NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Intelligent decision support system-based biomarker discovery

Belacel, Nabil; Cuperlovic-Culf, Miroslava

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

Publisher's version / Version de l'éditeur:

Bio-IT World Conference & Expo [Proceedings], 2011-04-12

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=f8d3aa03-aad5-43c7-8243-9d645e14f789
https://publications-cnrc.canada.ca/fra/voir/objet/?id=f8d3aa03-aad5-43c7-8243-9d645e14f789

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

NRC·CNRC

Institute for Information Technology

Intelligent decision support system-based biomarker discovery

Nabil Belacel, Miroslava Cuperlovic-Culf
{nabil.belacel, miroslava.cuperlovic-culf}@nrc-cnrc.gc.ca

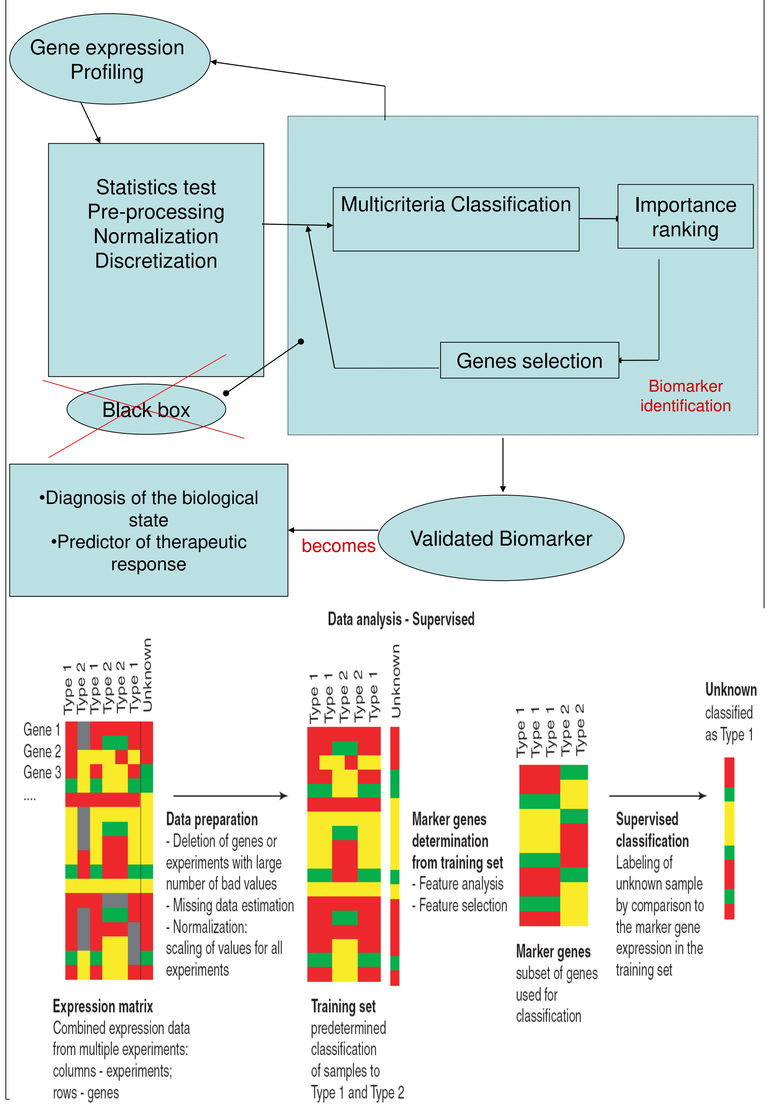
Knowledge Discovery Group, Institute for Information Technology, National Research Council of Canada, Moncton, NB, Canada

The successful treatment of cancer depends on early and accurate detection. Molecular diagnosis is expected to greatly improve current diagnostic processes. Biomarkers are molecular signatures that can be used to identify the presence or absence of a particular disease. The high throughput experimental methods, including microarrays, provide an excellent tool for parallel measurement of expression of biological molecules such as genes, proteins and metabolites. Though many methods already exist for the determination of markers and tumor diagnosis using high throughput data, more precise and accurate methods for feature selection as well as tumor classification are still needed. In this poster we present a general overview of a new strategy for gene expression-based biomarker discovery. By applying our strategy to data from tissue samples, we were able to identify the patterns of gene expression -biomarkers- unique to prostate and colon. The prostate biomarker panel consists of 8 genes. This panel was able to detect prostate cancer with over 96% specificity and sensitivity in leave-one-out cross validation. The colon biomarker panel consists of 17 genes. This panel was able to detect colon cancer with over 95% specificity and sensitivity in leave-one-out cross validation. Independent validation on clinical samples confirmed the discriminatory power of this gene panel as well, yielding over 95% accuracy of diagnosis. Independent validation on clinical samples confirmed, once again, the discriminatory power of this gene panel, yielding over 95% accuracy of diagnosis for both colon and prostate. We are currently seeking partners for further validation, and licensing of these technologies, and partners interested in collaboration leading to discovery of biomarker panels for other applications using our intelligent decision support system for biomarker discovery.

