

*Extracting Knowledge From Genomic Experiments By Incorporating The Biomedical Literature*

[James P. Sluka](#)

InPharmix Inc.

We present a technique to extract relevant information from the literature to aid in the analysis of a typical genomics data set. Analysis was conducted using PDQ\_MED, a program based on the assumption that if two genes are found to be related under an experimental paradigm, such as a gene chip experiment, then any literature which relates the two genes is of interest. PDQ\_MED searches MEDLINE for abstracts that contain two or more of the terms in the user's query set.

We have used PDQ\_MED to analyze 160 genes up-regulated in acute myeloid leukemia (AML) from the NCI-60 dataset. PDQ\_MED executed 12,880 queries to MEDLINE and identified nearly 300,000 abstracts that refer to at least one of the 160 terms. PDQ\_MED identified and analyzed a set of 81 terms that can be grouped together via the literature. In addition, there is literature directly linking 52 of the terms with AML. Overall, the literature analysis identified 1028 sentences that directly relate two or more of the query genes.