

Abstract

Locus Specific Databases in The 21st Century

Bruce Gottlieb, Lenore K. Beitel and Mark Trifiro

*Lady Davis Institute for Medical Research, Department of Medicine, McGill University
Montreal, Canada*

As we enter the 21st century most locus specific genetic databases (LSDB) remain firmly entrenched in the 20th century. Their function has been to act as passive repositories for listings of gene alterations that result in proteins whose putative altered structure presumably results in diseased or non-diseased phenotypes. The effectiveness of these databases has been predicated on a number of assumptions. Firstly, that a particular mutation is indeed responsible for an altered phenotype. However, this clearly needs to be experimentally proven but in most cases is not done. Secondly, it is assumed that the mutation has been inherited, i.e. is germline, and therefore will be an excellent predictor of the inheritance pattern of a particular condition. However, increasingly it is being observed, particularly in diseases such as cancer, that the mutations are often somatic rather than germline in origin. The third assumption is that there is a clear distinction between mutations and polymorphisms (gene alterations that occur in >1% of a population). Again, in many cases this distinction is becoming increasingly more difficult to determine. These assumptions have thus resulted in making most LSDBs in their present form, less than ideal for determining the significance of an alteration to a gene as the cause of a particular disease or condition. In this paper we will report on a number of recent developments that have occurred over the past few years that when introduced into LSDBs will make them both more relevant and useful. Among the developments that will be discussed, are new tools for protein modeling allowing for the creation of 4D as opposed to 3D molecular structures. This will allow for a much better assessment of the functionality of a particular mutation or polymorphism. In addition, the creation of models that simulate protein-protein interactions will be particularly significant in light of the large number of associated proteins now being identified. The huge array of associated proteins now being identified has opened up a whole new area of research into the functioning of locus specific diseases. Thus, while many have predicted that in the proteomic age LSDBs will become increasingly irrelevant, it seems that recent developments might render such judgments somewhat premature.