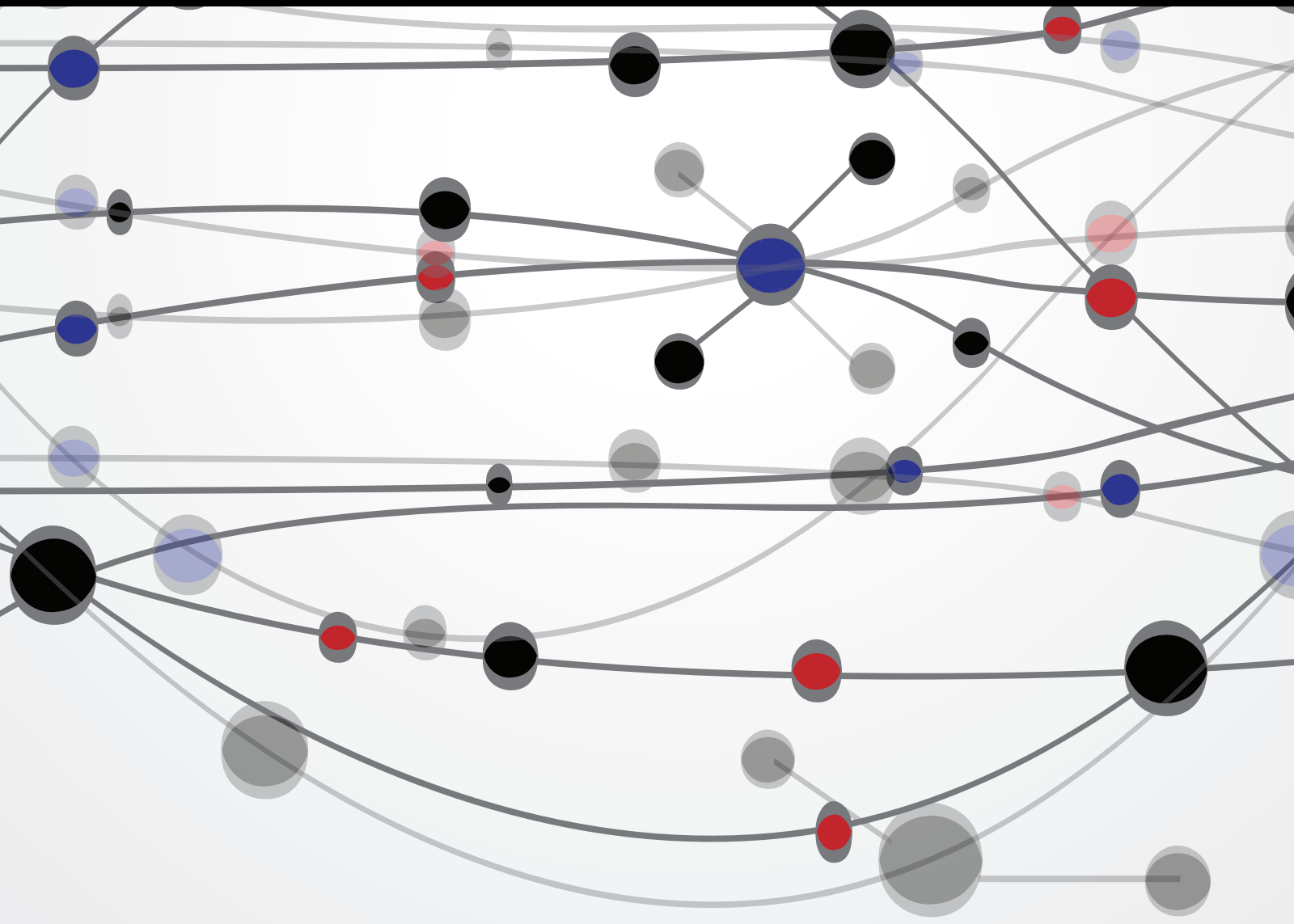
The Scientific World Journal

Bioinformatics and Biomedical Informatics

Guest Editors: Kayvan Najarian, Rachid Deriche, Mark A. Kon, and Nina S. T. Hirata



Bioinformatics and Biomedical Informatics

The Scientific World Journal

Bioinformatics and Biomedical Informatics

Guest Editors: Kayvan Najarian, Rachid Deriche,
Mark A. Kon, and Nina S. T. Hirata

Contents

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 591976, 1 page
http://dx.doi.org/10.1155/2013/591976

Editorial
Bioinformatics and Biomedical Informatics

Kayvan Najarian,[1] Rachid Deriche,[2] Mark A. Kon,[3] and Nina S. T. Hirata[4]

[1] *Department of Computer Science, Virginia Commonwealth University, 401 West Main Street, P.O. Box 843019, Richmond, VA 23284, USA*
[2] *Inria Sophia Antipolis-Méditerranée Research Center, 2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France*
[3] *Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215, USA*
[4] *Department of Computer Science, Institute of Mathematics and Statistics, Rua do Matão, 1010 05508-090 São Paulo, SP, Brazil*

Correspondence should be addressed to Kayvan Najarian; knajarian@vcu.edu

Received 15 May 2013; Accepted 15 May 2013

New and novel high-throughput systems for measurement of biological and physiological data have created a challenge: processing and analyzing the abundant resulting datastreams. There is a need for advanced computational methods to process these often very large datasets, and more generally to form decision-support systems for complex problems in medicine, biology, and related fields. During the last few decades, the vital role of these computational methods, in particular in emerging fields of life sciences such as translational medicine, has been further recognized. The need for more efficient computational methods has highlighted their roles as fundamental elements of almost any endeavor in today's life sciences. The current focus of these fields is on designing computational methods that facilitate the design and development of systems for clinical applications.

This special issue serves as a brief update to the current status of and advances in methods of biomedical informatics and bioinformatics. The computational approaches presented span a number of methods, from newly developed algorithms in DNA microarray analysis to biomedical signal/image processing techniques. The particular diseases as well as remedies addressed, from novel components of schizophrenia to new diagnoses of breast cancer, address major health issues that any society today is likely to be facing.

The paper by M. Logotheti et al. provides a comparative genomic study in patients suffering from schizophrenic and bipolar disorders using microarray expression profiling meta-analysis. A. Belle et al. present a survey of biomedical informatics, methods, and applications, as applied to computer-aided decision support systems. The paper by J. H. Phan et al. presents a meta-analysis-based feature selection method to combine multiple microarray datasets and improve reproducibility in identification of informative genes and subsequent clinical prediction. Finally, M. Burton et al. provide a cross-study comparison of classification methods of gene expression profiles for predicting metastasis in breast cancer.

With the current burst of new biomedical measurement systems (in particular portable monitoring devices) among many new data sources, there are now petabyte and higher capacity biomedical databases requiring processing via evermore efficient computational methods. Work on their designs in bioinformatics and biomedical informatics will accelerate further their quality and capacity, as imaging and monitoring technologies produce ever larger quantities of more detailed and more informative data. These works are continuing the advance of the above technologies and their frontiers.

Kayvan Najarian
Rachid Deriche
Mark A. Kon
Nina S. T. Hirata

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 896056, 10 pages
http://dx.doi.org/10.1155/2013/896056

Research Article

A Hierarchical Method for Removal of Baseline Drift from Biomedical Signals: Application in ECG Analysis

Yurong Luo,[1] Rosalyn H. Hargraves,[2] Ashwin Belle,[1] Ou Bai,[3] Xuguang Qi,[1] Kevin R. Ward,[4] Michael Paul Pfaffenberger,[1] and Kayvan Najarian[1]

[1] *Department of Computer Science, School of Engineering, Virginia Commonwealth University, 401 West Main Street, Richmond, VA 23284, USA*
[2] *Department of Electrical and Computer Engineering, School of Engineering, Virginia Commonwealth University, 401 West Main Street, Richmond, VA 23284, USA*
[3] *Department of Biomedical Engineering, School of Engineering, Virginia Commonwealth University, 401 West Main Street, Richmond, VA 23284, USA*
[4] *Department of Emergency Medicine and Michigan Critical Injury and Illness Research Center, University of Michigan, Ann Arbor, MI 48109, USA*

Correspondence should be addressed to Ashwin Belle; bellea@vcu.edu

Received 12 February 2013; Accepted 9 April 2013

Academic Editors: G. Koch, J. Ma, and V. Positano

Noise can compromise the extraction of some fundamental and important features from biomedical signals and hence prohibit accurate analysis of these signals. Baseline wander in electrocardiogram (ECG) signals is one such example, which can be caused by factors such as respiration, variations in electrode impedance, and excessive body movements. Unless baseline wander is effectively removed, the accuracy of any feature extracted from the ECG, such as timing and duration of the ST-segment, is compromised. This paper approaches this filtering task from a novel standpoint by assuming that the ECG baseline wander comes from an independent and unknown source. The technique utilizes a hierarchical method including a blind source separation (BSS) step, in particular independent component analysis, to eliminate the effect of the baseline wander. We examine the specifics of the components causing the baseline wander and the factors that affect the separation process. Experimental results reveal the superiority of the proposed algorithm in removing the baseline wander.

1. Introduction

The electrocardiogram (ECG) is an important physiological signal that helps determine the state of the cardiovascular system; however, this signal is often corrupted by interfering noise. Baseline wander is a commonly seen noise in ECG recordings and can be caused by respiration, changes in electrode impedance, and motion. Baseline wander can mask important information from the ECG, and if it is not properly removed, crucial diagnostic information contained in the ECG will be lost or corrupted. Therefore, it is vital to effectively eliminate baseline wander before any further processing of ECG such as feature extraction.

The simplest method of baseline wander (drift) removal is the use of a high-pass filter that blocks the drift and passes all main components of ECG though the filter. The main components of ECG include the P-wave, QRS-complex, and T-wave. Specifically, the PR-Segment, ST-Segment, PR-Interval, and QT-Interval are considered as the main segments of the ECG. Each of these intervals/segments has its corresponding frequency components, and according to the American Health Association (AHA), the lowest frequency component in the ECG signal is at about 0.05 Hz [1]. However, a complete baseline removal requires that the cut-off frequency of the high-pass filter be set higher than the lowest frequency in the ECG; otherwise some of the baseline drift will pass through the filter. The frequency of the baseline wander high-pass filter is usually set slightly below 0.5 Hz. Therefore, knowing that the actual ECG signal has components between 0.05 Hz and 0.5 Hz, the forementioned simple approach for baseline

removal distorts and deforms the ECG signal. In particular, it affects the ST-segment that has very low frequency components. Furthermore, ectopic beats occurring in the ECG during the course of different types of diseases and injuries change the frequency spectrum of both the baseline wander and the ECG waveforms. All the above-mentioned characteristics demand a more comprehensive approach that works for a wider range of applications and avoids distorting the main ECG waves when removing the baseline drift.

Digital filters are commonly employed method to eliminate baseline wander. Cut-off frequency and phase response characteristics are two main factors considered in the majority of these designs. The use of linear phase filters prevents the issue of phase distortion [2]. For finite impulse response (FIR) filters, it is rather straightforward to achieve linear phase response directly. Feed-forward and feed-back technologies such as infinite impulse response (IIR) filters can also provide minimum phase distortion [3]. In all of these methods, the cut-off frequency should be chosen so that the information in the ECG signals remains undistorted while the baseline wander is removed, which results in a trade-off. Usually, the cut-off frequency is set according to the slowest detected (or assumed) heart rate. However, if there are ectopic beats in the ECG signal, it is even more difficult to find this particular frequency. It is a prevalent phenomenon that the overlap between the baseline wander and low frequency components of the ECG compromises the accuracy of the extracted features.

Time-variant filters are designed to increase flexibility in the adjustment and control of the cut-off frequency. In such methods, the cut-off frequency of the filter is controlled by the low frequency characteristics of the ECG signal [4]. Cubic spline curve fitting [5], linear spline curve fitting [6], and nonlinear spline curve fitting [7] belong to another family of filters that remove the baseline wander but often require some reference points. For instance, the linear spline curve fitting method [5] forms a subsignal of the ECG for a single cardiac cycle starting 60 ms before the P-wave and ending 60 ms after the T-wave and fits a first order polynomial to this sub-signal after subtracting the mean of sub-signal. Multirate system wavelet transform has also been utilized for the ECG baseline wander removal. The approach using wavelet adaptive filter (WAF) [8] consists of two steps. First, a wavelet transform decomposes the ECG signal into seven frequency bands. The second step is an adaptive filter that uses the signal of the seventh lowest frequency band as the primary input and a constant as a reference unit for filtering. Another multi-rate system, empirical mode decomposition (EMD) [9], has also been adopted to eliminate the baseline wander. Compared with the wavelet technique that uses some predefined basis functions to represent a signal, EMD relies on a fully data-driven mechanism; that is, EMD does not require any a-priori known basis.

Adaptive filters as a cascade structure [10] have also been used for this application. The first step of this approach uses an adaptive notch filter to eliminate the DC component of the ECG. The second step forms a comb filter assuming that the signal is an event-related signal. Blind source separation (BSS), in particular independent component analysis (ICA) [11–13], is another choice to remove the baseline wander. As a specific type of BSS method, ICA has been extensively used in biomedical signals [14–16], such as the ECG and the EEG. It has been used as an effective method to decompose multichannel signals into fundamental components. As many more applications of ICA are being recognized, newer variations of ICA are being introduced. Standard ICA [17] (sICA) is a technique that is used to estimate source signals when several mixtures of signals are available. Both the source signals and the mixing process are unknown, and the sources are estimated only on the assumption that they are statistically independent. Comparing the formulation of the standard ICA, convolutive ICA (fICA) deems that the finite impulse response is closely associated with the mixing process, and the mixing process can be considered as a weighted and delayed mixture of sources [18, 19]. Fast and robust fixed-point ICA [20] is produced based on the idea that it is feasible to use contrast function to approximate negentropy. Through a fixed-point algorithm, the contrast function is maximized to extract latent sources with high speed. Temporally constrained ICA [21, 22] is a more flexible model to separate latent sources. By using prior knowledge or additional constraints, the targeted latent source is extracted. Moreover, there are many other forms of ICA for different applications such as topographic independent component analysis [23] and spatial and temporal independent component analysis [24].

In summary, the traditional methods are limited in either frequency delineation or reference choice, and the case of BBS in applications mentioned previously does not give sufficient evidence in noise removal. Based on these points of view, in the proposed method, a unified method utilizing an adaptive notch filter and BSS is used for baseline drift removal. Specifically, multichannel signals are constructed using a single-channel signal, and ICA is applied to the ECG. The main contributions of our work lie in combining the capabilities of adaptive filters and BSS, expanding the capabilities of the independent components for this application by customizing the ICA method towards the removal of the ECG baseline wander. Furthermore, the factors affecting the performance of the separation process are explored and improved in this paper.

The rest of the paper is organized as follows. The overall structure of the proposed method is illustrated in Section 2. The adaptive notch filter, as employed in the paper, is described in Section 3. The concepts and formulation of the ICA, the fast and robust fixed-point ICA, and the customized form are introduced in Section 4. Section 5 introduces the process of detecting the components that cause the baseline wander and verifies this process. This section also explores the factors that affect the separation of the results. Finally, Section 6 concludes the paper.

2. Method

Figure 1 shows the framework of the proposed method. As it can be seen in Figure 1, the first step of the proposed method is an adaptive notch filter, designed to form subsignals of the ECG, as described later. Next, as shown in Figure 1,

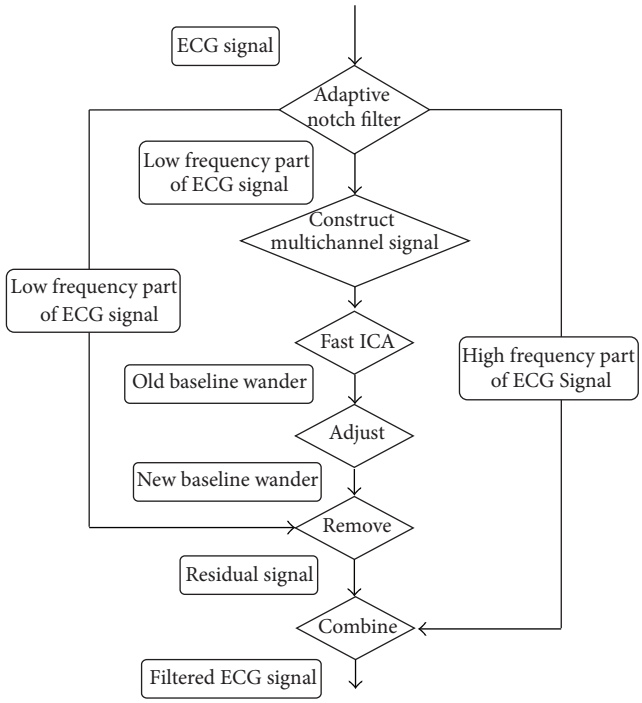


Figure 1: Schematic diagram of proposed method.

the proposed method utilizes ICA to remove the baseline drift. Considering the noisy nature of the typical raw ECG signal, in this study, subsignals in low frequencies of the ECG are formed and these filtered signals are, then, formed by an adaptive notch filter, then used as the input to the ICA algorithm. Moreover, with regard to the inputs fed to the ICA algorithm, in this study, only a single-channel ECG signal is available. Therefore, knowing that ICA requires multichannel signals to process as its input, in order to use ICA to remove baseline wander, one needs to build multichannel signals from the single-channel ECG. In order to address this issue in the proposed method, a systematic process was created in which delayed versions of the ECG are stacked to form the multi-channel signal. In addition, as shown in Figure 1, the independent component formed by the ICA as the output, which is originally labeled as the baseline wander, needs to be further adjusted to form a better estimate of the baseline wander. This is due to the fact that, while one of the components resembles the baseline drift, it is unlikely that any of the original components detected by the ICA is "purely" the baseline wander.

The specific steps shown in Figure 1 are further described below.

(a) Form sub-signals of ECG using an adaptive notch filter: as shown in Figure 1, the adaptive notch filter [25, 26] is designed and customized to form the sub-signal. The reason for using the adaptive notch filter is its flexibility as well as its relatively superior performance compared with other filters. As mentioned above, applying the ICA algorithm on a sub-signal of the ECG has the advantage of reducing the errors coming from multi-channel signals in estimating the baseline wander.

