DDT Vol. 8, No. 18 September 2003

Highly mutated regions could show basis for human disease

Emilia Rappocciolo, e.rappocciolo@elsevier.com

The search for the genetic basis of human disease is going to get a boost form a new mathematical model that has been developed by a group of scientists at The University of Texas, Southwestern Medical Center, in Dallas (http://www.swmed.edu) [1].

Disease risk

Genes carry all the information that determines an organism's characteristics, and scientists are well aware that, embedded in the DNA sequence, there are bits of information that can be used to predict risk of disease and genetic contribution to disease development.

Nearly all diseases that have been identified so far through linkage analysis are diseases in which mutation in one gene is necessary and sufficient to cause disease (http://ncbi.nlm.nih.gov/OMIM). A frequently used example to describe this situation is cystic fibrosis in which the mutation of a single base in the *CFTR* gene is directly linked to the development of the disease.

More commonly, disease is caused by a combination of mutations in several genes that contribute to its development in different degrees and fashions. If each locus contributes modestly to disease aetiology, more powerful methods of analysis are needed to point to all the genes involved in a particular condition.

Single nucleotide polymorphisms (SNPs)

SNPs are single basepair positions in genomic DNA at which sequence variations exist in normal individuals in some population(s), where the least frequent allele has an abundance of 1% or greater. Coding SNPs (cSNPs) are are found in coding sequences of the genome. In a complex disease a combination of several SNP alleles in sets of key genes are thought to be the cause. If a factor contributes to an increased risk for disease occurrence, that factor should be found at higher frequency in individuals affected by that condition, compared with healthy individuals [2]. In the past few years, the use of SNPs to map complex disease has been on the agenda of many laboratories, and several efforts have been made to recognize the prospects and drawbacks of such approach [3].

The paper by Monica Horvath *et al.* [1] opens new avenues in this search. The team, led by Harold Garner, Professor of Biochemistry and Internal Medicine at Southwestern, compared four public databases: HGMD (human genome mutation database), dbSNP, CGAP–GAI (cancer genome anatomy project–genomic annotation initiative) and TSC (SNP consortium database). Because there are 64 codons in the genetic code, cSNPs for each database were sorted into $3 \times 3 \times 4^3 = 576$ possible codon mutation classes (CMCs; see Figure 1). The raw number of mutations per CMC was then converted into a frequency statistic, which was adjusted by wild-type codon usage. When all the CMC frequencies for a wild-type codon are summed, the authors can estimate the relative mutational load for a specific trinucleotide. Following analysis of the four databases, this group showed that a few CMCs were enough to describe a considerable fraction of mutation in functional genes. For example, the top 5% of CMCs account for 27.4% of the observed variants in dbSNP, while the bottom 5% account for only 0.02% [1].

Potential candidates

Horvath says that this will change the way we look at sequence databases and will enable scientists to define motifs statistically, and subsequently pick genes that display those motifs as more probable candidates for complex disease. 'This method will give us an idea of where to look, and find the best place in which a candidate gene for complex disease can be found.' The genome can then be 'prioritized' into

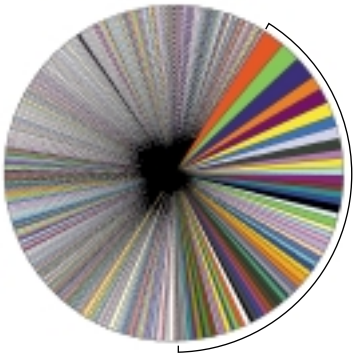

Figure 1. Double base single nucleotide polymorphism (dbSNP) coding region point mutations as categorized into 576 mutation classes. The section highlighted by a black line indicates SNP classes respresenting extremely probable genetic variation. Figure kindly provided by Monica Horvath at the University of Texas, Southwestern Medical Center, Dallas, USA (http://www.swmed.edu).

 PII: S1359-6446(03)02849-6

sections that are more likely than others to contain noteworthy genes.

Christopher Lee, Assistant Professor at the Department of Chemistry and Biochemistry, University of California, Los Angeles (UCLA; http://www.ucla. edu) says: 'This result is quite significant for guiding disease mutation searches. Rather than having to go the usual route of searching for polymorphisms that show linkage to an unknown disease mutation, this paper suggests the possibility of directly looking for the mutations that are most likely to cause disease.' He continued, 'For some time, people have been excited about – and arguing about – the notion that a subset of the most common mutations might cause a large proportion of human disease. This result appears to point in that direction. An obvious question is to look at some diseases whose mutational basis is very well studied, and ask what fraction of disease in the population can actually be predicted by the authors' high frequency mutations'. This is indeed what the scientists at Southwestern are investigating at present.

Low hanging fruit

The pharmaceutical industry is seeing rapid evolution in the development of drugs based on information contained in genomes. Identifying the genes that are more likely to be involved in the aetiology of complex disease will have significant impact on the speed of the drug discovery process. Scientists will be able to zoom into the human genome and discover areas that are more likely than others to contribute to disease, which can then provide them with the necessary information to develop new pharmacological tools and drugs. The complexity of the human genome can be thus reduced, enabling researchers to pick significant SNPs as 'low hanging fruits' that are accessible and easy to locate in the DNA sequence.

References

1 Horvath, M.M. *et al.* (2003) Low hanging fruit: a subset of human cSNPs is both highly non-uniform and predictable. *Gene* 312, 197–206
2 Brookes, A.J. (1999) The essence of SNPs. *Gene* 234, 177–186
3 Lee, C. (2002) Irresistible force meets immovable object: SNP mapping of complex diseases. *Trends Genet.* 18, 67–69

Slime mould clue to depression

Laura Spinney, BMN News

Studies of slime mould are leading British researchers to an explanation for the side-effects of one of the most commonly prescribed drugs for the treatment of bipolar affective disorder, lithium.

Lithium is commonly used to treat the disorder but how it works, and what triggers side-effects, remains unknown. By studying the effects of the drug on the slime mould *Dictyostelium discoideum*, researchers have succeeded in teasing apart the signalling pathways that give rise to its mood-stabilizing effects and, by contrast, its teratogenic side-effects. Simply by altering the dose, they show it is possible to switch between the two.

Lithium: signalling pathway

Adrian Harwood of the Medical Research Council Laboratory for Molecular Cell Biology, University College London

(http://www.ucl.ac.uk), and colleagues, got an inkling last year of the signalling pathway involved in lithium's mood stabilization when they treated cultured rat sensory neurons with the drug and found that it inhibited the collapse of their growth cones.

Simply by adding inositol, the researchers could reverse the effect – and the same turned out to be true for two other common mood stabilizers, carbamazepine and valproic acid. For the first time, they had direct evidence that the inositide-signalling pathway was involved in the therapeutic action of lithium. However, lithium has a range of effects, depending on its dose, some of which are distinctly aversive. If administered in the first trimester of pregnancy, for instance, it can lead to malformations in the foetus. It is also unknown why some patients develop resistance to the drug while others do not.

Of mould and man

D. discoideum is a slime mould that expresses both the inositide pathway and the Wnt-signalling pathway, which makes use of the protein GskA (the *Dictyostelium* homologue of the human protein, Gsk-3). Both pathways are highly conserved across mammals, including humans. The Wnt-signalling pathway has been implicated in lithium action in

  PII: S1359-6446(03)02851-4