GENEPATH: AN INTELLIGENT ASSISTANT FOR GENETIC DATA ANALYSIS AND PATHWAY DISCOVERY. Blaz Zupan[12], Peter Juvan[1], Janez Demsar[1], Ivan Bratko[1], John A. Halter[2], Adam Kuspa[2], and Gad Shaulsky[2], 1) University of Ljubljana, Slovenia, and 2) Baylor College of Medicine, Houston, TX.

Genetic analysis often defines a framework for the elucidation of biological mechanisms. Mutations are the main tool used by geneticists to investigate biological phenomena. Initially, mutations help to define and catalogue genes that participate in a biological mechanism. Relations between genes are then determined using combinations of mutations in two or more genes. Information on gene relations obtained in this way is then used to construct genetic pathways or networks that outline the molecular details of a biological mechanism.

While it may be fairly straightforward for a geneticist to find the relation between two genes from experimental data, systematic search for all possible gene relations while accounting for all the data becomes complicated when the amount of data increases as many genes, mutations and phenotypes are analyzed. For the purpose of automating the task we have developed a program called GenePath that takes genetic data and any prior knowledge of the genetic network as input, and outputs a list of gene-to-gene and gene-to-outcome relations and a corresponding consistent genetic network.

The particular feature of GenePath is that it mimics expert geneticists by using the reasoning patterns they would otherwise employ manually. For instance, to determine if one gene precedes another, a pattern "*IF mutation of GeneA changes the phenotype P (compared to the wild type) and adding the mutation of GeneB reverses the phenotype P THEN gene A precedes gene B in a pathway for an outcome P*" is used. GenePath currently incorporates about fifteen such patterns, and can determine which genes influence (or do not influence) the outcomes, and which genes block (or do not block) other genes. GenePath uses these patterns to systematically search over the data and obtain all possible gene relations. These serve as constrains for the genetic network that is generated by GenePath as a final result of the analysis.

The particular advantage of GenePath is its explanation mechanism. Users can query on any relation or part of the network discovered, and GenePath can provide the corresponding evidence in terms of experiments and patterns that were required for its inference. The program also facilitates storage and communication of data from genetic experiments in a consistent manner.

GenePath was implemented in the Prolog logic-based programming language, and has an easy-to-use web-based interface (`magix.fri.uni-lj.si/genepath`). GenePath was developed around a genetic pathway that regulates the transition from growth to development in the social amoeba *Dictyostelium discoideum*. It has been tested successfully on other Dictyostelium pathways as well as on two *C. elegans* pathways.