Goal of the Investigation: Early Disease Detection

Brian T. Luke (lukeb@ncifcrf.gov)

Early detection of a disease is very important since it greatly improves the individual's chance of responding well to treatment. For example, the 5-year survival rate from prostate cancer is nearly 100% if it is detected early. Similarly, the 5-year survival rate for ovarian cancer is 95% if caught early, but since 75% of the cases are first observed in the later stages of the disease, the overall 5-year survival rate is less than 50%. It would be nice if there was a single test to determine if an individual had cancer somewhere in their body, but unfortunately such a test does not exist.

Many methods are available to detect a certain type of cancer, such as the detection of occult fecal blood (FOBT) [Mak-04], a sigmoidoscopy or a colonoscopy for colorectal cancer. Unfortunately there is a low compliance to some screening tests (e.g., colonoscopy) and low sensitivity of others (e.g., FOBT) [Sch-02]. In addition, some tests require specialized training and equipment (e.g. colonoscopy) and cannot be effectively implemented as population based screening tools. If a marker in the blood can be determined, the clinical screening can be easily applied to a larger percentage of the population and treatment can be initiated in a timely fashion. Serum markers like PSA for prostate cancer [Cat-94, Pou-97, Tra-94] and CA125 for ovarian cancer [Bas-83] are already being used, though with limited success. In addition, there is increasing evidence that the immune response to cancer results in the detection of specific antibodies in sera [Han-03].

While all cancers have many factors in common, tissue differences and the body's response to different cancers make the test for ovarian cancer (CA125) very different from the test for prostate cancer (PSA). Even within the same tissue, all cancers are not necessarily the same. It is well known that there are two major types of lung cancer, small cell lung cancer (SCLG) and non-small cell lung cancer (NSCLC). It is also known that NSCLC has three major sub-types; adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell undifferentiated carcinoma (LCUC). Each of these has differences in the biochemical processes going on within the cancer cell and one should not expect that the detection, or necessarily the treatment, of these cancers will be the same.

Over 1400 protein products have been identified in human serum [Cha-04], making the blood proteome a valuable resource to search for new markers. If there is a significant change in the blood concentration of a particular protein product, it may be directly related to the particular cancer, indirectly related through a host-response mechanism, or virtually unrelated to the disease by a cascade of events. Therefore, even a detailed understanding of the biochemical mechanism of the disease in the cell does not necessarily lead to a set of proteins that should be measured in the blood. A striking example of this is the finding that serum levels of PSA do not correlate with the levels within prostate cancer cells [Caz-02].

Since the identity of the putative biomarker(s) is not known, an undirected search of the blood proteome needs to be conducted to find any protein products that are differentially expressed in patients with a particular type of disease versus those without the disease.

The ultimate goal is to be able to run a battery of tests and determine if an individual has one of many possible diseases. For example, it has been shown that increased blood concentrations of complement C3a anaphylatoxin are associated with colorectal polyps and cancer [Hab-06], as well as prostate cancer and benign prostate hyperplasia [Ada-02, Luk-07]. If the C3a concentration is sufficiently low, than all for of these conditions can be ruled out, as well any other conditions associated with elevated concentrations. If the C3a concentration is sufficiently high, then further tests would have to be run to determine which of these conditions is present in the individual. In this example, a single biomarker is used to divide all individuals into two groups; those that have one of several conditions and those that do not. For each test, a classifier is required that determines which of two or more groups an individual belongs to. Depending upon its outcome, further tests may be required. If the disease can be accurately diagnosed at an early stage, then the treatment can be effectively applied.

The remaining sections dealing with early disease detection are contained in several web pages. The discussion so far has focused on examining the blood concentration of biomolecules to construct a classifier. This is not the only avenue of examination, and others will be discussed in Methods of Examination. In general, the goal will be to examine one or more *features* and use the *intensities* of these features to construct a classifier that divides the individuals into two or more groups. Before specific classifiers are described, several Requirements for Any Classifier are presented. These include Ransohoff's concepts of *bias*, *chance*, and *generalizability* [Ran-05a, Ran-05b] as well as *coverage*, *uniqueness*, and *significance* [Luk-07].

Two categories of classifiers are compared; fingerprint-based classifiers and biomarker-based classifiers. Fingerprint-based classifiers work on the premise that if two individuals have a sufficiently similar pattern of intensities across a specific collection features and one has a particular disease, then the other should also. A biomarker-based classifier first identifies features that are specifically associated with a given disease state and only uses a small number of these features to construct a classifier. Specific examples of a fingerprint-based classifier are a decision tree (DT) [Ho-06, Liu-05, Yan-05, Yu-05] and the medoid classification algorithm (MCA) used by the laboratories of Petricoin and Liotta [Bro-05, Con-04, Orn-04, Pet-05, Sri-06, Sto-05]. As an example of constructing a biomarker-based classifier, the procedures contained in the BioMarker Development Kit (BMDK) are described. Intermediate procedures such as support vector machines (SVM) and linear descriminant analysis (LDA) will also be described. In examining different procedures it is important to clarify the difference between a "panel of markers" and a "battery of tests," and between a disease category and a disease state. These two discussions lead to the conclusion that "personalized medicine" should be replaced with "State-based diagnosis and treatment".

These classification procedures are used to distinguish between individuals with benign prostate hyperplasia (BPH) and healthy prostates using the spectral intensities published by Adam and coworkers [Ada-02]. This is done using a set of selected peak intensities to construct the classifier and compares the classification accuracy between the correct assignment of BPH and healthy individuals, and the same feature intensities when these assignments have been randomly permuted. These results [Luk-07] showed that the DT and MCA procedures were able to construct accurate results for the datasets with permuted assignments and demonstrate that fingerprint-based classifiers have problems with coverage, uniqueness and significance, and are not generalizable to the underlying population.

These conclusions are reinforced by examining the results of classification studies using artificial datasets that are constructed to contain no information. Using the statistical and classification procedures incorporated in BMDK, between 16 and 31 features representing putative biomarkers are selected for each artificial dataset. Since these features contain no information, the final BMDK classifier based on a distance-dependent K-nearest neighbor algorithm was not able to accurately distribute the individuals into the two assumed groups. In contrast, using the DT and MCA methods on the full datasets produced accurate classification results without the presence of any biological information. In other words good results for a classifier do not imply an underlying biological principle in sharp contrast to the assertions of Petricoin and coworkers [Pet-03], or that because a single dataset can produce many accurate fingerprint-based classifiers it is information rich.

Further tests using only the truncated list of putative biomarkers showed that the DT and MCA procedures are still able to produce very accurate classifiers. The SVM and LDA classifiers yielded accurate results if the number of subjects in each group was reasonably small, and produced poorer results for the larger datasets.

The overall conclusion is that fingerprint-based classifiers such as DT and MCA should not be used to build a final classifier from either the entire dataset or a selected set of potentially important features. It is possible to use an ensemble of fingerprint-based classifiers as another procedure to identify putative biomarkers, as is shown from the results of investigating datasets that contain a single putative biomarker of various strengths.

Since the examination of possible forms of the final classifier or procedures to identify putative biomarkers is by no means exhaustive, the artificial datasets used in these examinations are available to download and use in other procedures. If anyone would like to share the results of their examination of these datasets, please send me an email (lukeb@ncifcrf.gov) containing a URL that points to your analysis so that it can be included in this discussion.

(Last updated 4/27/07)