Biomarker discovery strategy

The discovery of new biomarkers is often carried out by comparing physiological changes between normal and disease states. The physiological and biochemical conditions in normal and disease states result in differential gene expression profiles, protein expression profiles and changes in metabolite profiles. After examining up-regulated and down-regulated genes, proteins and metabolites, we can identify new indicators and genetic patterns associated with particular diseases that can serve as biomarkers.

For successful implementation of a biomarker discovery strategy it is critical to access experimental and information technology expertise (bioinformatics, data mining, statistics, optimization,…). Several steps including data collection, analysis, validation and visualization are necessary to identify biomarkers from different omics platforms. We have developed a strategy for biomarkers discovery using new intelligent methods for the automated diagnosis, treatment and prognosis of different diseases from any omics data that become available.



References:

Nabil Belacel, Miroslava Cuperlovic-Culf and Rodney J. Ouellette. Molecular Method for Diagnosis of Prostate Cancer. (US Patent WO 2007/030919 A1)

Nabil Belacel, Miroslava Cuperlovic-Culf and Rodney J. Ouellette. Molecular Method for Diagnosis of Colon Cancer. (US Patent pending)

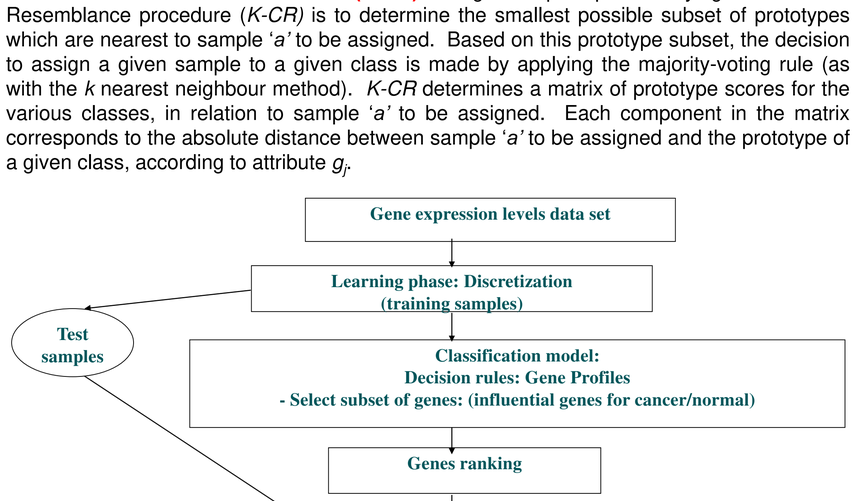
Cuperlovic-Culf, M; Belacel, N; Davey, M.; Ouellette, R.J. *Biomarkers Journal,* 15:693-706 (2010).

Nabil Belacel: The k Closest Resemblance Approach for Multiple Criteria Classification Problems. MCO 2004: 525-534

Nabil Belacel: Multicriteria assignment method PROAFTN: Methodology and medical application. *European Journal of Operational Research 125(1)*: 175-183 (2000)

Biomarker discovery strategy

The different steps of our Biomarker discovery strategy are summarized as follows:



- Pre-processing: Biomedical data are notoriously complex and variable. The data produced by any high throughput technology contains huge amount of information but is often affected by errors and noise due to sample preparation and instrument approximation. The goal of preprocessing methods is to clean up the data so that the techniques for biomarker determination such as data mining will be able to tease out key information and enhance the sensitivity and the specificity in diseases diagnosis. The preprocessing allows us to: (i) reduce noise, (ii) reduce the amount of data, and (iii) make the different omics attributes comparable.

- Biomarker determination: Once the omics data is preprocessed and qualified, a multistep process can begin to mine the data for expression changes and patterns that are related to biomarkers. The process of mining the omics data frequently leads to the identification of relatively small sets of genes, proteins or metabolites that are particularly relevant to the diagnosis, treatment or prognosis, called biomarkers. To determine these markers we apply data mining techniques to analyze and select a subset of features from different available omics data. In this step we identify key genes, markers, and signatures responsive to clinical needs using multicriteria classification method known as K-closest resemblance (K-CR) method.

- Multicriteria classification method (K-CR): The general principle underlying the K Closest Resemblance procedure (*K-CR)* is to determine the smallest possible subset of prototypes which are nearest to sample '*a'* to be assigned. Based on this prototype subset, the decision to assign a given sample to a given class is made by applying the majority-voting rule (as with the *k* nearest neighbour method). *K-CR* determines a matrix of prototype scores for the various classes, in relation to sample '*a'* to be assigned. Each component in the matrix corresponds to the absolute distance between sample '*a'* to be assigned and the prototype of a given class, according to attribute g_j.

