nature genetics
volume 4 no. 1 may 1993

The joys of reverse biology

For the legions of research groups scouring the human genome for important disease loci, cloning the gene involved can often be the least of their problems. This enterprise, which has been termed 'reverse biology', is winning more and more disciples who have experienced that enlightening moment in front of the computer as the latest homology algorithm mulls over the fruits of months and perhaps years of effort, in an instant selecting a new field of inquiry to be thrust upon the hapless investigators. In some cases, it might be signal transduction or membrane transport; but in others, even the computer is baffled by a virgin sequence that fails to make any significant matches. Many such experiences were relived by an international cast of speakers at the first ever *Nature Genetics* conference*, which was held last month to mark the first anniversary of its launch.

Perhaps the greatest suspense is currently tantalizing those who have just grasped one of the most sought after prizes in human genetics — the gene for Huntington's disease (HD). As described by Marcy MacDonald (Massachusetts General Hospital), the ten-year quest to find the Huntington's gene, which was successfully concluded in March (*Cell* 72, 371–382;1993), has left a host of new questions, chief among which is the precise function of the gene product, huntingtin. Sadly, despite some 11,000 nucleotides encoding a protein with more than 3,000 amino acids, no homologies have so far been found, and certainly no clues as to how such a ubiquitously expressed gene can lead to such selective neuronal loss.

Lovell-Badge — the joy of sex determination.

Although the first partial cDNAs for huntingtin were pulled out almost 18 months ago, the 5' end of the gene remained elusive until much more recently. And it was this end of the gene which truly incriminated it in the pathogenesis of the disease, revealing an expanded trinucleotide CAG repeat in HD patients. As MacDonald put it, there is a "rough but believable" correlation between the age-of-onset of the disease and the number of repeats. For example, juvenile-onset cases of HD (under two years of age) are associated with large expansions. Many other interesting features of these repeats are also emerging: the repeat can be seen to have either expanded or contracted in some affected children. The basis of apparent 'sporadic' cases of HD has also been attributed to newly expanded repeat lengths — two such cases were associated with relatively slight increases from roughly 35 repeats to more than 40 repeats, just above the normal range.

*Human Genetics: Mapping the Future. 1st International *Nature Genetics* conference, Omni Shoreham hotel, Washington DC, 1–2 April, 1993.

Collins, MacDonald and King reflect on past successes and those to come.

The models that will bear fruitful comparisons with HD include spinal and bulbar muscular atrophy — another neuronal disease associated with a coding CAG repeat sequence (Kenneth Fischbeck, University of Pennsylvania)—and the most common inherited form of mental retardation, fragile X syndrome (Steve Warren, Emory University, Atlanta). Fragile X is caused by an expanding CGG repeat just ahead of the coding portion of the FMR-1 gene, but little has been uncovered about the putative function of FMR-1. However, new analyses have resolved two 30-amino-acid domains in FMR-1 which closely resemble motifs in RNA-binding proteins, such as yeast MER1. Interestingly, an invariant isoleucine residue in this conserved sequence happens to be the site of the one point mutation in *FMR-1* described so far (*Nature Genet.* 2, 31–35;1993) providing strong support for the significance of both findings.

For two other neurological conditions, Alzheimer's disease and Charcot-Marie-Tooth (CMT) disease, genetic heterogeneity necessarily complicates the analysis. As John Hardy (University of South Florida) pointed out, the road to the discovery of the amyloid precursor protein (APP) gene defect in some early-onset familial cases of Alzheimer's disease (FAD), was strewn with mistakes, caused in large part by the inclusion of families of different aetiologies in the early linkage analyses. New descriptions of APP mutations (see page 11, this issue) indicate that although rare, these instances are not related, occurring as they do in different racial backgrounds. The newly mapped locus on chromosome 14 gives rise to a more prevalent cause of FAD, accounting for perhaps 90% of early-onset patients (< 50 years). In CMT, the molecular basis of some autosomal and X-linked forms remains to be described (Christine van Broeckhoven, University of Antwerp), but the peripheral myelin protein 22 appears to be the basis of CMT type 1a, as evidenced by an unusual duplication of more than 1 megabase of flanking DNA on chromosome 17, and the discovery of at least a couple of specific *PMP-22* mutations in non-duplication patients. Interestingly, a deletion of this same interval has recently been reported, giving rise to a related condition called hereditary neuropathy with liability to pressure palsies (*Cell*, 72, 143–151; 1993). *PMP-22* is highly expressed in nerve Schwann cells, but understanding the means by which an increased gene dosage can produce the same phenotype as a point mutation may have to wait until suitable mouse models are developed.

