Methods of Examination

Brian T. Luke (lukeb@ncifcrf.gov)

While blood concentrations of biochemicals are one possible means of examination, it is not the only one. Concentrations of biochemicals in the urine are another, but this can be extended to tears, mucous, or virtually any biofluid. Instead of directly measuring the concentration of specific compounds, mass spectra (with or without pre-fractionation) and 2D NMR of these biofluids can also be used to measure abundance. The advantage of these spectral methods is that the abundance of a compound can be examined from one individual to the next without knowing the identity of this compound. Therefore, examining the intensity or area of spectral peaks is called an undirected search since a list of compounds to examine was not created before hand, while examining blood concentrations of specific biocompounds, such as proteins using ELISA, is a directed search. In addition to direct measurements of concentrations or intensity measurements from spectra, microarray experiments are directed searches since the search is over a set of pre-defined compounds. Microarray results can directly examine tumor or stroma cells to determine, for example, if the individual has invasive or non-invasive breast cancer [Boe-07].

In general, the set of biochemicals whose concentration is directly measured or examined by microarray analysis, as well as the set of peaks present in various spectra, are known as *features*. For each individual, each of these features has a corresponding value. This value can be the concentration, the logarithm of the relative fluorescence intensity, or the intensity or area of the spectral peak. The search for a putative biomarker is over the set of N available features, and each individual is represented by an array of N numbers representing the *intensities* of these features.

As in any classification study, building a classifier starts with a training set. This training set contains the intensities for individuals whose histology is known. These intensities are used to construct a classifier that sufficiently determines the correct group of the individuals in this training set. For example, in the colorectal study [Hab-07] colonoscopies were performed to determine if the individual had polyps or colorectal cancer. Their serum was examined using surface-enhanced laser-desorption/ionization time-of-flight (SELDI-TOF) mass spectroscopy. As with any experimental procedure, studies are being performed to produce standardized procedures for generating reproducible spectra [Bon-05, Hor-05, Rai-05, Sem-05]. It is also important to properly process the spectra, which includes background subtraction, spectral truncation, and consistent spectral scaling. Finally, a peak-picking algorithm should be used to select the features since tests have shown that using individual mass-to-charge ratios (m/z) or binning the spectra generally leads to an inconsistent comparison between spectra. This study identified a set of four peaks with correlated intensities that corresponded to complement C3a anaphylatoxin and this biomarker was very useful in identifying those individuals with colorectal cancer and benign polyps [Hab-07].

(Last updated 4/4/07)