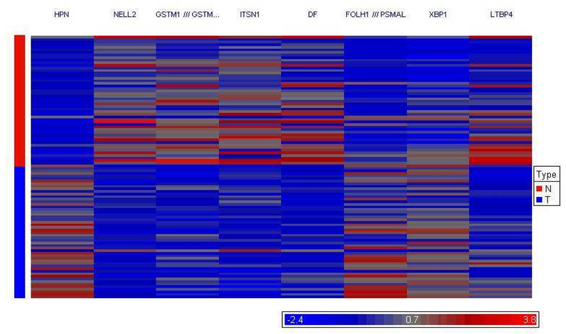
- Bibliomics / In silico experiment: Several different levels of in silico validation will be performed. First is testing of the power of determined biomarkers in independent data sets. Second is the analysis of biological knowledge about the markers from publications and databases. Finally we will investigate connections between proposed markers and other biological molecules through the analysis of biological pathways (Pathway Assist Studio pathway analysis software or similar) and networks (Cytoscape and MetaNetworks).

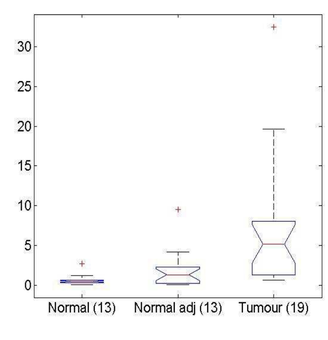
- Tissue/Biofluids Validation: Examination of panel accuracy on clinical tissue samples. Measurement of data using alternative methods (for example qPCR instead of microarrayas). Preclinical and clinical trials on noninvasive samples: urine for prostate; stool for colon cancer.

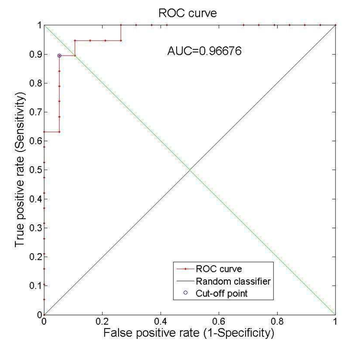
Molecular Cancer Diagnostic Applications

Biomarker Panel Prostate Cancer:

Using our strategy we have identified and validated eight gene biomarker panel that can diagnose prostate cancer with over 95% specificity and sensitivity. The application of a panel, rather than a single gene, reduces the risk that the gene expression alterations are caused by non-malignant transformations. The determined biomarker panel allows extremely accurate diagnosis on tissue biopsies samples and possibly (under validation) on urine samples. The novel approach takes advantage of the gene panels and allows reference free diagnosis directly from real-time PCR measurements. Gene expression diagnostic can be accomplished for very small samples such as core needle biopsies of cells obtained from urine samples.

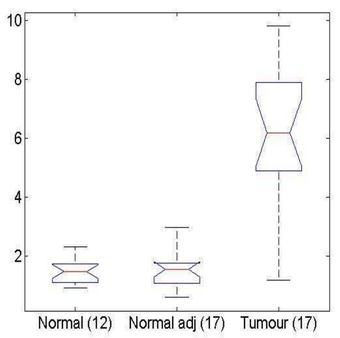


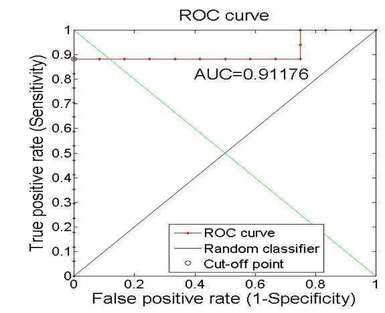




Biomarker Panel Colon Cancer:

The same strategy have been applied to colon cancer and a panel of 17 specific marker human genes is discovered and validated. The overexpression of some of these marker genes compared to their expression in normal colon tissue and the underexpression of the rest of these marker genes compared to normal colon tissue are indicative of cancerous colon tissue. By using these 17 marker genes as a diagnostic tool, small tissue samples, such as those obtained by core needle biopsies and from stool samples can be used.





Conclusions:

The National Research Council Institute for Information Technology (NRC-ITT) and the Atlantic Cancer Research Institute (ACRI) both based in Moncton, New Brunswick, Canada, have identified sets of biomarkers to be used in the detection of cancer. One set of 17 biomarkers is for the detection of colon cancer and the second set of 8 biomarkers is for the detection of prostate cancer. Although there are many biomarkers in the scientific literature for these fatal illnesses, almost all lack the sensitivity and specificity requisite for diagnosis. We have used a panel approach, with the rationale that applying a number of markers as a panel can provide more information and/or more accuracy than any single marker as a diagnostic, prognostic or therapeutic aid. Using novel methods and proprietary algorithms, two sets of biomarkers have been identified and validated for early detection of colon and prostate cancers. These panels which are patented or patent pending are available for licensing, and/or partnering.

Bio-IT World Conference & Expo, April 2011