Emerging evidence for genetic heterogeneity in Alport syndrome, which is characterized by renal failure resulting from malformation of the glomerular basement membrane (GBM) and deafness, was presented by Steven Reeders (Yale University School of Medicine). The physically linked pair of α3(IV) and α4(IV) collagen genes on chromosome 2q are linked to autosomal recessive forms of the disease, and together with the X-linked α5(IV) collagen, form the main structural meshwork of the GBM. The sixth isoform of type(IV) collagen has also been identified adjacent to the α5(IV) gene.

Positional cloning has made a considerable impact on the study of cancer, leading to the isolation of genes for familial polyposis coli (Kenneth Kinzler, Johns Hopkins University School of Medicine) and neurofibromatosis type 1 (NF1)(Francis Collins, University of Michigan). The spectrum of mutations in the APC tumour suppressor gene on chromosome 5q in both hereditary and sporadic forms of colon cancer is very similar, usually involving a premature truncation of the 2,843 amino acid APC protein, leaving 1,600 residues at the N terminus. And yet there are cases where two 'hits' do not seem to be required for colon cancer to ensue, suggesting perhaps that APC can act in a dominant-negative manner. APC does, in fact, show weak homology to myosin and other intermediate filament proteins, and *in vitro* assays have shown that the N-terminal region of APC binds to the wild-type protein. Interestingly, the binding sequences in p53 are located near the C terminus, and thus truncations of p53 would not be expected to produce the same dominant effect; in agreement with this idea, p53 lesions have been shown to be missense mutations (Curtis Harris, National Institutes of Health).

Just moments before confirming his appointment as director of the National Center for Human Genome Research, Collins described similar progress in elucidating the potential role of the

NF1 gene product, neurofibromin. Recent studies have shown that somatic deletions of *NF1* are associated with a variety of other tumours, including melanoma, where it appears to be a late event. Neurofibromin has a well-defined region of homology to a number of GTPase-activating proteins, but that leaves unexplained the putative function of the remainder of this large protein. However, immunofluorescence and biochemical studies now suggest that neurofibromin colocalizes with the intracellular microtubule network.

Collins has been a central figure in three positional cloning successes, most recently Huntington's disease, but now his attention, and that of many other investigators, is focused on hereditary breast cancer. Mary-Claire King (University of California, Berkeley), by contrast, has been studying this disease for nearly 20 years, but indicated that the search for the breast cancer gene is nearing an end. As others before her, King praised the contributions of the Genethon 'mega-YAC' library, which despite a significant degree of chimaerism, has allowed a consortium involving King, Collins and others to clone the critical region on chromosome 17q. Although this region has been narrowed down to about 1 Mb all viable candidate genes appear to have been excluded. It is estimated that one in 150 women carries the defective *BRCA1* allele, making this one of the most common — and tragic — of all genetic disorders, yet representing only a fraction of the total incidence of the disease. King admits that she cares little which field investigators will ultimately face when the gene is found, with the minor caveat: "God, let it not be immunology!"

There are eight known tumour suppressor genes, and valuable information may result from their homologous inactivation in mice (Tyler Jacks (Massachusetts Institute of Technology). Mice deficient for p53 are known to be more susceptible to a variety of tumours, such as lymphomas, but surprisingly, homozygotes are developmentally normal, although they die young. Jacks presented exciting new information to suggest that p53 may have a role in apoptosis (cell death). In normal cells, p53 levels increase dramatically in response to irradiation (due to post-translational stabilization). But although p53 is not required for cell death *per se*, p53-deficient lymphocytes are extraordinarily resistant to γ-irradiation. p53 may be required to facilitate cell death, for example when developing T cells have produced aberrant gene rearrangements, and that in the absence of the protein, such abnormal cells may proliferate and give rise to tumours.

Mouse studies have also proved invaluable in understanding the function of other tumour suppressor genes such as that for Wilms' tumour (*WT1*) and DNA-binding proteins such as PAX6, which is involved in aniridia (Veronica van Heyningen, MRC Human Genetics Unit, Edinburgh). Recent results have shown that defects at both loci can give rise to other disorders — Denys-Drash syndrome for *WT1*, and a solid tumour for *PAX6*. With a little more information about the expression of these and related genes, it may be possible to predict 'candidate diseases' for other mutations, for example Peter's Anomaly and Gillespie syndrome may both be linked to *PAX6* defects.

Considerable progress has been made in understanding the molecular basis of embryonic pituitary gland development, and the role of specific transcription factors such as *Pit-1*, which is required for development of three of the five pituitary cell types (Michael Rosenfeld, University of California, San Diego). Both mouse and human mutations in *Pit-1* have been associated with dwarfism and deficiencies of growth hormone and the recently cloned receptor for growth hormone releasing factor (GRF). Interestingly, many homologues of POU-domain transcription factors such as Pit-1 have been detected in other tissues, including the immune and nervous system. The most recent addition to the family is *Skn-1*, which is an enticing candidate for various mouse and human skin disorders (*Science* **260**, 78–82;1993).