(b) Construct multi-channel signals: applying ICA requires that the signals are multi-channel ones. However, in many ECG processing applications only the single-channel ECG signal is available and/or processed. The proposed method applies the methodology in [11] to construct multi-channel signals by delaying the single-channel signal. In our study, the multi-channel signals are constructed using sixty signals, which are delayed 10 sample points (~83 ms) of the original signal in succession.

(c) Adjust the baseline wander extracted by ICA: the baseline wander extracted by ICA is an approximation of the true baseline wander because (1) there will be some errors in the resulting component due to the fact that the estimation process used in the ICA (in particular in the first few attempts) may be nonoptimal; (2) in the ICA analysis there may be more than one maximum in the estimation function and, therefore, the true baseline wander may not be located accurately; (3) the constructed multi-channel signals cannot convey all information about the baseline wander and, as such, the proposed process may alleviate the issues associated with the non-optimal construction of multi-channel signals. The 10-sample shift of the signals provides large enough variations between the multisignal component to alleviate the issues concerning dependencies for ICA processing.

3. Adaptive Notch Filter

The adaptive notch filter [26] is based on the same theoretical foundations as adaptive noise cancellation [25]. There are two inputs in the structure of the adaptive noise cancelling. One is the primary input, containing the signal and the noise and the other one is the reference input, which is the reference signal related to the noise in the primary input. Using least mean square (LMS) criterion, the reference signal is gradually approached to the noise in the primary input. When the stability is achieved, the output is acquired through subtracting the reference input from the primary input. This type of filter can deal with inputs that are deterministic or stochastic, stationary or time-variant. If the inputs are stationary stochastic, the solution of the adaptive noise cancelling approaches closely Wiener filter [25]. As to the adaptive notch filter, the reference signal is the signal with one- or multifixed frequencies, which are treated as the frequencies to be excluded.

The advantages of adaptive notch filters lie in the following aspects: (1) if the frequency of the interference is not precisely known or the interference drifts in the frequency, the exact excluded frequency could be measured/adapted during the filtering process; (2) the filter is tunable since the null point moves with the reference frequencies; (3) the adaptive notch filter can be made very sharp at the reference frequency; (4) through adjusting the parameters, the adaptive notch filter can be considered as a time-invariant filter by

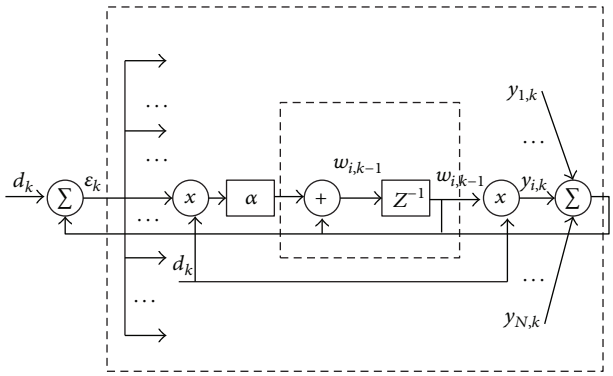


Figure 2: The diagram of adaptive noise cancelling.

lessening the influence of the time-varying components. The inference of adaptive notch filter is described in [25, 26]. The diagram of adaptive noise cancelling is shown in Figure 2. The system is an N-stage tapped delay line (TDL). The weight of the filter is updated according to the following equations:

$$\begin{aligned} y_k &= w_k^T x_k, \\ \varepsilon_k &= d_k - y_k, \\ w_{k+1} &= w_k + \partial \varepsilon_k x_k, \end{aligned} \tag{1}$$

where x is the reference input, d is the desired response, y is the output of the filter, w is the weight of the filter, ∂ is the adaptation constant, and k is the time index. As described in [26], the response from $E(z)$ to $Y(z)$ includes two parts. In practical applications, it is feasible to make the time-varying component to be insignificant ($\beta/N \approx 0$) by changing the values of N and setting β as follows:

$$\beta = \frac{\sin\left(N w_r T\right)}{\sin\left(w_r T\right)}, \tag{2}$$

where w_r is the frequency of the interference. If the reference input is considered to be the following form:

$$x = C\cos\left(w_r T + \theta\right), \tag{3}$$

the transfer function of adaptive notch filter can be expressed as follows:

$$H(z) = \frac{z^2 - 2z\cos\left(w_r T\right) + 1}{z^2 - 2\left(1 - N\partial C^2/4\right) z\cos\left(w_r T\right) + \left(1 - N\partial C^2/2\right)}. \tag{4}$$

Therefore, the parameter N can be set to the fixed value as described above. It can be seen that the above-mentioned filter is very flexible and can be adjusted using the adaption constants ∂ and C to provide the desired bandwidth and depth of a suitable notch filter.

4. Independent Component Analysis

After applying the notch filter, the main step used is ICA. First, the "standard" ICA is described. ICA can be briefly explained using a simple example of separating two source signals $s_1(t)$ and $s_2(t)$ that were mixed by an unknown linear process. Two different linear mixtures, $x_1(t)$ and $x_2(t)$, are given as follows:

$$\begin{aligned} x_1(t) &= c_{11}s_1 + c_{12}s_2, \\ x_2(t) &= c_{21}s_1 + c_{22}s_2, \end{aligned} \tag{5}$$

where c_{11}, c_{12}, c_{21}, and c_{22} are unknown coefficients. The objective of the problem is to recover the signal $s_1(t)$ and $s_2(t)$ from mixture signals $x_1(t)$ and $x_2(t)$ without knowing any prior information about the source signals $s_1(t)$ and $s_2(t)$ and the mixing process (i.e., c_{11}, c_{12}, c_{21}, and c_{22}), except that $s_1(t)$ and $s_2(t)$ are statistically independent.

In the generalized case, where there are more latent sources and more mixture of signals, the formal definition of ICA is as follows:

$$x_i(t) = c_{i1}s_1 + c_{i2}s_2 + \cdots + c_{in}s_n, \quad i \in [1, n], \tag{6}$$

where $s_i(t)$ is called latent source, $x_i(t)$ is the mixture signal, c_{ij} is the mixing coefficient between $x_i(t)$ and $s_j(t)$, and n is the number of latent sources and mixture signals. The above formulation can be expressed as the following matrix form:

$$X = C_{n\times n} \cdot S, \tag{7}$$

where X is the matrix of mixture signals, in which each column is one mixture signal; S is the matrix of latent signals, in which each column is one latent signal; and $C_{n\times n}$ is the matrix for mixing coefficients.

The feasibility of solving the ICA problem lies in the condition that the latent sources are independent of each other. According to the Central Limit Theorem, the distribution of a sum of independent random variables approaches a Gaussian distribution. This implies that the solution of ICA can be achieved when distribution diverges from Gaussianity. The deviation from Gaussianity can be determined using measures such as Negentropy.

Negentropy is one measure of non-gaussianity defined based on the concept of entropy, which is the fundamental concept of information theory. Entropy, E, as a measure of information in random variables is defined for a discrete random variable y as folows:

$$E(y) = -\sum_i P\left(y = c_i\right) \log P\left(y = c_i\right), \tag{8}$$

where c_i is the possible values of Y and $P(Y = c_i)$ means the probability when the value of Y is c_i. For a continuous random variable y, entropy E is defined as the following equation:

$$E(y) = -\int f(y) \log\left(f(y)\right) dy, \tag{9}$$

where f is the probability distribution function. Negentropy, J, is then defined as follows:

$$J(y) = E\left(y_{\text{gauss}}\right) - E(y), \tag{10}$$

where y_{gauss} is a Gaussian random variable with the same covariance matrix as y. A fundamental conclusion in information theory is that a Gaussian variable has the largest entropy among all random variables of equal variance. Hence, negentropy is always nonnegative, and it is zero only if Y has a Gaussian distribution.

The exact calculation of negentropy requires an accurate estimation of the probability distribution function, which may be computationally costly or data intensive. Therefore, it is often preferred to find simple approximations of negentropy. Simple approximations of negentropy have been introduced [27], which are based on the maximum entropy principle. In general, the following family of approximations is the most commonly used group:

$$J(y) = \sum_{i=1}^{p} k_i \left[E\left(G_i(y)\right) - E\left(G(v)\right)\right]^2, \tag{11}$$

where k_i are constants and v is a gaussian random variable with zero mean and unit variance. Often, the value of p and k_i can be set to one. Therefore, the above formulation becomes as follows:

$$J(y) = \left[E\left(G(y)\right) - E\left(G(v)\right)\right]^2. \tag{12}$$

The following formulations of G functions have proved very useful in practical applications:

$$\begin{aligned} G_1(y) &= \frac{1}{a_1}\log\cosh(a_1 y), \qquad g_1(y) = \tanh(a_1 y), \\ G_2(y) &= -\frac{1}{a_2}\exp\left(-\frac{a_2 y^2}{2}\right), \qquad g_2(y) = y\exp\left(-\frac{a_2 y^2}{2}\right), \\ G_3(y) &= \frac{1}{4}y^4, \qquad g_3(y) = y^3, \end{aligned} \tag{13}$$

where $1 \le a_1 \le 2$, $a_2 \approx 1$, and g is the first derivative of the function G.

Before applying the main processing operations of the ICA, it is often necessary to perform some preprocessing. Usually, the two different operations are conducted: centering and whitening. Centering requires that the random variable y is a zero-mean random variable, and it is performed by subtracting its mean vector. Whitening will make the random variable uncorrelated and set their variances equal to unity by using the eigenvalue decomposition of their covariance matrix:

$$E\left\{yy^T\right\} = DVD^T, \tag{14}$$

where D is the orthogonal matrix of eigenvectors and V is the diagonal matrix of eigenvalues. Now, assuming that z is a new random variable after whitening, consider the following:

$$z = DV^{-1/2}D^T y. \tag{15}$$

Whitening makes the problem change from estimating mixing matrix to estimating a new one $\widetilde{C}$:

$$z = DV^{1/2}D^T Cs = \widetilde{C}s. \tag{16}$$

Among several improvements of ICA, fast and fixed-point independent component analysis [20], as a direct extension of the standard ICA, was developed for calculating latent sources with high speed. The basic rule of fast and fixed-point independent component analysis is to find a direction, which can maximize non-Gaussianity of $w^T x$. Non-Gaussianity is decided according to the approximation of nongaussianity as mentioned above. The following is the basic description of the algorithm.

(a) Initialize a weight vector w in one direction.

(b) Change the weight vector according to the following criteria: $w' = E\{xg(w^T x)\} - E\{g'(w^T x)\}w$, and normalize the weight vector as $w = w'/\|w'\|$.

(c) If the weights have not converged, go back to step (b),

where w is the weight vector to calculate latent source $s = w^T x$ and convergence means that the old weight vector and the new weight vector are in the same direction.

In this study, the fast and fixed-point independent component analysis [20] is used as the implementation of ICA block shown in Figure 1.

5. Results

An ECG dataset of human volunteer undergoing lower body negative pressure (LBNP) [28] as a surrogate of hemorrhage was employed to verify the effectiveness of removing baseline wander. This data set was created under Institutional Review Board approval. The LBNP dataset consisted of a total of 91 subjects. Each subject had a single vector lead ECG recording collected at the sampling rate of 500 Hz. The baseline wander in ECG signals demonstrated significant level of variations in the amplitude over the course of the LBNP experiment. During LBNP, subjects are exposed to increasing negative pressure to their lower bodies. This causes a redistribution of blood volume to the lower extremities and abdomen causing a decrease in blood pressure and cardiac output and resulting in an increased respiratory rate.

The results of the proposed method are compared with a reference method, called robust locally weighted regression [29], which is often treated as one of the most robust and commonly used methods to remove baseline drift. The robust locally weighted regression method employs two techniques: the local fitting of polynomials and an adaptation of iterated weighted least squares to remove the baseline drift.

5.1. Results of Adaptive Notch Filter. One objective of the proposed system is the removal of unwanted frequencies around 0 Hz as well as 60 Hz. As the frequencies around zero are excluded, the filter acts as a high-pass filter. In order to lessen the influence of the time-varying components, one needs to first set a suitable parameter N to obtain a desirable level of time-varying component, β/N. Figure 3 shows the value of the time-varying component β/N for different values of N.

Figure 3 indicates that the value of N determines the degree at which the time varying component influences

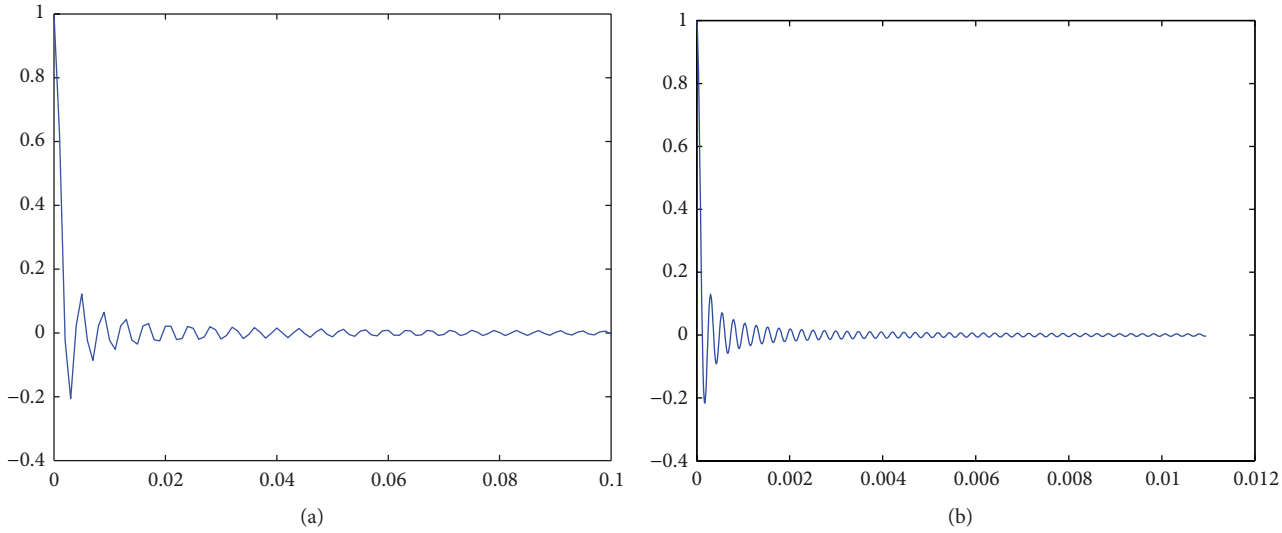


Figure 3: The resulting value of β/N (a) $N = 256$; (b) $N = 4096$.

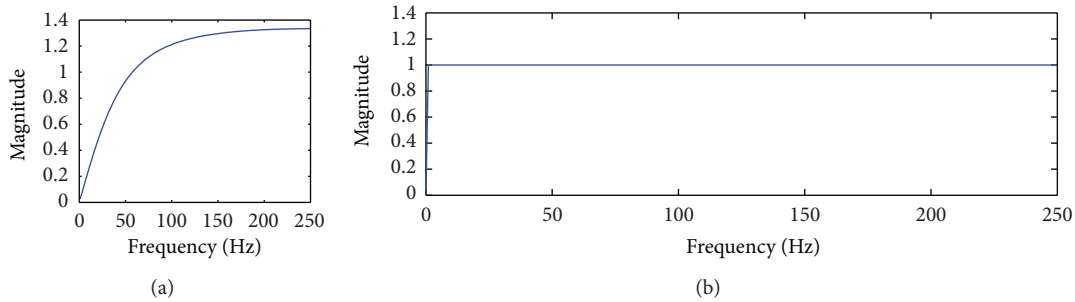


Figure 4: Transfer function for two choices of adaptive notch filters (a) $C = 1$; (b) $C = 0.01$.

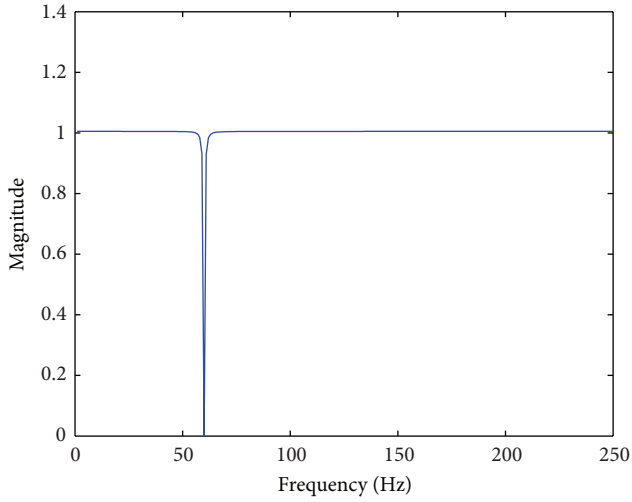


Figure 5: Transfer function of the adaptive notch filter around 60 Hz.

the filter. In general, with the increase in the value of N, this influence decreases gradually. In this study, the value of N was set to 10,000. The parameter ∂ identifies whether or not the adaptation converges [25]. The value of ∂ should be greater than 0 but less than the reciprocal of the largest eigenvalue, λ, of the matrix R, which is defined as the correlation matrix of signal [25]. In this study, the value of ∂ was set to 0.0001. The bandwidth of the filter can be approximated using the following equation [26]:

$$\mathrm{BW} = \frac{N\partial C^2}{2T} \ (\mathrm{rad/s}). \tag{17}$$

Figure 4 shows the transfer function of the resulting adaptive notch filter, and, as expected, this filter acts as a high-pass filter. Note that the value of C provides yet another degree of freedom for this filter design process, and, hence, Figure 4 presents the transfer function for two different filters formed using two different values of C, each resulting in a very different bandwidth. A main advantage of the adaptive notch filter used here is that changing the values of parameters N, ∂, and C can provide a wide spectrum of desired filters with diverse shapes of transfer function.

Adaptive notch filter for frequencies around 60 Hz is designed similarly. The parameter N was to 2048, ∂ to 0.001, and C to 0.1. Figure 5 depicts the transfer function of the resulting adaptive notch filter.

Table 1: Experimental results of removing the baseline wander.

