nature genetics May

Reading between the genes

The most extensive comparison so far of contiguous genomic DNA in humans and mice reveals a surprising degree of conservation in intergenic sequences.

WHATEVER the true number of genes in the human genome — let's say 75,000 give or take 20,000 — the burning question remains as to what function (if any) the often vast lengths of DNA that have no direct purpose for encoding peptides might serve. These mysterious sequences, which may make up as much as 97% of the human genome for example, range from apparently barren stretches of highly repetitive DNA threading the individual genes together to the numerous sequences known as introns that punctuate the coding regions of most genes.

Over the years, many people have come to regard these inter- and intragenic curiosities as simply 'junk' DNA, evolutionary relics that can tolerate endless amounts of change (polymorphism) without any untoward effect on the host organism. Of course, there are a few exceptions. Some genes are clearly regulated in part by crucial DNA motifs that may sit many kilobases from the gene itself, but maybe these are just the exceptions that prove the rule.

Or are they? As the amount of DNA sequenced from humans, mice and other organisms grows, it is becoming possible to compare matching pieces of DNA from different species to discern the amount of sequence conservation between them. In the geneticist's creed, evolutionary conservation is a key indicator of functional importance, even if that precise role is not immediately apparent to the hapless investigator. Until recently, such comparisons have been handicapped by the limited amount of contiguous genomic DNA sequence available. But with the continuing improvements in automated DNA sequencing technology, prospects are looking up.

Writing in the May issue of *Nature Genetics*, Ben Koop and Leroy Hood (*Nature Genet.* **7**, 48–53; 1994) describe their scrutiny of almost 100 kilobases of homologous human and mouse genomic DNA sequence from the T-cell receptor locus. T-cell receptors are membrane-bound heterodimers that closely resemble antibodies and are made up of two different pairs of polypeptide chains, either α/β or γ/δ. In addition to constant (C) regions, a series of variable (V), diversity (D) (in the case of the γ/δ chains) and junction (J) segments contribute to the impressive variability exhibited by different receptors.

Koop and Hood focused on the 100-kilobase Cα/Cδ regions of mouse and human, which by their calculation contain 50 and 61 Jα segments, respectively. Together, these coding regions account for only 6% of the total DNA. Of course, it is not surprising that these genes should show a high degree of conservation — of the order of 75% in fact. But, surprisingly, a survey of the entire region, including 94% non-coding DNA, produces an astonishing similarity value of 71% between mouse and human. This is the highest figure seen so far and contrasts with the results of other reasonably large sequence comparisons performed previously, such as those for globin or crystallin gene clusters.

As yet there are no firm answers to explain this degree of sequence conservation over such a large expanse of DNA. In the authors' view, it is probably the result of natural selection acting within and beyond the confines of coding elements, supporting the notion of chromosomes as 'information organelles' which perform additional functions such as the storage, copying and evolution of information. Of course, there may be more mundane explanations. Further large-scale sequence comparisons will be required to make the distinction.

Human and mouse homologies are being exploited in other ways, notably in the search for genes controlling complex traits, which can obviously be performed more easily in fast-breeding mice with controlled genetic backgrounds. One promising area is the search for genes underlying behavioural traits, including the tendency to abuse substances such as drugs and alcohol. In another report in this month's issue, Wade Berrettini and colleagues describe some potentially important genetic differences between two mouse strains which have a vastly different preference for morphine (*Nature Genet.* **7**, 54–58; 1994).

When given the choice between a bottle of sweetened quinine (the sweetener helps to remove a bitter aftertaste) and a suspension of morphine, the C57Bl/6J inbred mouse will happily consume as much suspension as 300 mg per kg per day, resulting in a characteristic tail muscle contraction called Straub tail and severe withdrawal symptoms. By contrast, the DBA/2J mouse can be tempted to take 10 mg per kg per day, but no more. These dramatic differences are genetically controlled, and thus by performing a quantitative trait loci analysis that has successfully been used to map murine loci for hypertension, cancer and epilepsy, Berrettini and colleagues have been able to identify the positions of three genes responsible for most of the variation in morphine consumption.

Their approach was simply to breed a second-generation intercross between the two strains, and then type the animals with the highest and lowest morphine consumption with some 150 microsatellite markers. In so doing, they detected three loci on mouse chromosomes 10, 6 and 1 (in order of significance) that together appear to account for almost 90% of the total genetic variance. As in humans, the addictive behaviour of the C57Bl/6J mice crosses pharmacological boundaries — that is, they are also partial to substances such as alcohol and cocaine. The search is now on to see whether genes in the homologous human chromosomal regions may explain human patterns of substance abuse.

One gene that will be found before then — although it is taking longer than expected — is the hereditary breast and ovarian cancer gene, *BRCA1*, on the long arm of chromosome 17. About 5% of breast cancer cases among women are inherited, and because many would argue that the disease is reaching epidemic proportions, that makes its familial form one of the most common inherited disorders of all. Men, too, contract breast cancer. Its incidence is 100 times less than in women, but are inherited forms of male breast cancer also attributable to *BRCA1*? The answer at this stage appears to be that they are not. A large consortium study, led by Michael Stratton, has looked at 22 families and found no evidence to support linkage of male breast cancer to the *BRCA1* locus (*Nature Genet.* **7**, 103–107; 1994). It would seem, then, that at least one other important gene is accountable for these cases, just as *BRCA1* is evidently not the only gene responsible for hereditary breast cancer in women.

Kevin Davies

Kevin Davies is the Editor of Nature Genetics.

.Also in May's *Nature Genetics*: gene defects in male pseudo-hermaphroditism and stationary night blindness; *p53* mutations in anaplastic Wilms' tumour; radiation hybrids and 'genomic sequence sampling' for whole genomes; and *VHL* mutations in renal carcinomas.