area of disproportion should be included at various ages. A description of function, symmetry (or lack of symmetry) and longitudinal measurement during growth are desirable.
 - A full, system-by-system physical examination recording structural and minor anomalies.
 - Photographs of the face (anterior–posterior and lateral), of the whole body in underwear or bathing suit (anterior–posterior and lateral, preferably against a standard measuring instrument) and of specific features unique to the condition.
 - If bony changes are noted anywhere, a full skeletal survey should be recorded.
-

the DNA samples and recording relevant clinical features. When papers describing mutations in genes responsible for human diseases are published, clinical details must be there. Experience tells us that if the full clinical descriptions have not been collected and recorded on the affected individuals whose DNA has led to identification of a mutation, they never will be (it is very difficult to contact people again). Authors need to insist that the clinical

information collected by their colleagues be published or at the very least be readily accessible. Ideally, it should be part of every publication, although an accessible website is another option. If this information is not available, opportunities to advance our understanding of biology and human disease and to help affected individuals will be lost. □

1. Amiel, J. *et al. Nat. Genet.* **33**, 459–461 (2003).
2. Dodé, C. *et al. Nat. Genet.* **33**, 463–465 (2003).

3. Fantes, J. *et al. Nat. Genet.* **33**, 461–463 (2003).
4. Haug, K. *et al. Nat. Genet.* **33**, 527–532 (2003).
5. O'Driscoll, M., Ruiz-Perez, V.L., Woods, C.G., Jeggo, P.A. & Goodship, J.A. *Nat. Genet.* **33**, 467–501 (2003).
6. Puls, I. *et al. Nat. Genet.* **33**, 455–456 (2003).
7. Robertson, S.P. *et al. Nat. Genet.* **33**, 487–491 (2003).
8. Slager, R.E., Newton, T.L., Vlangos, C.N., Finucane, B. & Elsea, S.H. *Nat. Genet.* **33**, 466–468 (2003).
9. Majewski, F., Ranke, M. & Schinzel, A. *Am. J. Med. Genet.* **12**, 23–35 (1982).
10. Hall, J.G., Froster-Iskenius, U.G. & Allanson, J.E. *Handbook of Normal Physical Measurements* (Oxford Medical Publications, Oxford, 1989).

The smoking gun of gene transfer

Gene transfers from chloroplasts to the nucleus occur naturally over evolutionary time scales but have always been inferred from sequence comparisons, never directly observed. A notable new report has caught plastid-to-nucleus gene transfer in the act, providing exciting opportunities to study its mechanistic details in the laboratory and its ecological implications in nature.

About 1.5 billion years ago, a free-living cyanobacterium took up permanent residence in a eukaryotic host cell. The results of that symbiosis are today's chloroplasts, the photosynthesizing organelles of plants. One of the most important processes that occurred en route to the establishment of the chloroplast as a stably inherited organelle (as opposed to a transient endosymbiont) was the transfer of genes from the cyanobacterial symbiont's genome to the host's nuclear chromosomes. Recent estimates suggest that as much as 18% of the nuclear genes in *Arabidopsis thaliana* ultimately stem from the ancestral plastid genome. But gene transfer from chloroplasts to the nucleus has always been indirectly inferred from gene-sequence comparisons—until now. In a recent report, chloroplast-to-nucleus gene transfer was observed *in flagrante*, capturing what may be nature's most prevalent kind of genetic promiscuity: lateral gene transfer from organelles to the nucleus.

Using experimental techniques pioneered by Pal Maliga, Huang *et al.* (*Nature* **422**, 72–76 (2003)) introduced a fragment of foreign DNA into the chloroplast genome of tobacco, *Nicotiana tabacum*. The fragment, which was stably integrated into the circular chloroplast chromosome by homologous recombination, contained an antibiotic resistance marker that is expressed in the chloroplast. This allowed them to obtain



A rare (1 in 16,000) kanamycin-resistant plant (photo courtesy of J. Timmis).

plants whose plastids—but only their plastids—were homogeneously transgenic (homoplasmic transformants). Their decisive trick was then simple: in addition to the plastid-specific resistance marker, their chloroplast DNA-integrated fragment also contained a kanamycin-resistance gene, which possessed a spliceosomal intron under the control of promoter specific to nuclear gene expression. These homoplasmic lines were used as the male (pollen) donor in crosses to wild-type female recipients. The resulting seed was sown on medium containing kanamycin. The expectation was that only such seedlings would survive in which the kanamycin-resistance marker specific to the nuclear expression machinery had been transferred in a functional state from the chloroplast to the nuclear genome. Among 250,000 progeny tested, they found 16 independent kanamycin-resistant plants (see figure) with heritable nuclear insertions of the chloroplast fragment, which in most cases segregated in a mendelian fashion. That translates to a rate of successful chloroplast-to-nucleus gene transfer accompanied by stable nuclear expression of one in 16,000 pollen grains tested.

As the authors point out, this surprisingly high frequency of gametes that acquired a fragment of chloroplast DNA is still an underestimate for the overall transfer rate of bulk DNA from the chloroplast to the nucleus, because only transfers of the engineered resistance gene that resulted in proper expression were detected. Transfers that did not encompass the region of chloroplast DNA into which the resistance gene had been inserted also went undetected. Hence, the true rate of bulk chloroplast DNA transfer is probably higher. On the other hand, *N. tabacum* is an allotetraploid, a circumstance that may buffer the potentially deleterious effects of large DNA insertions in the nuclear chromosomes, thus potentially permitting a higher transfer rate than might be expected in diploids. Yet there is no evidence one way or the other to suggest that the chloroplast-derived fragments are being inserted into active genes. Further study of the integration sites, their surrounding regions and the tendency of the integrated DNA to undergo deletion should provide additional clues.

The findings suggest that one in every few thousand plants that we see in the greenhouse, in cultivated fields or in nature has a large and freshly incorporated piece of chloroplast DNA somewhere in its nuclear chromosomes that was integrated only one generation ago. Given the geological time scales over which this rapid and efficient intracellular gene transfer mechanism has been operating, the biggest surprise of all is perhaps that there is any chloroplast-specific DNA left at all—but that is another story. □

—William Martin

Institute of Botany III, University of Dusseldorf, Dusseldorf, Germany. e-mail: w.martin@uni-duesseldorf.de