

Gentrepid: a webserver for candidate disease gene prediction

Erdahl T. Teber, Jason Y. Liu, Richard A. George, Sara Ballouz, Chi N. I. Pang, Boer Xu, Naresh P. Bains, Duncan B. Sparrow, Robin Otway, Guanglan Guo, Diane Fatkin and Merridee A. Wouters
Victor Chang Cardiac Research Institute, Sydney, Australia

Key words: disease gene prediction; common pathway scanning; common module profiling.

URL: www.gentrepid.org

The identification of genes responsible for human disease enables an understanding of disease mechanisms and is essential for the development of diagnostics and therapeutics. Genetic studies can successfully identify genomic regions linked to a particular disease phenotype. Taking the next step of isolating the disease-causing gene can be difficult. For linkage analysis, genomic regions are often large, particularly for complex diseases. Genome-wide association (GWA) studies have greater power to detect genetic variants that confer modest disease risks than linkage analysis does, but even these may identify hundreds of SNPs. Currently, published GWA studies list only the 20–50 most-significant SNPs and their neighboring genes (the “most-significant SNPs/genes” approach), while paying little attention to the rest. A system that predicts and prioritizes candidates using this genetic information would be a great boon.

Here we present the Gentrepid web server, a system for predicting candidate disease genes associated with specific phenotypes using genetic and experimental biological data. The Gentrepid web server allows a user to enter a disease interval and the particular phenotype they are interested in. Standard OMIM phenotypes are available and novel phenotypes can also be entered. Disease loci can be specified by STSs, SNPs or cytogenetic bands. The server interrogates our database, which contains current knowledge of all known disease genes, pathway and interaction data and complete pre-calculated domain annotations for all genes. This information is fed into our prediction tools, CMP and CPS¹, which then predicts novel disease genes within the user-defined interval. The output displays the known disease genes, and predictions made by the various methods, mapped to an interactive display of the human genome. Users’ novel data is password-protected and only accessible by the individual user.

Of the approaches that use links between genes and phenotypes G2D², BITOLA³, GeneSeeker⁴ and TOM⁵ are freely available to academic users as an online service. Of these, BITOLA and GeneSeeker, require significant user input. The fully-automated Gentrepid web server provides a valuable alternative service to geneticists.

References

- ¹ RA George, JY Liu, LL Feng et al., *Nucleic Acids Res* **34**, e130 (2006).
- ² C Perez-Iratxeta, P Bork, and MA Andrade, *Nat Genet* **31**, 316 (2002).
- ³ D Hristovski, B Peterlin, JA Mitchell et al., *Int J Med Inf* **74**, 289 (2005).
- ⁴ MA van Driel, K Cuelenaere, P Kemmeren et al., *Eur J Hum Genet* **11**, 57 (2003).
- ⁵ S Rossi, D Masotti, C Nardini et al., *Nucleic Acids Res* **34**, W285 (2006).