Another DNA-binding protein whose function is of intense interest to at least half the population is *SRY*, which, as shown by Robin Lovell-Badge (National Institute of Medical Research, London) and colleagues, encodes the testis-determining factor on the Y chromosome. The presence of SRY initiates the male pattern of differentiation of the genital ridge rather than the 'default' (female) pathway. However, SRY is neither necessary nor sufficient for Sertoli cell differentiation, and despite convincing evidence *in vitro* that SRY acts as a transcription factor — it has been shown to induce bends in DNA upon binding of up to 130°, which may physically influence the activity of other regulators — the *in vivo* downstream targets remain elusive. It should be said, however, that the attention of Lovell-Badge and colleagues has been diverted recently with their surprising discovery that in adult mouse testis, *SRY* is transcribed as a circular 1.3 kilobase RNA. What that means is still anybody's guess.

Venter and Cohen place their bets for 1994.

Future directions

Two of the most important rationales for cloning new genes involved in hereditary disease are the promise of improved diagnosis and treatment. The great potential of blastomere analysis for couples with a history of genetic disease outlined by Mark Hughes (Baylor College of Medicine) emphasizes how far the field has come in a very short time. By amplifying DNA from single cells removed from eight-cell blastomeres grown *in vitro*, reimplantation of the embryos can be confined to those that are not only healthy, but have a wild-type genotype. A successful birth has already been reported for a British couple who were carriers for cystic fibrosis (CF), and more sophisticated assays are being developed to detect many of the less common mutations. The method is also being applied to sex-linked conditions such as Lesch-Nyhan disease and chromosome aneuploidies. Ron Crystal (Cornell University) believes that the 1990s will see the development of the concept of 'the gene as the drug', and imminent clinical trials on the use of adenoviral vectors for CF will be an important barometer. In one of the first demonstrations of a physiological response from gene therapy, administration of the erythropoietin gene in an adenovirus gives rise to a significant increase in the haematocrit of cotton rats. Although many questions remain about the safety and regulation of such forms of therapy, the strategy may offer a cheaper alternative to current methods of treatment involving the recombinant erythropoietin protein.

In a fitting finale, two leading players in the genome project provided updates on the differing, but hopefully complementary, approaches to deciphering the human genome. For Daniel Cohen (Centre d'Etude Polymorphisme Humain), the goal is to produce comprehensive physical and genetic maps of the chromosomes. The current genetic map contains more than 2,000 highly polymorphic microsatellite markers, leaving only two holes of more than 20 cM. In total, Cohen anticipates that more than 10,000 markers will be available by the end of next year, leaving no gap larger than 3 cM. This map may then be combined with the physical map being generated by means of analysis of the much-maligned 'mega-YACs'. These YACs contain large (1–1.5 Mb) segments of DNA, but like all such clones are prone to chimaerism — about 20 or 30%, estimates Cohen. He also agrees that there are some discrepancies between his recently published map of chromosome 21 (*Nature* **359**, 380–387;1992) and those from other sources, and that future studies will resolve these differences. Cohen's group is now generating overlapping contiguous maps of YACs across the genome by hybridizing repetitive sequences to gridded YAC clones, storing the images digitally and performing pairwise comparisons. So far, around 1,000 contigs each of 20 YACs, spanning on average 2.5–3.5 Mb, have been assembled, which should correspond to 70–100% of the genome. These YACs are already being screened with hundreds of genetically mapped sequence-tagged sites (STSs).

Such cross-referencing of genetic and physical maps is very much in the mind of Craig Venter (The Institute of Genomic Research), except that in his case, the probes would be human gene sequences, not simply anonymous markers. According to Venter's latest estimates, his group is well on the way to characterizing the vast majority of human genes within 18 months, by generating up to 500,000 bases of sequence, or 1,000 'expressed sequence tags' (partial cDNA sequences or ESTs) per *day*. Using advanced super computing technology, it then takes a mere 15 seconds to conduct a thorough homology search for each sequence, and by analysing transcripts from more than 40 different cDNA libraries, redundancy will be minimized. The potential of these new 'molecular profiles' for comparing the pool of active genes at different stages of the cell cycle, or in monitoring cancerous changes, is considerable. The daunting task of mapping the burgeoning supply of ESTs remains, but progress is being made in this area too, such that approximately 20% of mapped human genes are already ESTs. If the potential for such projects is realized, then, as Venter suggested, perhaps we are witnessing 'the end of reverse biology'. We should know a great deal more by the time *Nature Genetics* hosts its second anniversary conference, in Toronto on April 7–8, 1994. □