***Web crawling to Match miRNA ID to expression***

***Abstract***

A web crawler was developed to match miRNA IDs to corresponding expressions. A miRNA ID corresponds to multiple expressions and an expression is from multiple miRNAs. The complexity of building a connection between miRNA IDs and expression is a O(n^3). A web crawler was designed to conduct matching between miRNA IDs and their expressions by automatically search databases hosted on internet.

**Methods**

A web crawler was designed to match each miRNA record to its expression by searching miRbase.org and targetscan.org. Each miRNA containing miRNA id and fold change is passed to mirbase.org and several Uniform Resource Locator (URL) pointing to targetscan.org are fetched. Then the webpage for each URL is searched. A table containing target gene is generated from each webpage at targetscan.org. Each record in the table is compared with the expression records. If the target gene in the table at targetscan.org matches the gene in the expression records, the miRNA id, fold change, gene id, representative sequence and miRNA id fetched from mirbase.org are saved into the output file. In each output record, miRNA id and fold change are from miRNA record, gene id and representative sequence are contained in both the webpage searched at targetscan.org and expression record, the second miRNA id is the id at miRbase.org due to the fact that each miRNA id in miRNA records may correspond to more than one miRNA ids at miRbase.org. For instance, to search the corresponding gene expression for has-mir-361, two miRNA ids at mirbase.org are found, one is hsa-mir-361-5p, one is has-mir-361-3p. Open the URL for each of two miRNA ids at targetscan.org, 5161 and 375 records are found respectively. Search gene id contained in each record at targetscan.org in expression table. If gene id is found, a record is saved into output file. For instance, a record (has-mir-361, -2.40987, VEGFA, NM\_001025366, has-mir-361-5p) represents a match between miRNA records and expression records and saved, which is one of 445 matched pairs.