Subject	Shift/elevation	Error_1	Error_2
1	290/0	2.0996	0.7847
2	250/1	28.1832	2.7037
3	300/4	193.9524	3.4495
4	300/1	24.3905	1.0727
5	300/3	89.1358	3.6282
6	290/2	17.9017	1.1614
7	300/1	28.955	1.0623
8	200/0	107.7542	13.4439
9	300/2	203.8138	4.0846
10	290/2	81.7942	2.2818
11	290/2	256.3747	8.7264
12	300/2	41.0977	2.4223
13	260/1	44.2238	2.279
14	260/2	101.7592	2.3317
15	310/2	700.1481	101.429
16	290/1	12.7575	1.3522
17	290/1	45.6429	2.6412
18	310/0	36.8833	11.8224
19	290/1	9.1224	1.88
20	290/2	181.3923	23.0193
21	300/2	25.4492	2.6421
22	370/4	252.4353	8.5616
23	260/2	304.7066	7.4637
24	290/1	116.9048	3.77
25	300/1	16.3922	1.05
26	290/0	3.4748	0.6671
27	300/2	144.4347	1.8579
28	290/0	23.9724	2.1641
29	290/1	14.5089	0.2205
30	290/1	155.3859	3.6707
31	300/1	50.6959	2.6757
32	300/1	27.2665	1.101
33	300/1	56.7045	1.999
34	290/2	324.4399	10.0313
35	300/1	42.6266	0.8791
36	290/2	539.7357	31.3238
37	290/1	19.8874	0.8131
38	290/1	14.3623	2.6499
39	260/1	8.8582	6.4787
40	300/4	135.0286	3.0056
41	290/1	29.5551	2.7541
42	300/1	43.3923	3.0052
43	290/0	51.0465	6.9241
44	290/0	31.9213	5.4646
45	290/1	9.7597	1.3328
46	270/1	22.7897	1.3598
47	290/2	93.5265	1.899
48	290/0	8.7422	1.5607
49	350/6	892.829	74.3034
50	300/3	209.4986	6.3436
51	300/3	60.5121	2.6645
52	290/1	3.7123	0.9486
53	290/4	247.2271	5.138
54	250/1	32.0128	2.8609
55	310/0	20.1471	1.336
56	310/0	5.2858	4.0839
57	290/0	7.1664	0.9526
58	300/1	35.4656	0.8932
59	290/1	10.9895	0.8653
60	300/3	115.7327	4.8387
61	300/1	26.7803	0.7141
62	290/2	9.3222	2.8809
63	290/1	16.9436	0.9469
64	300/0	27.7014	1.7794
65	290/1	55.1891	4.9226
66	310/6	620.3234	8.0999
67	400/2	23.6969	0.5595
68	290/2	36.63757	1.4766
69	290/2	241.5044	11.7279
70	290/1	5.5229	0.3386
71	290/2	173.1734	7.0318
72	300/2	77.4627	3.2468

5.2. Experimental Results and Problems Analysis. The results of both methods, that is, the proposed and the reference methods, are examined and compared in all 91 subjects. A unified "span" value, described in the reference method [29], which is designed to assess the quality of the methods in removing the baseline wander, is calculated for all cases. This value for all experimental results was 1500, which is the level identifying a very high quality of baseline removal.

The 91 cases, based on the closeness of the results of the two methods, are divided into two groups. The details of the results are shown for 72 out of 91 subjects in Table 1; for these subjects the proposed algorithm achieves almost identical results as the reference method. The results of the remaining 19 subjects, which will be discussed separately, show that the proposed method cannot be able to remove the baseline drift optimally.

In Table 1, "shift" and "elevation" are the values for adjustments to the original independent component (baseline wander) to form the new baseline wander in the horizontal and vertical directions; "error_1" represents the difference between the old baseline wander (sig_1) before shift and the baseline wander (sig) from the reference method calculated as follows:

$$\text{error}_1 = \frac{(\text{sig}_1 - \text{sig})^2}{n}, \tag{18}$$

where n is the number of sample points in the baseline wander, and finally "error_2" represents the difference between

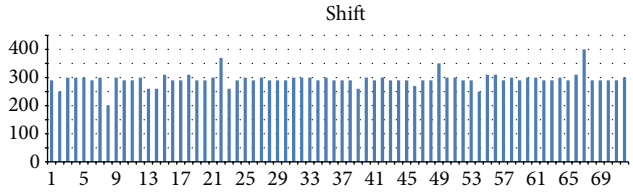


Figure 6: Value of "shift" that adjusts the old baseline wander to form the new one for all 72 subjects.

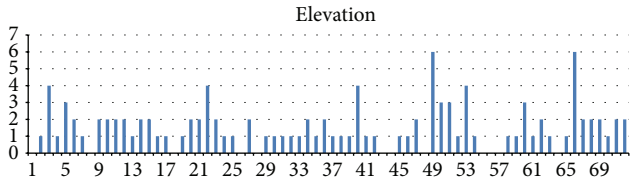


Figure 7: Value of "elevation" that adjust the old baseline wander to form the new one for all 72 subjects.

the new baseline wander (sig_2) and the baseline wander (sig) from the reference method calculated as follows:

$$\text{error}_2 = \frac{(\text{sig}_2 - \text{sig})^2}{n}. \tag{19}$$

As it can be seen in Table 1, for all cases error_2 is significantly smaller than error_1 which shows the impact of that method in "purifying" the baseline wander and creating a better estimate of the drift. In order to better assess the performance of the proposed method in removing the baseline wander, more analyses are conducted on the results.

Figures 6 and 7 show the shift and elevation for all 72 subjects. As can be seen, both of these variables are almost the same for all subjects and do not change across different subjects (x-axis) or vary in a small scope. This observation illustrates the reason to adjust the parameters between the old baseline wander and the new baseline wander.

Figure 8 shows the error reduction in 72 subjects after adjusting shift and elevation value. It can be seen that in all of these cases the errors decrease significantly after adjusting the baseline wander compared with the baseline wander. The average percentage of error reduction Aver E reaches up to 90.13%. The formulation of the average percentage of error reduction is shown in the following:

$$\begin{aligned} \text{percentage}\,(i) &= \frac{(\text{error}_1 - \text{error}_2)}{\text{error}_1}, \quad i \in [1, n], \\ \text{Aver}\,E &= \sum_{i=1}^{i=n} \text{percentage}\,(i), \end{aligned} \tag{20}$$

where i is the index of subject and n is the total number of subjects.

Sample signals before baseline removal and after baseline removal with the proposed method as well as the reference method are shown in Figure 9. As shown in Figure 9, the results of the two methods in all above-mentioned 72 subjects

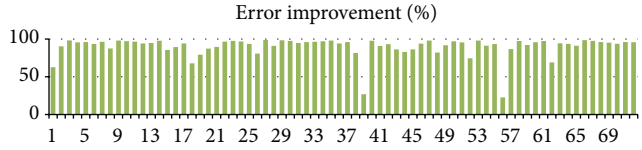


Figure 8: Improved percentages of error after adjustment.

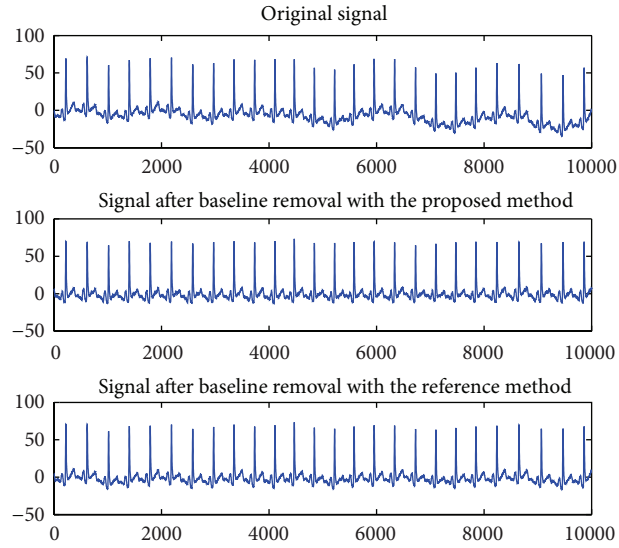


Figure 9: Comparison between the proposed method and the reference method.

are very similar. In addition, as it can be seen, both methods are very effective in removing the baseline drift.

However, as mentioned above, on the ECG of the remaining 19 subjects, the results of the proposed method and the reference method are not as similar; that is, the value of error_2 (which shows the difference between the two methods) is significant. This is because in these signals the inherent pattern observed from ECG is highly distorted hence leading to spurious estimations. As mentioned before, we have visually inspected all 91 cases. By examining the signals for these 19 cases, it was discovered that the high value of error_2 does not seem to come from the inability of the proposed method to remove the baseline wander. In such case, the possible reason and improvement are discussed in the following part.

As a comparison between the proposed method and reference method, some such sample results are shown in Figure 10. In these cases, due to the presence of significantly stronger baseline drifts, the reference method seems not to be eliminating almost all the baseline drift. The reason for this might lie in the fact that the reference method relies heavily on the parameters set that may work very well for some ECG signals but not for others. As shown in Figure 10, our proposed method shows more effective performance in removing the baseline around times such as 4700, 5500, 7500, and 9500. Another major advantage is that the proposed

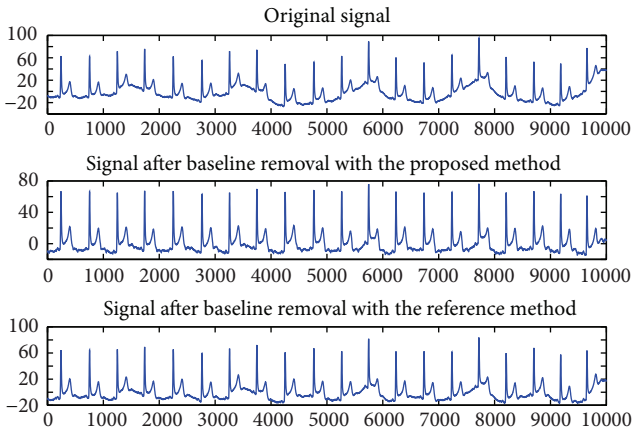


Figure 10: Comparison of the proposed method to the reference method.

method is computationally faster than the reference method while achieving the same quality of results.

5.3. Further Experimental Analysis of Method. As mentioned above, in the experiment, multi-channel signals are constructed through a single-channel signal. The multi-channel signals are constructed using sixty signals, which are 10 sample point delayed successions of the original signal. By observation, the number of the constructed signals greatly impacts the success of finding the true baseline wander. Moreover, the degree of delay has a close relationship with the smoothness of the baseline wander. Experimentally, it can be considered that more channels and smaller delayed signals may achieve better results, meaning that the constructed multi-channel signals may convey enough information in order to accurately extract the baseline wander.

In addition, as discussed above, the LBNP dataset shows a significant level of variations in the baseline drift. Therefore, in further analysis of the method, the sub-signals were segmented to verify whether the slow changes in the trend of the baseline wander affect the results of the proposed method in separating the baseline wander. The sub-signals were chosen to be only 10,000 sample points long from the beginning of the original signal in LBNP dataset. Experimental results showed that the slow changing trend of the baseline wander did not affect the performance of the proposed method in extracting the baseline wander. In other words, the baseline drift with slow changing trends can also be successfully extracted using the proposed method.

6. Conclusion

While using the blind source separation paradigm, the ECG baseline wander or drift may be removed. The present paper demonstrates a hierarchical method utilizing ICA to significantly improve the performance of this process and achieve improved performance. Compared with the existing methods, the proposed method has the following advantages. (1) The proposed method provides more flexibility with regard to parameter estimation and selection. (2) When following the steps proposed for adjustment of ICA process, the fundamental assumption of baseline noise coming from an independent source can be further verified, which supports the validity of using the method for ECG baseline removal. Such an assumption, verified by additional experimental results, would present a chance to remove other types of noise. (3) The filtering process proposed for forming the multi-channel signals provides a highly flexible method to form the input to ICA.

References

[1] L. T. Sheffield, A. Berson, D. Bragg-Remschel et al., "Recommendations for standard of instrumentation and practice in the use of ambulatory electrocardiography," *Circulation*, vol. 71, no. 3, pp. 626A–636A, 1985.

[2] J. A. van Alsté, W. van Eck, and O. E. Herrmann, "ECG baseline wander reduction using linear phase filters," *Computers and Biomedical Research*, vol. 19, no. 5, pp. 417–427, 1986.

[3] L. P. Harting, N. M. Fedotov, and C. H. Slump, "On baseline drift suppressing in ECG-recordings," in *Proceedings of the IEEE Benelux Signal Processing Symposium*, pp. 133–136, 2004.

[4] L. Sornmo, "Time-varying digital filtering of ECG baseline wander," *Medical and Biological Engineering and Computing*, vol. 31, no. 5, pp. 503–508, 1993.

[5] C. R. Meyer and H. N. Keiser, "Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques," *Computers and Biomedical Research*, vol. 10, no. 5, pp. 459–470, 1977.

[6] C. Papaloukas, D. I. Fotiadis, A. P. Liavas, A. Likas, and L. K. Michalis, "A knowledge-based technique for automated detection of ischaemic episodes in long duration electrocardiograms," *Medical and Biological Engineering and Computing*, vol. 39, no. 1, pp. 105–112, 2001.

[7] V. S. Chouhan and S. S. Mehta, "Total removal of baseline drift from ECG signal," in *Proceedings of International Conference on Computing: Theory and Applications (ICCTA '07)*, pp. 512–515, March 2007.

[8] K. L. Park, K. J. Lee, and H. R. Yoon, "Application of a wavelet adaptive filter to minimise distortion of the ST-segment," *Medical and Biological Engineering and Computing*, vol. 36, no. 5, pp. 581–586, 1998.

[9] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in Biology and Medicine*, vol. 38, no. 1, pp. 1–13, 2008.

[10] J. Raimon, P. Laguna, N. V. Thakor, and P. Caminal, "Adaptive baseline wander removal in the ECG: comparative analysis with cubic spline technique," in *Proceeding of Computers in Cardiology*, pp. 143–146, October 1992.

[11] Z. Barati and A. Ayatollahi, "Baseline wandering removal by using independent component analysis to single-channel ECG data," in *Proceedings of International Conference on Biomedical and Pharmaceutical Engineering (ICBPE '06)*, pp. 152–156, December 2006.

[12] M. P. S. Chawla, H. K. Verma, and V. Kumar, "Artifacts and noise removal in electrocardiograms using independent component analysis," *International Journal of Cardiology*, vol. 129, no. 2, pp. 278–281, 2008.

[13] M. P. S. Chawla, H. K. Verma, and V. Kumar, "Independent component analysis: a novel technique for removal of artifacts and baseline wander in ECG," in *Proceedings of the 3rd National Control Instrumentation System Conference*, pp. 14–18, 2006.

[14] M. Keralapura, M. Pourfathi, and B. Sirkeci-Mergen, "Impact of contrast functions in Fast-ICA on twin ECG separation," *IAENG International Journal of Computer Science*, vol. 38, no. 1, pp. 38–47, 2011.

[15] K. Arfanakis, D. Cordes, V. M. Haughton, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Combining independent component analysis and correlation analysis to probe interregional connectivity in fMRI task activation datasets," *Magnetic Resonance Imaging*, vol. 18, no. 8, pp. 921–930, 2000.

[16] Y. Ye, Z. L. Zhang, J. Zeng, and L. Peng, "A fast and adaptive ICA algorithm with its application to fetal electrocardiogram extraction," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 799–806, 2008.

[17] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[18] M. Milanesi, N. Vanello, V. Positano et al., "Frequency domain approach to blind source separation in ECG monitoring by wearable system," in *Proceedings of Computers in Cardiology*, pp. 767–770, September 2005.

[19] M. Milanesi, N. Vanello, V. Positano, M. F. Santarelli, and L. Landini, "Separation and identification of biomedical signals based on frequency domain independent component analysis," *WSEAS Transactions on Systems*, vol. 4, no. 10, pp. 1752–1761, 2005.

[20] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[21] C. J. James and O. J. Gibson, "Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 9, pp. 1108–1116, 2003.

[22] J. Lee, K. L. Park, and K. J. Lee, "Temporally constrained ICA-based foetal ECG separation," *Electronics Letters*, vol. 41, no. 21, pp. 1158–1160, 2005.

[23] A. Hyvärinen, P. O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Computation*, vol. 13, no. 7, pp. 1527–1558, 2001.

[24] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human Brain Mapping*, vol. 13, no. 1, pp. 43–53, 2001.

[25] B. Widrow, J. R. Glover, J. M. McCool et al., "Adaptive noise cancelling: principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

[26] J. R. Glover, "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 6, pp. 484–491, 1977.

[27] A. Hyvarinen, "New approximations of differential entropy for independent component analysis and projection pursuit," in *Advances in Neural Information Processing Systems*, vol. 10, pp. 273–279, 1998.

[28] S. Y. Ji, A. Belle, K. R. Ward et al., "Heart rate variability analysis during central hypovolemia using wavelet transformation," *Journal of Clinical Monitoring and Computing*, pp. 1–14, 2013.

[29] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Assocaition*, vol. 74, pp. 829–836, 1979.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 685917, 14 pages
http://dx.doi.org/10.1155/2013/685917

Research Article

A Comparative Genomic Study in Schizophrenic and in Bipolar Disorder Patients, Based on Microarray Expression Profiling Meta-Analysis

Marianthi Logotheti,[1,2,3] Olga Papadodima,[2] Nikolaos Venizelos,[1] Aristotelis Chatziioannou,[2] and Fragiskos Kolisis[3]

[1] *Neuropsychiatric Research Laboratory, Department of Clinical Medicine, Örebro University, 701 82 Örebro, Sweden*
[2] *Metabolic Engineering and Bioinformatics Program, Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, 48 Vassileos Constantinou Avenue, 11635 Athens, Greece*
[3] *Laboratory of Biotechnology, School of Chemical Engineering, National Technical University of Athens, 15780 Athens, Greece*

Correspondence should be addressed to
Aristotelis Chatziioannou; achatzi@eie.gr and Fragiskos Kolisis; kolisis@chemeng.ntua.gr

Received 2 November 2012; Accepted 27 November 2012

Academic Editors: N. S. T. Hirata, M. A. Kon, and K. Najarian

Schizophrenia affecting almost 1% and bipolar disorder affecting almost 3%–5% of the global population constitute two severe mental disorders. The catecholaminergic and the serotonergic pathways have been proved to play an important role in the development of schizophrenia, bipolar disorder, and other related psychiatric disorders. The aim of the study was to perform and interpret the results of a comparative genomic profiling study in schizophrenic patients as well as in healthy controls and in patients with bipolar disorder and try to relate and integrate our results with an aberrant amino acid transport through cell membranes. In particular we have focused on genes and mechanisms involved in amino acid transport through cell membranes from whole genome expression profiling data. We performed bioinformatic analysis on raw data derived from four different published studies. In two studies postmortem samples from prefrontal cortices, derived from patients with bipolar disorder, schizophrenia, and control subjects, have been used. In another study we used samples from postmortem orbitofrontal cortex of bipolar subjects while the final study was performed based on raw data from a gene expression profiling dataset in the postmortem superior temporal cortex of schizophrenics. The data were downloaded from NCBI's GEO datasets.

