

Datasets Available for Download

Brian T. Luke (lukeb@ncifcrf.gov)

The datasets available for downloading are found in the links below. [Structure of the Datasets](#) contains a description of their structure.

[Random Intensity Datasets](#): This link lets you download one or more of the 30 datasets that contain 300 features and between 30 and 300 Cases and Controls. The datasets are constructed with random peak intensities so that they contain no biological information. Each dataset has an associated document that describes the results of an analysis using the [BioMarker Development Kit](#) (BMDK), and classifiers based on a [decision tree](#) (DT) and a [medoid classification algorithm](#) (MCA).

[Putative Biomarkers from Random Intensity Datasets](#): Each of the 30 datasets available from the preceding section was analyzed using the [10 methods](#) available in BMDK. They produced between 16 and 31 putative biomarkers. This link contains these 30 datasets but only with intensities for the putative biomarkers. These putative biomarkers contain no real information and therefore should not be able to construct an accurate classifier.

(Last updated 5/3/07)