MEDSEQ: Development of a Service for the Analysis of Disease Genes

German Integrated Map (GIM)

Jens Hanke, Gerrit Lehmann, Harald Pankow, Inga Zastrow, Dave Brett , Jens G. Reich

Max−Delbrück−Center for Molecular Medicine
Dept. of Bioinformatics
Robert−Rössle−Str. 10, D−13125 Berlin−Buch, Germany

The study of disease genes is of particular importance within the human genome project. We have established a service for medical users that allows a thorough analysis of disease genes, including approaches for sophisticated gene location as well as integrated sequence analysis using complex models. Additionally, complete genome cross−referencing allows a rapid and complete analysis for the recognition of potential disease genes. The public database MEDSEQ (http://www.bioinf.mdc−berlin.de/marker.html) offers:

i) complete EST clustering with homology results, paralog information and SNP mapping,
ii) a complete human transcription map with genetic and physical scaling,
iii) cross−referencing to other model organisms with a connection to mapped human and mouse phenotypes,
iv) disease relevant genomic regions can be compared to sequence and function−specific information from various organisms.

With the integrated map, researchers can localize a gene of interest based on mendelian traits to a genomic interval. The large growth in sequence information makes the finding of candidate genes purely from bioinformatic techniques of greater importance. In the current state we looked for all human positional cloned associated disease genes which were published in the last year (98) and had previously been mapped to a region closer than 3 cM. From 8 published genes we were able to identify 4 genes based only on information gained from our integrated map database (we took only the current data status in the year 98) and extensive literature investigation. In principle these finding make possible the discovery of new disease associated genes within a 3cM region without labour intensive position cloning techniques. The recent rapid accumulation of data from sequencing, mapping and functional identification of various organisms should enable MEDSEQ to discover many more such candidate genes.