1. Introduction

Schizophrenia (SZ) and bipolar disorder (BD) are approached and studied as diseases with aberrant functions of the neurotransmitter systems, as neurodevelopmental diseases or generally complex diseases caused by multiple genetic and environmental factors. Recently they have started to be studied as systemic diseases; thus a combination of disturbed biological systems and genes of small contribution is believed to cause their expression [1, 2].

Altered membrane composition of the cells, aberrant membrane phospholipid metabolism [3, 4], dysfunctional tyrosine, and other amino acid (AA) transport systems [5–11] evidence the systemic nature of SZ disease. Moreover, failure of niacin skin test implying reduced arachidonic acid (ARA) in cell membranes of schizophrenics [12] and abnormalities in muscle fibers [13] constitute such indications. The same holds for BD, which can also be considered a systemic disease. Aberrant tyrosine, and other AA transport systems, in cells from BD disorder patients [14, 15], aberrant signal transduction [16], and abnormal membrane composition and metabolism support the notion of BD being a systemic disease as well [17, 18].

Studying these disorders through this holistic approach, we presume the membrane phospholipid hypothesis, namely, that aberrant AA transport mechanisms and the disturbed

cell membrane composition are highly correlated. AAs are transported though cell membranes with specific transporter/protein transport systems, which perform active transport of AAs from one side of the cell membrane to the other [19]. These AA transporters are embedded in the cell membranes; thus their structure and functionality interact with the membrane composition and functionality, as well as with membrane fluidity and enzymatic activity [9, 20]. Particularly, a membrane defect would impact, for example, the functionality of the tyrosine transporters as well as the permeability of the membranes [2, 5].

The Membrane Theory. The membrane theory of mental diseases is related with two primary abnormalities: an increased rate of removal of essential fatty acids (EFA) from the membrane phospholipids, combined with a reduced rate of incorporation of fatty acids (FA) into membrane phospholipids [21]. Some SZ study findings that relate the expression of the disease with the membrane hypothesis are studies based on postmortem and blood samples showing reduction of docosahexaenoic acid (DHA) and ARA in cell membranes independently of the disease state and magnetic resonance spectroscopy (MRS) studies revealing decreased levels of phosphomonoesters (phospholipid membrane synthesis precursors) and higher levels of phosphodiesters (phospholipid metabolism products) in SZ patients compared to control patients [22]. Also, the niacin skin flush test is indicative of a membrane dysfunction resulting in an inflammatory dysfunction [12]. In addition, phospholipase A2 (PLA) calcium (Ca) dependent type has been shown to have an increased activity and PLA Ca independent type a decreased activity. The latter is considered quite important finding, as the A2 enzyme catalyzes the breakdown of FA [23].

Similar findings suggest cell membrane dysfunction in BD. 31P-MRS magnetic resonance spectroscopy (MRS) measures phosphorus metabolites in the organs [24]. Phosphomonoester levels are measured in BD depressed patients with MRS. Phosphomonoesters are measured as being higher in these patients compared to control subjects and lower in asymptomatic patients. Abnormal functionalities in signal transduction pathways are also repeated in several studies including overactivated phosphatidylinositol and G-protein pathways, as well as altered membrane protein kinase C and adenyl cyclase enzyme pathways. PLA enzyme activity and Ca release are involved in the membrane hypothesis of BD [17].

Amino Acid Transporters. The transport of AAs into the cell membranes of the blood brain barrier (BBB) is mediated by many transport systems. Three basic active transporters result in the AA flux from and into all types of cells (including brain cells). The primary active transport mechanism is an adenosine triphosphatase (ATPase) that exchanges sodium (Na) and potassium (K) ions, contributing in the maintenance of the ion gradients of the cells, known as sodium-potassium adenosine triphosphatase (Na,K-ATPase). These ion gradients in combination with other ions and gradients are utilized by the secondary active transport mechanisms for the influx of specific AAs into the cells. The secondary active transport through these AA influxes sets also an AA concentration gradient in the cells, which, in combination with Na+ exchange, is further utilized by the tertiary active transport mechanisms for transport of another group of AAs in and out of the cells. AAs may be transported via different AA transport mechanisms. An alteration in any of the active transport mechanisms could result in an aberrant AA transport into the cells [10, 25].

Aim of the Study. The aim of our meta-analysis was to interpret the results of comparative genomic profiling studies in schizophrenic patients as compared to healthy controls and in patients with BD and try to relate and integrate our results with an aberrant AA transport through cell membranes.

2. Materials and Methods

2.1. Microarray Datasets. Four human datasets were used, by downloading submitted raw data (Cel files) from corresponding studies, available at the Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI) [26].

(1) The first study has the GEO Accession number GSE12654 and the microarrays preparation followed the guidelines of MIAME in the way it is described in [27]. RNA from postmortem brain tissues (Brodmann's Area 10) of 15 schizophrenic and 15 BD affected patients and 15 control healthy subjects was hybridized on Affymetrix HG-U95 Arrays. After quality control stage in this study, 11 schizophrenic, 11 BD and 15 control subjects were used for further bioinformatic analysis.

(2) The second study has the GEO Accession number GSE5389, and the microarrays preparation followed the guidelines of MIAME in the way it is described in [28]. RNA extracted from human postmortem brain tissue (Brodmann's Area 11) from 15 adult subjects with BD and 15 healthy control subjects was hybridized to Affymetrix HG-U133A GeneChip to identify differentially expressed (DE) genes in the disease state. After quality control in this study, 10 BD and 11 control subjects were used for further bioinformatic analysis.

(3) The third study has the GEO Accession number GSE21935, and the microarrays preparation followed the guidelines of MIAME in the way it is described in [29]. 60 postmortem RNA samples derived from brain tissue (Brodmann's Area 22) of schizophrenic and control patients were hybridized to the Affymetrix HG-U133 Plus 2.0 Array. After quality control stage samples from 19 control and 23 SZ subjects were subjected to bioinformatic analysis.

(4) The fourth study has the GEO Accession number GSE12649, and the microarrays preparation followed the guidelines of MIAME in the way it is described in [30]. RNA samples were extracted from postmortem brain tissue (Brodmann's Area 46) of 35 BD subjects, 35 SZ subjects, and 35 healthy control subjects.

The RNA was applied to the Affymetrix HG-U133A GeneChip. After quality control stage in this study, 35 SZ, 33 BD samples, and 34 control samples were finally subjected to bioinformatic analysis.

2.2. Analysis of Microarray Data. The raw signal intensity data of each study were imported into the Gene Automated and Robust MicroArray Data Analysis (Gene ARMADA) software tool [31] for versatile, microarray data analysis. In order to extract the signal intensities from the raw data, specific steps were followed: background correction was performed with the gcRMA method and was followed by Quantile normalization. The negative intensity values were treated with the minimum positive and noise method and then summarization followed with the Median Polish method. The data were transformed in $\log_2$ values. In each analysis two experimental conditions were always selected: the disease condition and its corresponding control condition. Genes that were characterized as absent in more than 40% of the samples in each experimental condition were excluded from further analysis. The missing values were imputed using the k-nearest neighbor (k-NN) algorithm. All the steps of the microarray analysis were common for all the extracted datasets.

2.3. Statistical Analysis. The probe sets that were differentially expressed in the disease samples compared to the control healthy samples were selected by two-tailed Student's t-test. The lists of the DE probe sets were defined by applying the following criteria in each dataset: (i) 1.3 or greater-fold change (FC) of the mean expression in all studies, except for the fourth study of BD samples compared to controls with FC > 1.2 (small number of DE genes with stricter cutoff) and (ii) P value threshold below 0.05. The P value distribution for each gene list was used to estimate the False Discovery Rate (FDR) levels. The final gene list corresponds to an FDR < 0.05. The statistical analysis was also performed in the Gene ARMADA software.

2.4. Prioritized Pathway/Functional Analysis of Differentially Expressed Genes. In order to derive better insight into the biological processes related to the DE genes, the lists of significant genes from each microarray analysis were subjected to statistical enrichment analysis using the Statistical Ranking Annotated Genomic Experimental Results (StRAnGER) web application [22]. This bioinformatic tool is using gene ontology term (GOT) annotations and KEGG pathways as well as statistical overrepresentation tests further corrected by resampling methods, aiming to select in a prioritized fashion those GOTs and pathways related to the DE genes, that do not just have a high statistical enrichment score, but also bear a high biological information, in terms of differential expression. Specifically gene ontology (GO) based analysis and KEGG-based analysis result in a list of GO terms and KEGG pathways, respectively, based on hypergeometric tests with values <0.05, which have been reordered according to bootstrapping to correct for statistical distribution-related bias.

2.5. Prioritizations of Putative Disease Genes. In order to prioritize the gene list of interest according to the functional involvement of genes in various cellular processes, thus indicating candidate hubgenes, after inferring the theoretical topology of the GOT-gene interaction network delineated, we used the online tool GOrevenge [32] with the following settings: Aspect: BP (Biological Process), Distance: Resnik, Algorithm: BubbleGene, and Relaxation: 0.15. By adopting these settings we are able to exclude from the interaction network the bias relating to the presence of functionally redundant terms, describing the same cellular phenotypic trait, and thus assessing the centrality, namely, the correlation of the specific genes to certain biological phenotypes in an objective way.

Finally, BioGraph [33] is a data integration and data mining platform for the exploration and discovery of biomedical information. The platform offers prioritizations of putative disease genes, supported by functional hypotheses. BioGraph can retrospectively confirm recently discovered disease genes and identify potential susceptibility genes, without requiring prior domain knowledge, outperforming other text-mining applications in the field of biomedicine.

3. Results and Discussion

3.1. Differentially Expressed Probesets. After the microarray analysis and the statistical selection, lists of DE probesets for each dataset occurred. From the first and fourth studies' analysis, four lists of significantly differentiated probesets were generated: two after comparison of SZ and control subjects and two after comparison of BD and control subjects. The second study (comparison of BD patients to control subjects) resulted also in a list of DE probesets and the third study in another list of DE probesets (SZ subjects compared to control subjects). The differentiated probesets from each case are depicted in representative volcano plots (Figure 1).

3.1.1. Differentially Expressed Genes in Each Study. In postmortem studies the alterations in the gene expression are usually lower than twofold [29]. For each study, transcripts of interest and of particular expression alterations are described in the following paragraphs. The lists of DE genes for each study are presented in Supplementary Tables 1–6 (available online at doi:10.1155/2013/685917). Information about the protein products arising from the DE genes has been provided mainly from the Reference Sequence (RefSeq) database of NCBI [34].

First Study. Statistical analysis of the gene expression profile of SZ and BD patients as compared to controls is summarized in Table 1. The number of DE genes is 196 and 134 respectively.

In SZ patients, transcripts related to the membrane hypothesis show altered expression. Lipases LPL and LIPA, downregulated phosphodiesterases ENPP2 and PDE8A, downregulated phosphoinositide PIK3R4, PNPLA4 phospholipase are related to membrane metabolic processes. ENPP2 and PDE8A dysregulation could also be related to

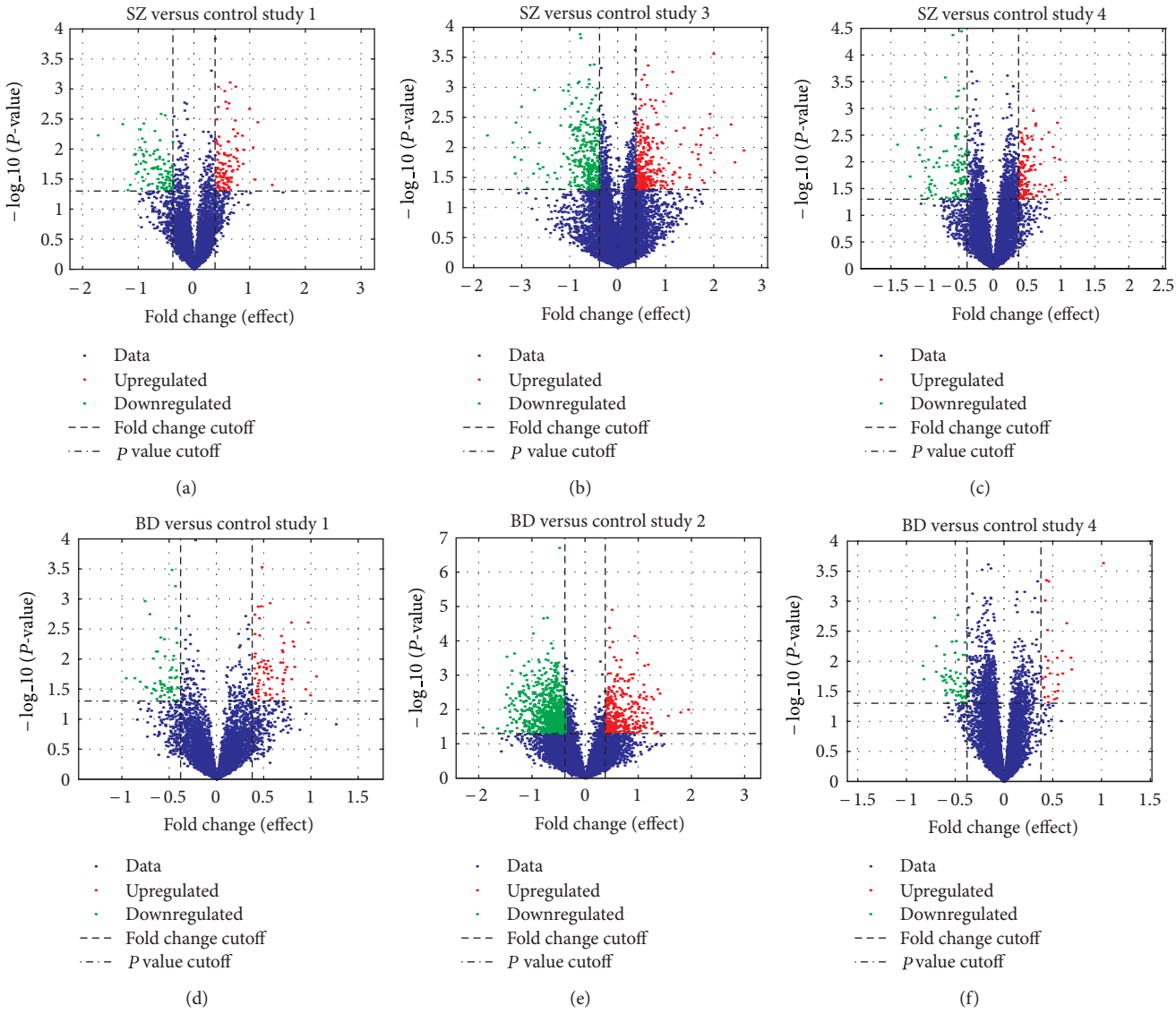


Figure 1: Volcano plots of DE probesets, generated from two-tailed Student's t-test. Upregulated genes in the disease state are depicted with red-colored spots and downregulated genes with green-colored spots. The first three plots (a, b, c) represent DE genes in SZ patients from first, third, and fourth studies, respectively, and the following three plots (d, e, f) represent DE genes in BD patients from first, second, and fourth studies, respectively. FC ratio between gene expression in disease state and healthy state is depicted in the horizontal axes for each dataset in $\log_2$ scale, and P values in $-\log_{10}$ scale are depicted in vertical axes. All plots are similar in most studies, except for plot (e), which shows more green and red spots. This fact means that the number of DE genes is similar in most studies but in study 2 there is a greater number of statistically significant genes in comparison to other plots.

previous MRS studies revealing different levels of phosphodiesters in SZ patients [23]. Some genes encoding proteins of signal transduction pathways, for example, downregulated G protein-coupled receptors GPR37 and GPRC5B, downregulated kinase activity encoding genes PIK3R4 and AATK, or SST somatostatin and CX3CR1 chemokine receptor can also be related to membrane dysfunctions [17]. Genes encoding ion homeostasis seem to be dysregulated as well. NPY, GRIN2A, and CACNA1C all annotated to Ca ion transport (provided by Gene Ontology Annotation UniProt Database) are DE. Also expression of manganese ion binding genes and copper ion binding genes (provided by Gene Ontology Annotation UniProt Database), such as MT1X, is affected. KCNQ2 encoding K voltage-gated channel is overexpressed.

In BD patients of the same study transcript ATP1A3, expressing Na,K-ATPase is downregulated. This ATPase is very important for the normal regulation of the primary active transport mechanism of the cells [29]; thus it affects indirectly the normal function of the AA active transport into the cells. Other dysregulated genes contribute to abnormal K binding and transport (provided by Gene Ontology Annotation UniProt Database): SLC12A5, KCNK3, and KCNK1 are downregulated. POLR2 K encoding phosphodiesterase 6D is upregulated. This fact complies with dysregulated

Table 1: Number of DE genes and probesets, in SZ and BD patients as compared to healthy controls. Genes are characterized as overexpressed when they present positive FC > |0.37| in $\log_2$ scale and as downregulated when they present negative FC respectively. Out of 63000 probesets and 10000 genes of the Affymetrix HG-U95 platform, we derived a much smaller number of probesets and genes.

Disease versus control	Overexpressed genes	Downregulated genes	Total DE genes	Total probesets
SZ versus control	103	93	196	203
BD versus control	74	60	134	134

Table 2: Number of DE genes and probesets, occurring from comparison of BD gene expression profile and control expression profile. Genes are characterized as overexpressed when they present positive FC > |0.37| in $\log_2$ scale and as downregulated when they present negative FC, respectively. Out of 45000 probesets and 33000 genes of the Affymetrix HG-U133A GeneChip, we derived a much smaller number of probesets and genes.

Disease versus control	Overexpressed genes	Downregulated genes	Total DE genes	Total probesets
BD versus control	303	732	1035	1162

membrane phospholipid metabolism, as phosphodiesters are products of this metabolic pathway [17]. SLC7A8 gene is overexpressed. The importance of this gene relies on the fact that it is encoding transmembrane Na-independent AA transport proteins of the L system. LAT1 protein complex, which is specifically expressed from SL7A8 gene, is a tertiary active transporter and mediates tyrosine, tryptophan, and other neutral AA transport systems through cell membranes [19].

