



Genetic algorithm-based efficient feature selection for classification of pre-miRNAs

P. Xuan¹, M.Z. Guo¹, J. Wang², C.Y. Wang¹, X.Y. Liu¹ and Y. Liu¹

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, P.R. China

²School of Computer and Information Science, Southwest University, Chongqing, P.R. China

Corresponding author: M.Z. Guo
E-mail: maozuguo@hit.edu.cn

Genet. Mol. Res. 10 (2): 588-603 (2011)

Received November 7, 2010

Accepted January 8, 2011

Published April 12, 2011

DOI 10.4238/vol10-2gmr969

ABSTRACT. In order to classify the real/pseudo human precursor microRNA (pre-miRNAs) hairpins with *ab initio* methods, numerous features are extracted from the primary sequence and second structure of pre-miRNAs. However, they include some redundant and useless features. It is essential to select the most representative feature subset; this contributes to improving the classification accuracy. We propose a novel feature selection method based on a genetic algorithm, according to the characteristics of human pre-miRNAs. The information gain of a feature, the feature conservation relative to stem parts of pre-miRNA, and the redundancy among features are all considered. Feature conservation was introduced for the first time. Experimental results were validated by cross-validation using datasets composed of human real/pseudo pre-miRNAs. Compared with *microPred*, our classifier *miPredGA*, achieved more reliable sensitivity and specificity. The accuracy was improved nearly 12%. The feature selection algorithm is useful for constructing more efficient classifiers for identification of real human pre-miRNAs from pseudo hairpins.

Key words: Feature selection; Genetic algorithm; Pre-miRNA; Information gain; Conservation