Second Study. Statistical analysis of the gene expression profile of BD patients as compared to controls is summarized in Table 2. The number of DE genes is 1035.

Many transcripts regulating ion transport are shown to be downregulated in this study: SCN1A, KCNK1, TRPC1, ATP6V1A, and ATP5G3. Many metallothionein encoding genes (provided by Gene Ontology Annotation UniProt Database) (MT1X, MT2A, MT1E, MT1M, MT1H, MT3, MT1A, and MT1G) are overexpressed. The latter genes combined with downregulated genes COX11, PAM, and RNF7 seem to result in abnormal copper ion binding, because their protein products are involved in this pathway (provided by Gene Ontology Annotation UniProt Database). Genes, encoding ATPases related to Ca++ (ATP2B1, ATP2B2) and H+ (ATP5G3, ATP6AP2, ATP6V1A, ATP6V1D, ATP6V1G2) transporting (provided by Gene Ontology Annotation UniProt Database), are downregulated. The protein encoded by the overexpressed ATP1B1 gene is a member of the family of Na+/K+ and H+/K+ ATPases, as well as a member of the subfamily of responsible proteins for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membranes [29]. PLA2G5 gene encodes an enzyme that belongs to PLA family. It catalyzes the membrane phospholipid hydrolysis to free FA, and in this study it is overexpressed. Overexpressed PLA2G4A also encodes an enzyme of A2 family. It hydrolyzes phospholipids to ARA (provided by RefSeq). ARA is subsequently metabolized into eicosanoids. Prostaglandins and leukotrienes belong to the eicosanoids, and they are lipid-based cell hormones that regulate inflammation pathways and cellular thermodynamics. The catalyzed hydrolysis also results in lysophospholipids that are further utilized as platelet-activating factors. High Ca++ levels and phosphorylation activate the enzyme (provided by RefSeq). 37 genes encoding proteins involved in magnesium ion binding (provided by Gene Ontology Annotation UniProt Database) show altered expression. Phosphoinositide-3-kinases encoded by downregulated genes PIK3C3, PIK3CB, and PIK3R1 encode phosphoinositide 3-kinases (PI3 K). These kinases are involved in signaling pathways, and their receptors are located on the outer cell membranes [17].

Third Study. Statistical analysis of the gene expression profile of SZ patients as compared to controls is summarized in Table 3. The number of DE genes is 122.

The membrane-related protein encoded by the overexpressed ABCA1 gene is a member of ATP-binding cassette (ABC) transporter proteins superfamily. ABC proteins mediate transport of many molecules across extra- and intracellular membranes. ABC1 transporter subfamily's substrate is cholesterol; thus its function is affecting the cellular lipid removal pathway. This gene is related to Tangier's disease and familial high-density lipoprotein deficiency (provided by RefSeq). Apart from ABCA1 gene, also SLC27A3, HSD11B1, CHPT1, and GM2A genes encoding proteins associated with lipid metabolic processes (provided by Gene Ontology Annotation UniProt Database) present a different expression in SZ patients compared to controls. In the DE list CACNB2 is present as an overexpressed gene. This gene encodes a subunit of a voltage-dependent Ca channel protein which is a member of the voltage-gated Ca channel superfamily (provided by RefSeq). CACNA1B, encoding another Ca channel that regulates neuronal release of neurotransmitter, has been proved to be involved in BD and SZ (provided by RefSeq).

Fourth Study. Statistical analysis of the gene expression profiles of SZ and BD patients as compared to controls is summarized in Table 4. The number of DE genes is 216 and 205, respectively.

In SZ patients of these study genes ATP2B2 and ATP2B4 are downregulated and upregulated, respectively. These genes encode proteins that belong to the family of P-type ATPases.

Table 3: Number of DE genes and probesets, occurring from comparison of SZ gene expression profile and control expression profile. Genes are characterized as overexpressed when they present positive FC > |0.37| in $\log_2$ scale and as downregulated when they present negative FC, respectively. Out of 54921 probesets and 38500 genes of Affymetrix HG-U133 Plus 2.0 Array, we derived a much smaller number of probesets and genes.

Disease versus control	Overexpressed genes	Downregulated genes	Total DE genes	Total probesets
SZ versus control	88	34	122	128

Table 4: Number of DE genes and probesets, occurring from comparison of SZ or BD gene expression profile and control expression profile. In case of SZ vs control samples genes are characterized as overexpressed when they present positive FC > |0.37| in $\log_2$ scale and in case of BD vs control when they present FC > |0.26| in $\log_2$ scale. Genes are characterized as downregulated when they present the negative FCs respectively. Out of 45000 probesets and 33000 genes of the Affymetrix HG-U133A GeneChips, we derived a much smaller number of probesets and genes.

Disease versus control	Overexpressed genes	Downregulated genes	Total DE genes	Total probesets
SZ versus control	113	103	216	227
BD versus control	69	136	205	210

These enzymes regulate primary ion transport. These two specific ATPases are very important for the homeostasis of Ca in the cell, as they catalyze cellular efflux of bivalent Ca ions from cells against great concentration gradients (provided by RefSeq). Ca ion homeostasis and Ca ion transport (provided by Gene Ontology Annotation UniProt Database) are also dependent on some other genes dysregulated in this study, such as upregulated NPY, RYR3, and ITPR2 and downregulated CXCL12. Two metallothionein encoding genes MT1X and MT1H are overexpressed. After pathway analysis, these genes, in concert with the differentiated expression of several other genes, seem to affect zinc ion binding and copper ion binding (provided by Gene Ontology Annotation UniProt Database).

In BD patients of the fourth study ATP1A2 is overexpressed. The protein expressed by this gene is a member of P-type cation transport ATPases and belongs to the subfamily of Na,K-ATPases. It belongs to integral membrane proteins, responsible for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membrane. These gradients are very important for osmoregulation, for Na-coupled transport of many organic and inorganic molecules, and for nerve and muscle electrical excitability. The catalytic subunit of Na,K-ATPase is encoded by multiple genes (provided by RefSeq). PLA2G16 is downregulated. The protein encoded by this gene belongs to a superfamily of PLA enzymes. PLA regulates adipocyte lipolysis and release of FA through a G-protein coupled pathway involving prostaglandin and prostaglandin receptors. It belongs to the phospholipase C enzymes that are activated by G-coupled regulatory pathways, such as serotoninergic 5-HT2 pathways (provided by RefSeq). Finally overexpressed metallothioneins MT1X, MT1M, MT1H, and MT1M may result in copper ion binding dysfunctions, as they are involved in this biological function (provided by Gene Ontology Annotation UniProt Database).

3.1.2. Common Differentially Expressed Genes in the Examined Studies. In the first and fourth study SZ gene expressions and BD gene expressions are compared to the same control

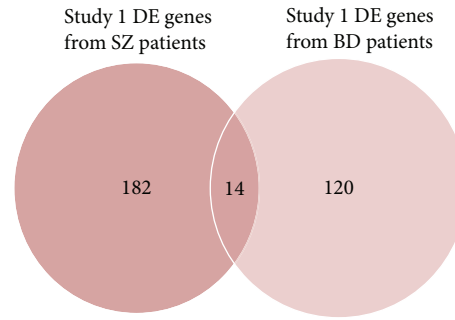


Figure 2: Venn diagram drawn based on DE genes in SZ and BD patients compared with controls of the first study from Brodmann's Area 10 (cognitive functions, goal formation functions). The common DE genes are represented by the intersection of the two circles.

gene expressions. Common DE genes in SZ and BD patients compared to the same control subjects, for example, in the first (Figure 2) and fourth (Figure 3) examined studies are depicted in Tables 5 and 6, respectively. The genes present in lists of statistical significant genes derived from SZ patients' expression profiles are given in Table 8. The common genes in all DE genes of BD patients compared to control groups from all related studies are presented in Table 7. MT1X gene is overexpressed in all studies, in all gene expression comparisons, except for the second study, where it is not among the statistical significant genes as shown in Figure 4.

Among the common DE genes in BD and SZ patients of the first study HTR2C is an interesting gene. Serotonergic pathway is highly related to psychiatric disease expressions. The neurotransmitter serotonin (5-hydroxytryptamine, 5-HT) causes many physiological functions after binding to receptor subtypes, such as 5-HT2 family of seven-transmembrane-spanning, G-protein-coupled receptors. These receptors activate phospholipase C and D signaling pathways. This gene encodes the 2C subtype of serotonin receptor, and its RNA editing is predicted to alter AAs within the second intracellular loop of the 5-HT2C receptor and generate receptor isoforms that differ in their ability to

Table 5: The fourteen common DE genes in schizophrenic and BD samples compared to control samples derived from the first study.

Gene symbol	FC ($\log_2$) SZ versus control	FC ($\log_2$) BP versus control	Gene title
SLC25A1	*−0.624219*	*−0.627028*	"Solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1"
HTR2C	*−0.511652*	−0.515884	5-hydroxytryptamine (serotonin) receptor 2C
SYP	*−0.506666*	*−0.644315*	Synaptophysin
SERINC5	*−0.476598*	*−0.564567*	Serine incorporator 5
CGRRF1	**0.388505**	**0.443519**	Cell growth regulator with ring finger domain 1
SF3B1	**0.434178**	**0.435295**	Splicing factor 3b, subunit 1, 155 kDa
ADD2	**0.476098**	**0.529755**	Adducin 2 (beta)
GRK5	**0.554659**	*−0.593328*	G protein-coupled receptor kinase 5
UCHL3	**0.587522**	**0.701958**	ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)
DARC	**0.642385**	**0.498777**	Duffy blood group, chemokine receptor
SEPT11	**0.651131**	*−0.551204*	septin 11
MT1X	**0.754667**	**0.966154**	Metallothionein 1X
CEBPD	**0.774212**	**0.726239**	CCAAT/enhancer binding protein (C/EBP), delta
LGALS3	**0.892986**	**0.636527**	Lectin, galactoside-binding, soluble, 3

Downregulation of genes in each disease state compared with controls is represented with negative FC values (fold decrease) and upregulation with positive FC values. Most statistically significant genes, common in SZ and BD, are differentiated in similar way.

Table 6: Common DE genes in SZ and BD patients as compared to control samples derived from the fourth study. Top twenty genes (BD) are shown.

Gene symbol	FC ($\log_2$) SZ versus control	FC ($\log_2$) BD versus control	Gene title
DERL1	*−0.9218*	*−0.59278*	Der1-like domain family, member 1
DDX27	*−0.58735*	*−0.55081*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27
NELL1	*−0.48395*	*−0.49181*	NEL-like 1 (chicken)
WDR41	*−0.561422*	*−0.47103*	WD repeat domain 41
SST	*−0.56168*	*−0.47692*	Somatostatin
ZYX	*−0.55832*	*−0.4319*	Zyxin
SSR1	*−0.79829*	*−0.41544*	Signal sequence receptor, alpha fibronectin
FSD1	*−0.4133*	*−0.39578*	Type III and SPRY domain containing 1
TRIM27	*−0.51857*	*−0.39195*	Tripartite motif-containing
TESC	*−0.546183*	*−0.364501*	27 Tescalcin
HES1	**0.383441**	**0.32929**	Hairy and enhancer of split 1
MT1H	**0.477326**	**0.329479**	(Drosophila) metallothionein 1H
GJA1	**0.694821**	**0.332313**	Gap junction protein, alpha 1, 43 kDa
TRIL	**0.405464**	**0.343382**	TLR4 interactor with leucine-rich repeats
MT1X	**0.60052**	**0.35402**	Metallothionein 1X
AGXT2L1	**0.816962**	**0.375859**	Alanine-glyoxylate aminotransferase 2-like 1
GREB1	**0.623598**	**0.418634**	Growth regulation by estrogen in breast cancer 1
EMX2	**0.975302**	**0.545582**	Empty spiracles homeobox
GPC5	**0.772653**	**0.591493**	2 glypican 5
ALDH1L1	**1.0583**	**0.599394**	Aldehyde dehydrogenase 1 family, member L1

Downregulation of genes in each disease state is represented with negative FC values (fold decrease) and upregulation with positive FC values. Most statistically significant genes, common in SZ and BD, are differentiated in similar way.

Table 7: Genes present in all gene lists from all studies including comparison of gene expression between BD samples and control samples.

Gene symbol	FC BD versus control (Study 1)	FC BD versus control (Study 2)	FC BD versus control (Study 4)	Gene title
SDC4	**0.403522**	**0.79702**	**0.323976**	Syndecan 4
MT1X	**0.440635**	**1.1129**	**0.35402**	Metallothionein 1X channel
KCNK1	*−0.416116*	*−0.5935*	*−0.280259*	Potassium, SubfamilyK, Member 1
MT1H	**0.684202**	**1.09618**	**0.329479**	Metallothionein 1H
POLR3C	**0.563585**	**1.28172**	*−0.335122*	Polymerase (RNA) III (DNA directed) PolypeptideC (62 kDa)

Downregulation of genes in each disease state is represented with negative FC values (fold decrease) and upregulation with positive FC values. Most statistical significant genes, common in all BD studies are differentiated in similar way.

Table 8: Genes present in DE gene lists from all studies including comparison of gene expression between SZ samples with control samples.

Gene symbol	FC SZ versus control (Study 1)	FC SZ versus control (Study 3)	FC SZ versus control (Study 4)	Gene title
SRGN	**0.777085**	**0.42152**	—	Serglycin
PRPF4B	**0.563723**	—	**0.415853**	PRP4 pre-mRNA processing factor 4 homolog B (yeast)
MT1X	**0.754667**	—	**0.60052**	Metallothionein 1X
GYG2	**0.754525**	—	*0.686934*	Glycogenin 2
NR4A2	*−0.90769*	—	*−0.550066*	Nuclear receptor subfamily 4, group A, member 2
NPY	*−0.568144*		*−0.406243*	Neuropeptide Y
SST	*−0.83089*	—	*−0.561683*	Somatostatin
PALLD	—	**0.509794**	**0.401231**	Paladin, cytoskeletal Associated protein
AQP4	—	**0.449303**	**0.714565**	Aquaporin 4
ARPC1B	—	**0.392173**	*−0.597327*	Actin-related protein 2/3 complex, subunit 1B, 41 kDa
PVALB	—	*−0.403296*	*−0.432033*	Parvalbumin
HSD11B1	—	*−0.413042*	*−0.538573*	Hydroxysteroid(11-beta)dehydrogenase1
PHLDA2	—	*−0.452704*	*−0.455578*	Pleckstrin homology-like domain, family A

Downregulation of genes in each disease state is represented with negative FC values (fold decrease) and upregulation with positive FC values. Most statistical significant genes, common in SZ studies are differentiated in similar way.

interact with G proteins and the activation of phospholipase C and D signaling cascades, thus modulating serotonergic neurotransmission in the central nervous system. Studies in humans have reported abnormalities in patterns of 5-HT2C editing in depressed suicide victims. Three transcript variants encoding two different isoforms have been found for this gene. This gene is downregulated in both diseases [17]. Serotonin neurotransmitter has been proved to play an important role in emotional, sexual, and eating behavior and in other symptoms of mental diseases, such as hallucinations. Many drugs used for the treatment of these diseases are serotonin agonists. Upregulated ADD2, GGRRF1, and MT1X encode proteins related to metal ion binding. HTR2C, DARC, and GRK5 products participate in signal transduction pathway.

The protein encoded by SDC4 gene is a transmembrane heparan sulfate proteoglycan that functions as a receptor in intracellular signaling. Downregulated KCNK1 gene encodes one of the members of the superfamily of K channel proteins, and it has been previously reported as dysregulated in BD patients [35]. The downregulation of this gene may affect the passive transport of K into the cells.

NPY (neuropeptide) and GABA-system-related SST (somatostatin) are downregulated in two of our SZ studies. These genes have been reported in many studies as candidate psychosis genes [36]. They have also been related to SZ. Earlier studies reveal also downregulation of these specific genes. Neuropeptide genes are involved in working memory functions [37]. In psychiatric diseases working memory and neurodegeneration have been suggested as possible abnormal functions of the prefrontal cortex. These genes seem to be implicated in these functions [36]. PALLD gene, myocardial infarction-related gene, has also been reported as dysregulated in SZ [38]. The protein encoded by AQP4 gene is involved in the regulation of the water homeostasis. Upregulation of this gene has been already reported and has

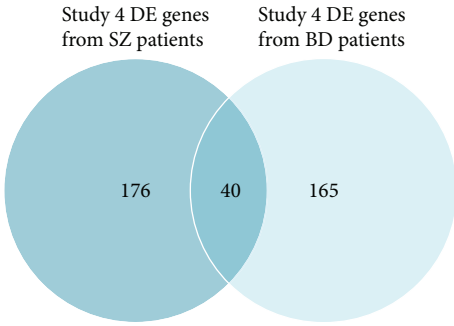


Figure 3: Venn diagram drawn based on DE genes in SZ and BD patients compared with controls of the fourth study from Brodmann's Area 46 (attention and working memory functions). The common DE genes are represented by the intersection of the two circles.

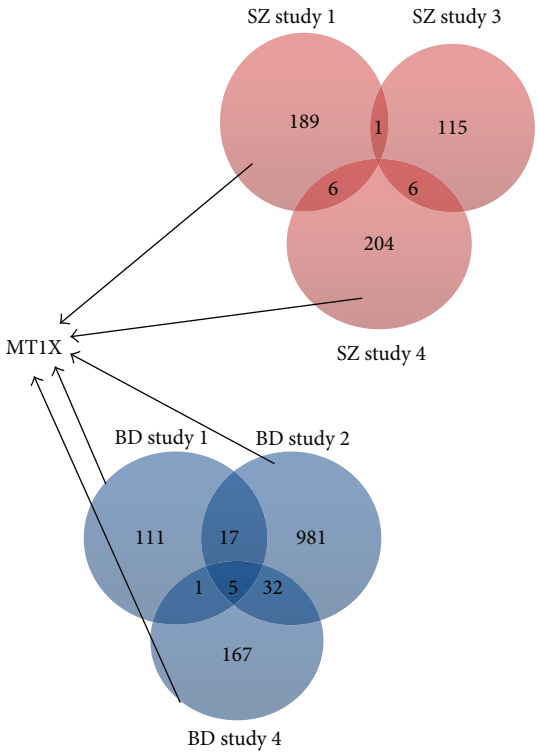


Figure 4: Venn diagram drawn based on DE genes in SZ and BD patients compared with controls. Red circles represent number of DE genes of SZ samples and blue circles represent number of DE genes of BD samples. MT1X is DE in all studies apart from study 3. All studies include samples from frontal cortices, apart from study 3.

been related to white matter hyperintensity, observed in MRS studies of BD patients [27]. Generally there are no common genes in all three SZ datasets. This could be explained by the fact that there are region-specific alterations in SZ, and our SZ raw data were extracted from different brain regions.

3.2. Pathway Analysis. The lists of statistical significant genes of each study were submitted to StrAnGER web application elucidating overrepresented GO terms. The results of GO-analysis for each dataset are presented in Supplementary Tables 7–12.

In the first study, StRAnGER analysis in the SZ-related DE gene list indicated that K ion binding and transport are two of the statistical significant altered GO terms. These processes are very important for the maintenance of K ion gradients in the cells. K ion transport regulates the fluxes of K ions from and into the cells via some transport proteins or pores [19, 25].

StRAnGER analysis in the BD-related DE gene list indicated altered synaptic pathways. Synaptic pathways and genes have been reported earlier as possible dysfunction factors in BD [39]. G-protein pathways are also related to neurotransmitter receptors and particularly to serotonergic receptors, most studied in BD as part of serotonergic pathway [17]. Ca transport, protein tyrosine kinase, and phosphoinositide binding are involved in signal transduction pathways. Several studies of BD patients have shown abnormalities in the phosphoinositol/protein kinase C (PKC) signaling system. One such study has demonstrated significantly higher concentrations of 4,5-bisphosphate (PIP2) in the platelet membranes of patients in the manic phase of BD; they also found that the levels of PIP2 increased when cycling from the euthymic state into the manic state. Additionally, the activity of platelet PKC was found elevated in patients, during a manic episode of BD. Additionally several independent studies have shown increased concentrations of the stimulatory alpha subunit (G_{as}) of G-protein in the brains of BD patients, specifically in the frontal, temporal, and occipital cortices. Other studies have suggested there is also increased presence/activity of G-proteins in the leukocytes of untreated manic patients and the mononuclear leukocytes of bipolar, but not unipolar, patients. Currently, there is no evidence to indicate that the increased concentration of G_{as} is caused by gene mutations; it has been suggested that they could be caused by a change in any of the biochemical pathways leading to the transcription and translation of the G_{as} gene [40]. Copper ion binding belongs to the significant GOTs as well.

In the second study copper ion binding, magnesium ion binding, chloride channel activity, chloride transport, postsynaptic membrane, and inositol or phosphatidylinositol phosphatase activity represent significantly differentiated GOTs.

In study 3 and 4 defense response, immune response, and inflammatory response GOTs are present in the over-represented GOTs. The inflammatory system is strongly related to these mental disorders, and the immune underlying mechanisms remain mainly obscure [41]. Lipid metabolic process is also a statistically significant GOT altered in study 3.

Dysregulated neurotransmitter systems in the central nervous system of BD and SZ patients have been systematically reported [2, 4]; thus central nervous system development is among the GO terms resulting from pathway analysis of study 3 BD DE list. Copper ion binding, chloride ion binding, and signal transduction pathways seem to be affected.

Table 9: Overrepresented GO terms extracted from the union of 68 common genes either of BD patients or of SZ patients.

GO annotation	GOT *P*-value	Enrichment
Protein amino acid phosphorylation	0.000254537	6/424
ATP binding	0.000266417	10/1063
Protein binding	0.000833654	19/3248
Transferase activity	0.001557772	8/925
Nucleotide binding	0.001893973	10/1348
Cytoplasm	0.002264782	10/1379
Extracellular region	0.005570074	5/547
Metabolic process	0.0076371	4/414
Multicellular organismal development	0.011932632	5/644
Endoplasmic reticulum	0.018083004	4/514
Zinc ion binding	0.024540778	8/1430
Plasma membrane	0.034457754	4/610

Copper ion binding is present in almost all lists of significantly altered GO terms. Signaling pathways are among the KEGG pathways that appear more often as overrepresentative pathways (Supplementary Table 13).

We also performed GO analysis in the 68 genes, shown schematically in Figure 4, that were present in at least two of the BD or SZ DE lists. Table 9 summarizes the GO terms of this pathway analysis. ATP binding is essential for the maintenance of the ion gradients in the cell. ATP is universally an important coenzyme and enzyme regulator [19].

3.3. Identification of Candidate Hub Genes. In order to expand our knowledge regarding which genes have critical role among the common DE genes in BD datasets, we used the online tool GOrevenge [32], which performs prioritization of the gene list taking into consideration the centrality of each gene, as described in the GO tree. The 68 genes found differentiated in at least two BD- or SZ-related studies were submitted to GOrevenge, and the analysis was performed based on GO annotations for Homo sapiens as described in materials and methods section. A prioritized list of genes, containing candidate linker genes, that is, genes participating in many different cellular processes, was derived (Table 10). Among them, three genes, namely, APOE, RELA, and NPY, have also been found as statistically significantly differentiated in at least two of either SZ or BD DE gene lists.

3.4. Prioritizations of Putative Disease Genes. By setting SZ and BD as concept, the relation of each gene with the BD and SZ was assessed, and the 68 genes found differentiated in at least two BD- or SZ-related studies were prioritized by BioGraph algorithm as shown in Tables 11 and 12, respectively. The genes are prioritized according to their score which is a statistical enrichment measure of the relevance of each gene with the inquired context (here specified as either BD or SZ) to the total relations (references) of the gene in the universe of terms. In this way, the user can derive which of its genes are already associated and in what extent with a given disease or generally biological term and which of them represent novel findings with respect to the investigated pathological phenotype. APOE, RELA, and NPY have also high scores and are among the ten top genes related either to the BD or SZ after the prioritization of genes in BioGraph. These three genes have been shown to play a major role in the examined studies, after different bioinformatic analyses. NPY has been reported as a candidate psychosis gene, as aforementioned.

Table 10: GOrevenge prioritization. The second column refers to the number of GO terms remaining after GOrevenge pruning, reflecting the centrality of each gene, while the third column refers to the original number of biological process category GO terms of each gene. Top 20 genes are shown. Genes presented in italics are among the statistically significant differentiated genes in at least two of either SZ or BD DE gene lists.

Gene symbol	Remaining GO terms	Original GO terms
TGFB1	56	126
CTNNB1	53	117
BCL2	50	121
SHH	45	142
AKT1	44	73
PSEN1	39	70
WNT5A	38	98
APOE	38	54
BMP4	37	128
TNF	37	88
FGF10	36	102
IL1B	35	75
AGT	34	63
P2RX7	33	68
SFRP1	32	81
RELA	32	50
TGFB2	32	66
BMP2	32	59
PPARG	31	51
EP300	31	46

APOE regulates cholesterol of the central nervous system; thus any alteration in APOE levels may result in abnormal brain function. APOE has been mostly related to Alzheimer's disease [42].

Genotyping studies and Western plot analysis have shown differences of APOE in SZ patients. Abnormal cholesterol metabolism has been associated with SZ as well. High levels of three different apolipoproteins in brains of patients with psychiatric disorders may indicate aberrant central nervous system lipid metabolism. Additionally, APOE has been implicated in inflammation pathways, after studies on mice revealing possible action of APOE as inflammatory response inhibitor. Inflammation pathways are considered candidate mechanisms responsible for the pathogenesis of several mental disorders and mainly of SZ [42].

RELA, v-rel reticuloendotheliosis viral oncogene homolog A (avian), is also involved in immune and inflammatory responses, as it encodes the main component of the

Table 11: Prioritization of the genes presented in table 11, by BioGraph exploiting unsupervised methodologies for the identification of causative SZ-associated genes. Genes with the higher nineteen scores are shown.

Gene symbol	Score
PVALB	0.172895
SYN2	0.084975
APOE	0.013519
RELA	0.00034
CRK	0.000246
NTRK2	0.000219
MAPT	0.000136
TRIP13	0.000127
NPY	$7.39E-05$
MT1X	$6.19E-05$
NR4A2	$4.25E-05$
SDC4	$3.57E-05$
PGK1	$3.29E-05$
PRPF4B	$3.21E-05$
SST	$2.35E-05$
TRPC1	$2.28E-05$
LGALS3	$2.19E-05$
DUSP6	$1.96E-05$
BGN	$1.66E-05$

Table 12: Prioritization of the genes presented in table 12, by BioGraph exploiting unsupervised methodologies for the identification of causative BD-associated genes. Genes with the higher nineteen scores are shown.

Gene symbol	Score
PVALB	1.930909595
NTRK2	0.520432786
MAPT	0.000852042
RELA	0.000381239
CRK	0.0002833
NPY	0.000109408
APOE	$8.79036E-05$
SYN2	$6.07336E-05$
NR4A2	$5.57465E-05$
TRPC1	$4.28846E-05$
SDC4	$3.78467E-05$
HSD11B1	$3.34794E-05$
TRIP13	$2.26339E-05$
SLC12A5	0.000021501
LGALS3	0.000020488
MT1X	$1.88525E-05$
SST	$1.75935E-05$
DUSP6	0.000015482
AQP4	$1.50416E-05$

NF-κB complex. NF-κB has been related indirectly to SZ, as it is highly correlated to SZ involved cytokines: interleukin-1β (IL-1β), IL-1 receptor antagonist (IL-1RA), IL-6, and tumor necrosis factor-α (TNF-α). NF-κB is a regulator of cytokines' expression, and proinflammatory cytokines activate NF-κB. NF-κB is present in synaptic terminals and participates in regulation of neuronal plasticity. NF-κB regulates genes that encode subunits of N-methyl-D-aspartate receptors, voltage-dependent Ca channels and the Ca-binding protein calbindin, cell survival factors, including Bcl-2, Mn-SOD, and inhibitor of apoptosis proteins (IAPs) and cell death factors, including Bcl-x(S) and Bax. All these genes are related to neurotransmission, and altered expression of several of them has been reported in previous SZ postmortem brain studies [43].

4. Conclusions

The aim of the study was to interpret the results of comparative genomic profiling studies in schizophrenic patients as compared to healthy controls and in patients with BD and try to relate and integrate our results with an aberrant AA transport through cell membranes. Starting from genomewide expression data, the analysis focused on genes and mechanisms involved in AA transport through cell membranes. We performed transcriptomic computational analysis on raw data derived from four different studies. Moreover, a multistage, translational bioinformatic computational framework is employed, previously utilized for the molecular analysis of transcriptomic data of atherosclerotic mice models [44], exploiting different methods in order to identify critical altered molecular mechanisms and important central players. In this way, the results derived here do not rely solely on a single stage of significance. They are complying to a systematic screening of the results, exploiting various statistical measures, in a unified analysis pipeline. These measures either exploit the stringent FDR estimations at the single gene level, further filtered to keep those common in between diseases or studies comparisons. Moreover, the consensus gene lists thus derived are corrected through a rigorous, bootstrapping framework, applied in the statistical enrichment analysis of the significant biological processes. Moreover, critical regulatory genes, prioritized by their total number of GO annotations, to the resulting significant GOTs list, are highlighted. It is also examined, whether these genes have been associated with the disease phenotypes of SZ or BD in the broader biomedical literature. The results were eventually analyzed, complying with a meta-analysis context, giving emphasis on common functional patterns mined amid the various studies.

Our bioinformatic analyses of the downloaded datasets demonstrate genes and GOTs associated with ion transport dysregulation (K, Na, Ca, and other ion transports and bindings) resulting in a disturbed primary active transport, suggesting a deficit in transmembrane Na+ and K+ gradients maintenance. Characteristic downregulation of Na+ and K+ transporting ATPases, enzymes responsible for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membrane, is indicated in the DE gene lists of two of our datasets. They are also upregulated in one dataset (BD patients' expression profiles). Also downregulation of P-type ATPases is reported in the datasets.

Altered distribution of specific ions in the cells may affect distributions of other ion groups. A statistical integration of many studies has previously related published data of Na,K-ATPase activity in erythrocytes of BD patients with the expression of the disease [45]. Decreased activity of Na,K-ATPase has been also related to SZ in previous studies [38]. The disturbed primary active transport observed in our study indicates difficulty in maintaining transmembrane ion gradients. This fact should result in disrupted, secondary, active AA transporter Systems A, X-AG, N, and y+, as they couple AA transport to the electrical and chemical gradients initiated by primary active transport. AA exchangers, systems ASC, y+L and L, that transport AAs by antiport mechanisms, may suffer from a deficit of secondary, actively transported AAs they need for the exchange, resulting in a disrupted transport of AAs mainly transported through this third mechanism.

Genes and pathways related to Ca transport agree with abnormalities in Ca signaling, that have been implicated in BD; findings show elevated intracellular Ca concentrations in the platelets, lymphocytes, and neutrophils of BD patients. Ca is very important in most intracellular signaling pathways and in the regulation of neurotransmitter synthesis and release [40].

Phospholipase activity may be dysregulated in BD and SZ diseases, as indicated by altered expression of the genes encoding this enzyme in this study. This alteration has obvious impacts on the phospholipid metabolism of the membrane, as it is a crucial enzyme in this metabolic pathway [23].

A consistent upregulation of MT1X and generally of metallothionein genes is consistent in different datasets. The functional role of metallothioneins in the brain has not been very well characterized [36]. The main function of metallothioneins is to protect neurons from pathological stressing factors. Abnormal expression of genes encoding these proteins may indicate an endogenous reaction to constant oxidative stress [46]. Several studies have suggested involvement of metallothioneins in functions of the central nervous system, such as neuroprotection, regeneration, and cognitive function. Other studies reported that metallothioneins are involved in cellular response, immunoregulation, cell survival, and brain functional restoration. Metallothioneins are mainly produced in astrocytes. Metallothionein overexpression has been also reported as a contributing factor in brain pathologies, such as excitotoxic injury, amyotrophic lateral sclerosis, Alzheimer's disease, and Parkinson's disease. Animal studies have associated substance dependences and learning procedures with metallothioneins. Other prefrontal cortex (PFC) studies have revealed overexpression of metallothioneins in SZ patients. All these studies indicate involvement of metallothioneins in neuroprotection and cognitive functions. A possible neurodegenerative function in the PFC may affect cognitive function in BD and SZ patients. Overexpression of these genes could then be a defense mechanism against these adverse processes. Metallothioneins have also been proposed as possible medical treatment as they have been tested in animal models and have been proved nontoxic [36].

The observed small number of common DE genes among the different studies reflects heterogeneity among the datasets analyzed, which could be explained by both biological and technical reasons. The brain area under study, the microarray platform used, and the selection of patients and controls could contribute to the heterogeneity and should be taken into consideration and duly addressed, ideally at the stage of the experimental design, whenever analogous meta-analysis tasks are envisioned. Highlighting genes that present different expression in different cases, but in the context of a multitiered systematic framework, like the one presented here, could result in molecular interactions, linked with causative, universal, and molecular pathways in mental disorders.

Abbreviations

ATPase:	Adenosine triphosphatase
AA:	Amino acid
ARA:	Arachidonic acid
ARMADA:	Automate Robust Microarray Data Analysis
BD:	Bipolar disorder
DHA:	Docosahexaenoic acid
EFA:	Essential fatty acids
FA:	Fatty acids
FC:	Fold change
FDR:	False discovery rate
GEO:	Gene expression omnibus
GO:	Gene ontology
GOT:	Gene ontology term
k-NN:	*k*-nearest neighbor
MRS:	Magnetic resonance spectroscopy
NCBI:	National Center for Biotechnology Information
PLA:	Phospholipase A2
Na,K-ATPase:	Sodium-potassium adenosine triphosphatase
StRAnGER:	Statistical Ranking Annotated Genomic Experimental Results
SZ:	Schizophrenia
Ca:	Calcium
Na:	Sodium
K:	Potassium
G_{as}:	alpha subunit of G protein
DE:	Differentially expressed
PFC:	Prefrontal cortex.

References

[1] A. Sawa and S. H. Snyder, "Schizophrenia: diverse approaches to a complex disease," *Science*, vol. 296, no. 5568, pp. 692–695, 2002.

[2] M. L. Persson, J. Johansson, R. Vumma et al., "Aberrant amino acid transport in fibroblasts from patients with bipolar disorder," *Neuroscience Letters*, vol. 457, no. 1, pp. 49–52, 2009.

[3] D. F. Horrobin, "Schizophrenia as a membrane lipid disorder which is expressed throughout the body," *Prostaglandins Leukotrienes and Essential Fatty Acids*, vol. 55, no. 1-2, pp. 3–7, 1996.

[4] T. M. Du Bois, C. Deng, and X. F. Huang, "Membrane phospholipid composition, alterations in neurotransmitter systems and schizophrenia," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 29, no. 6, pp. 878–888, 2005.

[5] F. A. Wiesel, J. L. R. Andersson, G. Westerberg et al., "Tyrosine transport is regulated differently in patients with schizophrenia," *Schizophrenia Research*, vol. 40, no. 1, pp. 37–42, 1999.

[6] F. A. Wiesel, G. Edman, L. Flyckt et al., "Kinetics of tyrosine transport and cognitive functioning in schizophrenia," *Schizophrenia Research*, vol. 74, no. 1, pp. 81–89, 2005.

[7] R. Vumma, F. A. Wiesel, L. Flyckt, L. Bjerkenstedt, and N. Venizelos, "Functional characterization of tyrosine transport in fibroblast cells from healthy controls," *Neuroscience Letters*, vol. 434, no. 1, pp. 56–60, 2008.

[8] E. Olsson, F. A. Wiesel, L. Bjerkenstedt, and N. Venizelos, "Tyrosine transport in fibroblasts from healthy volunteers and patients with schizophrenia," *Neuroscience Letters*, vol. 393, no. 2-3, pp. 211–215, 2006.

[9] L. Flyckt, N. Venizelos, G. Edman, L. Bjerkenstedt, L. Hagenfeldt, and F. A. Wiesel, "Aberrant tyrosine transport across the cell membrane in patients with schizophrenia," *Archives of General Psychiatry*, vol. 58, no. 10, pp. 953–958, 2001.

[10] L. Flyckt, G. Edman, N. Venizelos, and K. Borg, "Aberrant tyrosine transport across the fibroblast membrane in patients with schizophrenia -indications of maternal inheritance," *Journal of Psychiatric Research*, vol. 45, no. 4, pp. 519–525, 2011.

[11] C. N. Ramchand, M. Peet, A. E. Clark, A. E. Gliddon, and G. P. Hemmings, "Decreased tyrosine transport in fibroblasts from schizophrenics: implications for membrane pathology," *Prostaglandins Leukotrienes and Essential Fatty Acids*, vol. 55, no. 1-2, pp. 59–64, 1996.

[12] P. E. Ward, J. Sutherland, E. M. T. Glen, and A. I. M. Glen, "Niacin skin flush in schizophrenia: a preliminary report," *Schizophrenia Research*, vol. 29, no. 3, pp. 269–274, 1998.

[13] L. Flyckt, J. Borg, K. Borg et al., "Muscle biopsy, macro EMG, and clinical characteristics in patients with schizophrenia," *Biological Psychiatry*, vol. 47, no. 11, pp. 991–999, 2000.

[14] R. Vumma, J. Johansson, T. Lewander, and N. Venizelos, "Tryptophan Transport in human fibroblast cells: a functional characterization," *International Journal of Tryptophan Research*, vol. 4, pp. 19–27, 2011.

[15] D. Raucoules, J. M. Azorin, A. Barre, and R. Tissot, "Plasma levels and red blood cell membrane transports of L-tyrosine and L-tryptophan in depressions. Assessment at baseline and recovery," *Encephale*, vol. 17, no. 3, pp. 197–201, 1991.

[16] Y. Bezchlibnyk and L. T. Young, "The neurobiology of bipolar disorder: focus on signal transduction pathways and the regulation of gene expression," *Canadian Journal of Psychiatry*, vol. 47, no. 2, pp. 135–148, 2002.

[17] P. M. Kidd, "Bipolar disorder as cell membrane dysfunction. Progress toward integrative management," *Alternative Medicine Review*, vol. 9, no. 2, pp. 107–135, 2004.

[18] D. F. Horrobin and C. N. Bennett, "Depression and bipolar disorder: relationships to impaired fatty acid and phospholipid metabolism and to diabetes, cardiovascular disease, immunological abnormalities, cancer, ageing and osteoporosis. Possible candidate genes," *Prostaglandins Leukotrienes and Essential Fatty Acids*, vol. 60, no. 4, pp. 217–234, 1999.

[19] R. Hyde, P. M. Taylor, and H. S. Hundal, "Amino acid transporters: roles in amino acid sensing and signalling in animal cells," *Biochemical Journal*, vol. 373, no. 1, pp. 1–18, 2003.

[20] L. Bjerkenstedt, L. Farde, L. Terenius, G. Edman, N. Venizelos, and F. A. Wiesel, "Support for limited brain availability of tyrosine in patients with schizophrenia," *International Journal of Neuropsychopharmacology*, vol. 9, no. 2, pp. 247–255, 2006.

[21] D. F. Horrobin, "The membrane phospholipid hypothesis as a biochemical basis for the neurodevelopmental concept of schizophrenia," *Schizophrenia Research*, vol. 30, no. 3, pp. 193–208, 1998.

[22] A. Chatziioannou and P. Moulos, "Exploiting statistical methodologies and controlled vocabularies for prioritized functional analysis of genomic experiments: the StRAnGER web application," *Frontiers in Neuroscience*, vol. 5, pp. 1–14, 2011.

[23] D. L. Scott, S. P. White, Z. Otwinowski, W. Yuan, M. H. Gelb, and P. B. Sigler, "Interfacial catalysis: The mechanism of phospholipase A2," *Science*, vol. 250, pp. 1541–1546, 1990.

[24] A. Klemm, R. Rzanny, R. Fünfstück et al., "31P-Magnetic resonance spectroscopy (31P-MRS) of human allografts after renal transplantation," *Nephrology Dialysis Transplantation*, vol. 13, no. 12, pp. 3147–3152, 1998.

[25] S. Bröer, "Adaptation of plasma membrane amino acid transport mechanisms to physiological demands," *Pflügers Archiv. European Journal of Physiology*, vol. 444, no. 4, pp. 457–466, 2002.

[26] T. Barrett, D. B. Troup, S. E. Wilhite et al., "NCBI GEO: archive for high-throughput functional genomic data," *Nucleic Acids Research*, vol. 37, no. 1, pp. D885–D890, 2009.

[27] K. Iwamoto, C. Kakiuchi, M. Bundo, K. Ikeda, and T. Kato, "Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders," *Molecular Psychiatry*, vol. 9, no. 4, pp. 406–416, 2004.

[28] M. M. Ryan, H. E. Lockstone, S. J. Huffaker, M. T. Wayland, M. J. Webster, and S. Bahn, "Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes," *Molecular Psychiatry*, vol. 11, no. 10, pp. 965–978, 2006.

[29] M. R. Barnes, J. Huxley-Jones, P. R. Maycox et al., "Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia," *Journal of Neuroscience Research*, vol. 89, no. 8, pp. 1218–1227, 2011.

[30] K. Iwamoto, M. Bundo, and T. Kato, "Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis," *Human Molecular Genetics*, vol. 14, no. 2, pp. 241–253, 2005.

[31] A. Chatziioannou, P. Moulos, and F. N. Kolisis, "Gene ARMADA: an integrated multi-analysis platform for microarray data implemented in MATLAB," *BMC Bioinformatics*, vol. 10, article 1471, p. 354, 2009.

[32] K. Moutselos, I. Maglogiannis, and A. Chatziioannou, "GOrevenge: a novel generic reverse engineering method for the identification of critical molecular players, through the use of ontologies," *IEEE Transactions on Bio-Medical Engineering*, vol. 58, no. 12, pp. 3522–3527, 2011.

[33] A. M. L. Liekens, J. De Knijf, W. Daelemans, B. Goethals, P. De Rijk, and J. Del-Favero, "Biograph: unsupervised biomedical knowledge discovery via automated hypothesis generation," *Genome Biology*, p. R57, 2011.

[34] K. D. Pruitt, T. Tatusova, W. Klimke, and D. R. Maglott, "NCBI reference sequences: current status, policy and new initiatives," *Nucleic Acids Research*, vol. 37, no. 1, pp. D32–D36, 2009.

[35] N. Matigian, L. Windus, H. Smith et al., "Expression profiling in monozygotic twins discordant for bipolar disorder reveals

dysregulation of the WNT signalling pathway," *Molecular Psychiatry*, vol. 12, no. 9, pp. 815–825, 2007.

[36] K. H. Choi, M. Elashoff, B. W. Higgs et al., "Putative psychosis genes in the prefrontal cortex: combined analysis of gene expression microarrays," *BMC Psychiatry*, vol. 8, article 87, 2008.

[37] T. Hashimoto, D. Arion, T. Unger et al., "Alterations in GABA-related transcriptome in the dorsolateral prefrontal cortex of subjects with schizophrenia," *Molecular Psychiatry*, vol. 13, no. 2, pp. 147–161, 2008.

[38] N. Petronijević, D. Mićić, B. Duricić, D. Marinković, and V. R. Paunović, "Substrate kinetics of erythrocyte membrane Na, K-ATPase and lipid perosides in schizophrenia," *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, vol. 27, no. 3, pp. 431–440, 2003.

[39] C. A. Ogden, M. E. Rich, N. J. Schork et al., "Candidate genes, pathways and mechanisms for bipolar (manic-depressive) and related disorders: an expanded convergent functional genomics approach," *Molecular Psychiatry*, vol. 9, no. 11, pp. 1007–1029, 2004.

[40] H. K. Manji and R. H. Lenox, "The nature of bipolar disorder," *Journal of Clinical Psychiatry*, vol. 61, no. 13, pp. 42–57, 2000.

[41] S. Hope, I. Melle, P. Aukrust et al., "Similar immune profile in bipolar disorder and schizophrenia: selective increase in soluble tumor necrosis factor receptor I and von Willebrand factor," *Bipolar Disorders*, vol. 11, no. 7, pp. 726–734, 2009.

[42] E. A. Thomas and J. G. Sutcliffe, "The neurobiology of apolipoproteins in psychiatric disorders," *Molecular Neurobiology*, vol. 26, no. 2-3, pp. 369–388, 2002.

[43] R. Hashimoto, K. Ohi, Y. Yasuda et al., "Variants of the RELA gene are associated with schizophrenia and their startle responses," *Neuropsychopharmacology*, vol. 36, no. 9, pp. 1921–1931, 2011.

[44] O. Papadodima, A. Sirsjö, F. N. Kolisis, and A. Chatziioannou, "Application of an integrative computational framework in trancriptomic data of atherosclerotic mice suggests numerous molecular players," *Advances in Bioinformatics*, vol. 2012, Article ID 453513, 9 pages, 2012.

[45] S. W. Looney and R. S. Ei-Mallakh, "Meta-analysis of erythrocyte Na, K-ATPase activity in bipolar illness," *Depression and Anxiety*, vol. 5, no. 2, pp. 53–65, 1997.

[46] E. Mocchegiania, C. Bertoni-Freddarib, F. Marcellinic, and M. Malavolta, "Brain, aging and neurodegeneration: role of zinc ion availability," *Progress in Neurobiology*, vol. 75, pp. 367–390, 2005.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 769639, 8 pages
http://dx.doi.org/10.1155/2013/769639

Review Article

Biomedical Informatics for Computer-Aided Decision Support Systems: A Survey

Ashwin Belle,[1] Mark A. Kon,[2] and Kayvan Najarian[1]

[1] *Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA*
[2] *Department of Mathematics and Statistics, Boston University, Boston, MA 02215, USA*

Correspondence should be addressed to Kayvan Najarian; knajarian@vcu.edu

Received 30 November 2012; Accepted 9 January 2013

Academic Editors: J. Bajo, Y. Cai, and J. B. T. Rocha

The volumes of current patient data as well as their complexity make clinical decision making more challenging than ever for physicians and other care givers. This situation calls for the use of biomedical informatics methods to process data and form recommendations and/or predictions to assist such decision makers. The design, implementation, and use of biomedical informatics systems in the form of computer-aided decision support have become essential and widely used over the last two decades. This paper provides a brief review of such systems, their application protocols and methodologies, and the future challenges and directions they suggest.

1. Introduction

Over the past several decades the uses and applications of biomedical informatics for computer-aided medical diagnostics and decision support systems have become ubiquitous in clinical settings. Adaptations of decision support systems powered by biomedical informatics in either complex or simple forms were seen as early as the 1970s. A 1994 survey [1] indicates that the literature relevant to this field dates back to as early as the mid-1950s.

With advances in technologies related to medical signal and image acquisition, it can be seen that there has been an escalation of complexity in collected medical data. Apart from medical data being inherently more complex, the sheer volume of such data collected per patient is growing rapidly. Currently medical devices and high-throughput measurement systems produce thousands of images and large volumes of other data per patient in seconds, making it difficult for physicians to parse through the information while providing timely diagnoses and prognoses. There is a present significant need for development and improvement of computer-aided decision support systems in medicine, with an expected amplification in the future.

Clinical implementations of biomedical informatics methods in the form of computer-based decision support systems were seen as early as 1971, when Dombal's system AAPhelp, developed at Leeds University, attempted to automate the diagnosis of acute abdominal pain [2]. In 1974 a system called INTERNIST-I [3], a rule-based expert system designed to aid the diagnosis of complex medical problems in internal medicine, was developed. These represent prominent developments in early implementations of biomedical informatics systems, among many computer-based diagnostic decision support systems. Since their inception there has been a substantial evolution, with wide acknowledgment of their success in improving practitioners' performance and patient outcomes.

With broad research conducted in the area, there have been review studies on related topics. A book by Greenes [4] outlines general concepts and future directions for clinical decision support systems. Similarly, an article by Madabhushi et al. [5] describes development of computer-aided prognosis systems for predicting patient and disease outcomes using multiscale, multimodal medical data. Miller's article in 1994 provides a comprehensive list of important work conducted on diagnosis and decision support between 1954 and 1993

[1]. Similarly, a more recent article by Pearson et al. provides a systematic review of computerized clinical decision support systems between 1990 and 2007 [6]. In this work 56 different studies are considered, of which 38 are on systems used in therapy initiation, 23 involve computer-based monitoring of patients during therapy, and three study conditions for termination of therapy. From their outcomes, the authors infer that the most consistently effective computer-based systems are those that initiate advice to fine-tune existing therapies by improving patient safety, adjusting the doses, durations forms of prescribed drugs, or increasing the laboratory testing rates for patients on long-term therapies.

Some previous studies also provide insights into more specific subgenres of biomedical informatics methods and their implementations in the form of computer-aided diagnosis systems. Tourassi discusses systems that provide diagnostic interpretations based on image texture analysis [7] and Stivaros et al. [8] focus on the impacts of decision support systems in clinical radiological practice.

This paper provides a general survey of applications and methodologies in biomedical informatics that have been implemented as computer-aided decision support systems and discusses the resulting challenges, for example, in validation of such systems, and in adoption levels among end users. The paper is organized as follows. some major application areas for the above-mentioned systems are described, followed by a discussion of important methodologies employed. Then there is a brief look at the validation and success criteria for these systems, followed by the conclusion and discussion of future directions.

2. Applications

There are a number of application areas medicine for which computer-aided decision support systems have become designed and implemented. Some of the major application areas are discussed below.

2.1. Radiology. Here, computer-based image processing and analysis have been an active research area. Combining visualization, image processing, and machine learning for decision-making has provided an added advantage for clinical applications. With multiple technologies for medical imaging such as computed tomography (CT), X-rays, magnetic resonance imaging (MRI), and functional MRI (FMRI), numerous biomedical informatics methods have been designed for application-specific solutions. A study by Van Ginneken et al. surveys over 150 publications before 2001 on computer-aided diagnosis in chest radiography [9]. This survey emphasizes the continued interest in computer-aided diagnosis for chest radiography. There are also several studies on developing decision-making systems using automated analysis of CT scans. These include Chen et al.'s [10, 11] study, which focuses on developing a computer aided diagnostic system that automatically analyses brain CT scans of patients with traumatic brain injury (TBI). The system also automatically estimates the level of the intracranial pressure (ICP) within the brain. Another study by Davaluri et al. discusses the development of computer-assisted decision-making systems for pelvic injuries [12]. Wu et al. focus on fracture detection in traumatic pelvic injury patients and discusses an automated method for quantifying the size of fractures from CT images of patients with pelvic injuries [13]. Stivaros et al. provide an overview of underlying design and functionality of radiological decision support systems, with supporting examples of the development and evolution of such systems in the past 40 years [8].

2.2. Emergency Medicine and Intensive Care Units. One of the most active areas of research in the realm of biomedical informatics and decision support systems is emergency medicine. For patients in intensive care units (ICU) and emergency rooms, it is critical that diagnosis and treatment are provided in a timely manner. Since critical care units typically experience a heavy strain on resources, it becomes important to manage and dispense resources to critically ill patients who need it the most. Computer-aided decision support systems play a vital role in reducing diagnosis time, improving resource allocation efficiency, and decreasing patient mortality. Ji et al. describe a study that provides a comparative analysis of computer-assisted decision-making systems for traumatic injuries [14]. Systems such as one developed by Frixea et al. show how case-based reasoning techniques for the estimation of patient outcomes and resource utilizations can improve patient care dramatically in ICUs [15]. Kumar et al.'s study [16] presents a clinical decision support system which combines both case-based reasoning and rule-based reasoning and that performs well with real and simulated ICU data. Raschke et al. describe a computer alert system which is designed to recognize averse drug events (AEDs) in hospital settings [17]. This system is reported to be capable of generating alerts for patients with increased risk of AEDs. The study states that during the 6-month trial of the system, a total of 265 (44%) of the 596 true positive alerts were unrecognized by the physicians prior to the alert notification, hence showing a great promise for applications in continuous patient monitoring.

2.3. Cardiovascular Medicine. Having continuous or interventional monitoring of cardiovascular signals for diagnosing ailments or predicting impending cardiac events can be an extremely useful tool. Currently there are several research biomedical informatics studies attempting to develop computer-aided solutions for various aspects of cardiovascular medicine. A study conducted by Polat et al. describes a computer-aided diagnosis system that automatically identifies and classifies arrhythmia from the analysis of patients' electrocardiograph (ECG) signals [18]. The authors claim 100% accuracy in classification within the dataset used. Watrous reviews various studies which use auscultation signal of the heart for analysis and provide diagnostics decision support to physicians [19, 20]. Shandilya et al. present their work on the design and development of a nonlinear method for analysis of ventricular fibrillation using ECG signals to predict high yields accuracy for defibrillation success [21]. The study also describes the incorporation of PetCO2 signal to noticeably increase the predictive models robustness.

2.4. Dental Applications. Computerized clinical diagnosis and decision support systems have also seen much success in the field of dentistry. Firestone et al. describe a clinical decision support system on observer performance which was a knowledge-based system performing image analysis on radiographic images [22]. This study involved 102 approximal surface radiographic images and sixteen general practitioners for identifying the presence of caries and whether restoration was required. The paper states that those dental practitioners who used the system to produce their diagnoses showed significant increases in their ability to diagnose caries correctly, with an increased overall diagnostic accuracy and recommendation for restoration of detected cavitated surfaces. Similarly, Olsen et al. propose a computer-aided caries detection system using image analysis of data from intraoral cameras [23]. This paper describes a feasibility study of using advanced image processing and machine learning techniques to identify caries from digital images.

2.5. Cancer. Biomedical informatics has begun to play an important role in cancer detection and treatment. In a study conducted by Lisboa and Taktak, a systematic review of several studies involving decision-making tools in the field of cancer is presented [24]. In particular, the review focuses on those studies that apply artificial neural network methods. Using 27 studies which were either clinical trials or randomized controlled trials, the paper reports that 21 of those studies show benefits in treatment while the remaining 6 did not. Another study by Jesneck discusses an approach to optimize computer-aided decision-making for cancer diagnosis by combining heterogeneous information from different modalities [25]. The authors claim that their proposed method at times outperforms two popular machine-learning techniques, that is, linear discriminant analysis and artificial neural networks. A study by Madabhushi et al. briefly discusses four different computer-aided support systems for cancer diagnosis and prognosis [26]. The first system is an image-based risk score algorithm for predicting the outcome of the estrogen receptor marker for breast cancer patients based on digitized biopsy. The second system is discussed in the paper segments and determines the extent of lymphocytic infiltration from digitized histopathology. The third method described distinguishes patients with varying Gleason grades of prostate cancer, from needle biopsy specimens. The final system integrates quantitative image features extracted from digitized histopathology with protein expression measurements obtained from mass spectrometry, in order to distinguish between low and high risk patients with prostate cancer recurrence following radical prostatectomy. Jiang et al. published a paper evaluating the reduction of interobserver variability in the interpretation of mammograms while using computer-aided diagnosis tools [27]. The authors state that using computer-aided diagnosis tools has the potential to reduce variability amongst expert opinions as well as improve diagnostic accuracy for the interpretation of mammograms. Similarly, another study by Cheng et al. summarizes and compares the methods used in various enhancement and segmentation algorithms, mammographic feature extraction, classifiers, and their performances for detection and classification of microcalcification clusters [28]. A paper by Mazurowski et al. describes an optimization framework for improving case-based computer-aided decision systems used for screening mammography [29]. The paper claims that the proposed method significantly improves the overall performance and breast mass detection rates of such systems. Cai et al.'s paper describes a study based on classification of cancer subtypes and survival prediction in diffuse large B-cell lymphoma (DLBCL) using levels of genes [30]. Research by Rangayyan et al. describes refined methodologies that have been developed in computer-aided breast cancer diagnostic systems [31]. The research presents new detection techniques for identifying subtle signs of breast cancer addressing difficult problems such as focal architecture distortion and global bilateral asymmetry.

2.6. Pediatric Medicine. Computer-aided diagnosis and decision support systems have become popular for a variety of applications in neonatal and pediatric care units. A study by Ramnarayan et al. discusses the potential of diagnostic and decision support systems in pediatric settings with a case study of a web-based pediatric differential diagnostic tool [32]. Ramnarayan also explains the various usages of such diagnostic aid systems and outlines its future direction for research in another article [33]. Frizea et al. discuss an artificial intelligence-based system which uses case-based reasoning for estimating medical outcomes and resource utilization. The paper explains how such a system was initially intended for adult ICU care units and then was modified to function in neonatal ICUs. The paper reports that the results from a short clinical pilot study performed in neonatal ICU were very encouraging and captured the interests of physicians for their potential clinical usefulness. Tan et al. published a review paper on clinical decision support systems for neonatal care [34]. The objective of this review was to find whether the use of clinical decision support systems had any effect on the mortality and morbidity rate of newborn infants, and to see if there was any change in the performance of the physicians treating these infants. Mack et al. also published a similar review study of decision support systems available in pediatric intensive care units [35]. The paper provides a look into the factors that are involved in the applications of such systems in pediatric practices, including liability, human factors, audit trails, engineering, and alert fatigue. The paper concludes that selecting and implementing such systems in clinical practice requires a great deal of caution, though when done correctly it has good potential for benefiting and improving clinical practice in pediatric intensive care units.

3. Methodology

There are several fundamental computational methodologies used toward developing these biomedical informatics and computer-aided diagnosis support systems. The types of techniques and methods are based on application areas and required performance metrics. Some of the major aspects of such systems are discussed below.

3.1. Expert Systems, Case-Based Reasoning, and Rule-Based Systems. Methods such as rule-based systems (fuzzy and crisp), expert systems, and case-based reasoning are formed from the knowledge accumulated from experts of a given field. Opinions, diagnoses, and prognoses, among other components, are compiled to form rule-based analysis structures, based on which specific concepts for diagnosis solutions are developed. Kumar et al. present a hybrid decision support system which was designed based on both case-based and rule-based reasoning [16], which is applied to ICU facilities for aiding physicians in decision making. Another study by Innocent describes an approach to computer-aided medical diagnosis systems for clinical contexts using fuzzy logic [36]. In this system, knowledge from experts is compiled into fuzzy cognitive maps and logical structures to estimate a stage of disease using temporal information in symptom durations.

3.2. Signal and Image Processing. Some computerized diagnostic aid systems use a variety of patient data for analysis in developing diagnostic suggestions. These systems analyze raw patient signals and images to extract useful features and trends based on which diagnostic and decision support information is computed and presented to physicians. For instance, Polat et al. describe a signal processing system that analyzes ECG to classify cases of arrhythmia in diseased persons [18]. The signal processing system, developed by Shandillya et al., detects the ideal time to defibrillate patients undergoing cardiac arrest or ventricular fibrillation [21, 37]. Davaluri et al. proposes an image processing system which uses CT images of patients with pelvic injuries to produce a quantitative and qualitative assessment of detected hemorrhaging [38]. Similarly, Wu's work on developing a computer-assisted fracture detection system automatically processes several CT slices of pelvic injury patients to identify and quantify potential fractures [39].

3.3. Machine Learning. Due to the continuous advancements in the field of machine learning, more complex and sophisticated biomedical informatics systems are being designed. Systems that have the ability to predict and classify diseases fundamentally rely on some type of machine learning methodology. There is no one superior machine learning technique that can be applied toward all learning problems; instead the best method depends on the type of application. For example, Lisboa's study provides a systematic review of neural networks in decision support systems for cancer diagnosis and treatment [24]. Jesneck et al. describe how a customized machine learning technique outperforms standard techniques such as artificial neural networks and linear discriminant analysis in their study using cancer datasets [25]. Ji et al. compare a variety of machine learning techniques used in decision-making systems for traumatic injury assessment [40].

4. Impact of Computer-Aided Decisions in Bioinformatics

In the last two decades, bioinformatics has emerged as a vibrant and rapidly growing field. However, as shown above, the majority of computer-aided decision support systems is implementations of biomedical informatics systems, so that very few of the currently used computer-aided support systems are based on bioinformatics approaches, which is understandable given the age of the field.

A study by Maojo et al. provides a comparison of histories, fundamental foundations, and scientific approaches of the two complementary yet separate fields, that is, of medical informatics and bioinformatics [41]. With most computerized clinical diagnostic aids being developed under the umbrella of biomedical informatics, Maojo et al. explain how inclusion of knowledge from bioinformatics can strengthen applications development for healthcare. The authors emphasize that future research designed as a hybrid of both informatics subdisciplines is the key to making significant advances in clinical practice and biomedical research.

The effort to combine multimodal data and to combine biomedical informatics and bioinformatics has already shown a great promise. As mentioned, Madabhushi et al. describe research on computer-aided prognosis and diagnosis systems using multi-modal data fusion, including computerized image analysis and digitized patient data such as tissue and genomic information for predicting outcomes and survival [5]. These projects use protein expression and other data, processed by typical biomedical informatics methods, to diagnose and develop prognoses for cancer cases. Huang et al. analyzed and published a time series microarray gene expression profiles dataset to predict how patients respond to pegylated interferon treatment [42, 43]. Computer-aided decision systems adapted with bioinformatics knowledge have begun to show positive impact on virological research. For instance a paper by Huang et al. describes a computational method in identifying the underlying mechanisms for HIV-1 resistance in some people based on gene expression profiles and the analysis of the network of virus-host interaction [44]. Similarly, another study describes a novel approach in diagnosing liver cirrhosis and hepatocellular diseases using a network based analysis [45].

5. Validation and Criteria for Success

With numerous clinical implementations of decision support systems for a variety of medical applications, it is essential to have a systematic method to verify, validate, and compare different systems and their performances. For instance, Berner et al. compare the performance of four computer-based diagnostic systems applied towards internal medicine applications, namely: Dxplain, Iliad, Meditel, and QMR [46]. These systems have all been noted in various publications in their phases of development, evaluation, and applications [47, 48]. The authors have tested these systems on identical diagnostically challenging cases and measured the performances of each of these systems on several developed measurement scales. Estimates of performance were provided with a prospectively determined set of test specifications, using cases with a range of content and difficulty. Another study by Manotti et al. assesses the performance of another decision support system pertinent to oral anticoagulant treatment [49]. In this paper the authors describe a clinical trial of

Table 1: Stregnths and weaknesses of existing computer-aided decision support systems and research in different application areas.

Application areas	Strengths	Weaknesses
Cancer	(i) An abundance of molecular assays and data are available for many cancer cases; these can be used towards developing strong decision support systems	(i) More should be done to integrate knowledge from molecular-based and image-based sources available for cancer detection (ii) There is a need to develop better schemes and methods for validating the effectiveness of the existing and upcoming systems in this area
Radiology	(i) A variety of effective computational techniques exists for many applications in radiology (ii) It is one of the fastest growing fields using applications of computer-aided decision systems	(i) Most of the research in this area suffers from lack of comprehensive datasets (ii) Most of these studies do not include knowledge of illness/injury/complication into the decision-making process
Emergency medicine	(i) Although there are only a few systems that have been adopted into clinical practices, the existing systems have shown a positive impact on the cost and quality of healthcare (ii) There is a significant potential for computer-aided systems in this area since emergency medicine and trauma are very time and resource critical aspects of healthcare	(i) Accuracies of existing systems may not be sufficient for clinical uses (ii) A variety of illnesses and injuries have not yet been addressed by computer-aided decision support systems (iii) There is a lack of comprehensive validation of the short-/long-term impacts on these systems using sufficiently large datasets
Cardiovascular medicine	(i) Since heart disease is among the leading causes of death, computer-aided decision systems here have potentially very high impact on world health (ii) While most cardiovascular-based intelligent decision support systems suffer from high false positives, they often help detect disease at early stages	(i) These systems usually incorporate only a portion of available patient information. More variety in information sources may be required in the decision-making process to reduce false positives (ii) There is a lack of a comprehensive validation process. Existing research claims need to be tested in more real-world settings
Dental	(i) Existing systems have shown capability for detecting dental complications at early stages (ii) Such early detection facilitates better practice of preventive care	(i) Some of the technologies used for capturing the information for computer-aided decision support systems are relatively expensive and hence preventing them from being widely adopted in practice

the system with several patients across multiple clinics, to test whether the computer-based decision support system is efficient in stabilizing patients undergoing oral anticoagulant treatment by initiating and maintaining therapy. With statistical analysis of performance measures the paper reports that the decision support system improves the quality of anticoagulant treatment, both during long-term treatments and in early, unstable phases of treatment.

Several publications also explain the various criteria that need to be considered for successful development and application of a computer-assisted decision support system. Along these lines, Kaplan reviews the literature related to clinical decision support systems with an emphasis on evaluation criteria [50]. In the paper the author explains that with the success seen so far there is a general enthusiasm amongst physicians and researchers with the potential of computerized clinical decision support systems to improve the quality of healthcare. Nonetheless, there is a lack of theoretical understanding especially from a nonphysician's perspective of such systems and also as to why certain diagnostic aid systems may not be effective. Similarly Dreiseitl and Binder consider the effects of decision support systems on physicians' opinions, in particular to see whether they, doctors, value its opinion when it contradicts theirs [50]. They conclude that physicians are fairly susceptible to accepting recommendations of such decision support systems, making quality assurance and validation of more paramount importance. Ramnarayan et al. highlight the importance of developing a reliable and valid composite scoring system to measure the impact of diagnostic decision support on the quality of healthcare [32]. They claim that the scoring systems they describe can be further used in assessing outcome measures of other study types, involving computer-assisted diagnostic systems. Song et al. discuss the various approaches, goals, and characteristics of computer-aided healthcare workflows [51]. The authors analyze the workflow application issues and software challenges in the perspective of medical informatics and software engineering. Niès et al. published a paper listing four key characteristics pertaining to the content of diagnosis support that are associated with the success of computerized clinical decision support systems [52]. The paper provides a systematic review of published trials to identify the characteristics of the adopted methodologies and technicalities of those studies that assess the efficacy of clinical decision support systems.

6. Conclusion and Future Directions

Table 1 describes the overall strengths and weakness of existing computer-aided decision support systems, and research in some of the application areas discussed in this paper.

With the sheer number of biomedical informatics methods implemented as computer-assisted diagnosis and decision support systems, along with the vast amount of research in this field, such systems are inevitably becoming an inherent part of medicine. The systems are becoming capable of solving more complex and sophisticated clinical problems. By establishing systematic processes for validation and verification, these computer-aided systems can become much more reliable and thereby improve quality of diagnostic decisions, as well as reduce variance among physicians' opinions. The unique capabilities of these systems allow care givers and researchers to gain insight into current clinical issues in ways that would have been impossible in the past.

Furthermore, it is becoming advantageous to fuse information derived from medical data with multiple modalities to provide more robust diagnoses and treatment plan suggestions [5, 10, 40]. The current fusion of biomedical informatics and bioinformatics techniques will accelerate the formation of a new generation of system-biologic computer-aided decision support systems, that will process and combine information in molecular data, signals and images, and demographics, among others. These and many other sources of patient data will allow such systems to form much more specific and personalized recommendations.

Applying advances in computational methods and techniques towards such systems can help in problems such as overfitting of outputs towards specific types of data, susceptibility to incomplete/missing data, and presence of conflicting information from different sources. These advances in the computational methods can also improve the quality of information accessed from feature extraction and feature selection—this improvement is often a critical step prior to classification and/or clustering.

While computerized diagnostic and prognostic decision support systems have proved to be instrumental in medicine, it appears that an even more significant contribution of these systems can be expected when they further evolve to process and integrate newer and even broader types of patient data.

References

[1] R. A. Miller, "Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary," *Journal of the American Medical Informatics Association*, vol. 1, no. 1, pp. 8–27, 1994.

[2] F. T. de Dombal, D. J. Leaper, J. R. Staniland, A. P. McCann, and J. C. Horrocks, "Computer-aided diagnosis of acute abdominal pain," *British Medical Journal*, vol. 2, no. 5804, pp. 9–13, 1972.

[3] J. D. Myers, "The background of INTERNIST I and QMR," in *Proceedings of ACM Conference on History of Medical Informatics*, pp. 195–197, 1987.

[4] R. A. Greenes, *Clinical Decision Support: The Road Ahead*, Academic Press, 2007.

[5] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle, and G. Lee, "Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 506–514, 2011.

[6] S. A. Pearson, A. Moxey, J. Robertson et al., "Do computerised clinical decision support systems for prescribing change practice? A systematic review of the literature (1990–2007)," *BMC Health Services Research*, vol. 9, article 154, 2009.

[7] G. D. Tourassi, "Journey toward computer-aided diagnosis: role of image texture analysis," *Radiology*, vol. 213, no. 2, pp. 317–320, 1999.

[8] S. M. Stivaros, A. Gledson, G. Nenadic, X. J. Zeng, J. Keane, and A. Jackson, "Decision support systems for clinical radiological practice—towards the next generation," *British Journal of Radiology*, vol. 83, no. 995, pp. 904–914, 2010.

[9] B. Van Ginneken, B. M. Ter Haar Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.

[10] W. Chen, C. Cockrell, K. R. Ward, and K. Najarian, "Intracranial pressure level prediction in traumatic brain injury by extracting features from multiple sources and using machine learning methods," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM '10)*, pp. 510–515, December 2010.

[11] W. Chen, C. Cockrell, K. Ward, and K. Najarian, "Predictability of intracranial pressure level in traumatic brain injury: features extraction,statistical analysis and machine learning based evaluation," *Journal of Data Mining and Bioinformatics*. In press.

[12] P. Davuluri, J. Wu, K. R. Ward, C. H. Cockrell, K. Najarian, and R. S. Hobson, "An automated method for hemorrhage detection in traumatic pelvic injuries," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11)*, pp. 5108–5111, 2011.

[13] J. Wu, Y. Tang, P. Davuluri et al., "Fracture detection and quantitative measure of displacement in pelvic CT images," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '11)*, pp. 600–606, 2011.

[14] S. Y. Ji, R. Smith, T. Huynh, and K. Najarian, "A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries," *BMC Medical Informatics and Decision Making*, vol. 9, no. 1, article 2, 2009.

[15] M. Frize and R. Walker, "Clinical decision-support systems for intensive care units using case-based reasoning," *Medical Engineering and Physics*, vol. 22, no. 9, pp. 671–677, 2000.

[16] K. A. Kumar, Y. Singh, and S. Sanyal, "Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU," *Expert Systems with Applications*, vol. 36, no. 1, pp. 65–71, 2009.

[17] R. A. Raschke, B. Gollihare, T. A. Wunderlich et al., "A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital," *Journal of the American Medical Association*, vol. 280, no. 15, pp. 1317–1320, 1998.

[18] K. Polat, B. Akdemir, and S. Güneş, "Computer aided diagnosis of ECG data on the least square support vector machine," *Digital Signal Processing*, vol. 18, no. 1, pp. 25–32, 2008.

[19] R. L. Watrous, "Computer-aided auscultation of the heart: from anatomy and physiology to diagnostic decision support," in

Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06), pp. 140–143, September 2006.

[20] R. L. Watrous, W. R. Thompson, and S. J. Ackerman, "The impact of computer-assisted auscultation on physician referrals of asymptomatic patients with heart murmurs," *Clinical Cardiology*, vol. 31, no. 2, pp. 79–83, 2008.

[21] S. Shandilya, K. Ward, M. Kurz, and K. Najarian, "Non-linear dynamical signal characterization for prediction of defibrillation success through machine learning," *BMC Medical Informatics and Decision Making*, vol. 12, p. 116, 2012.

[22] A. R. Firestone, D. Sema, T. J. Heaven, and R. A. Weems, "The effect of a knowledge-based, image analysis and clinical decision support system on observer performance in the diagnosis of approximal caries from radiographic images," *Caries Research*, vol. 32, no. 2, pp. 127–134, 1998.

[23] G. F. Olsen, S. S. Brilliant, D. Primeaux, and K. Najarian, "An image-processing enabled dental caries detection system," in *Proceedings of the International Conference on Complex Medical Engineering (ICME '09)*, pp. 1–8, April 2009.

[24] P. J. Lisboa and A. F. G. Taktak, "The use of artificial neural networks in decision support in cancer: a systematic review," *Neural Networks*, vol. 19, no. 4, pp. 408–415, 2006.

[25] J. L. Jesneck, *Optimized Decision Fusion of Heterogeneous Data for Breast Cancer Diagnosis*, 2007.

[26] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle, and G. Lee, "Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 506–514, 2011.

[27] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, A. Y. Toledano, and K. Doi, "Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications," *Radiology*, vol. 220, no. 3, pp. 787–794, 2001.

[28] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, no. 12, pp. 2967–2991, 2003.

[29] M. A. Mazurowski, P. A. Habas, J. M. Zurada, and G. D. Tourassi, "Decision optimization of case-based computer-aided decision systems using genetic algorithms with application to mammography," *Physics in Medicine and Biology*, vol. 53, no. 4, pp. 895–908, 2008.

[30] Y. D. Cai, T. Huang, K. Y. Feng, L. Hu, and L. Xie, "A unified 35-gene signature for both subtype classification and survival prediction in diffuse large B-cell lymphomas," *PloS one*, vol. 5, no. 9, p. e12726, 2010.

[31] R. M. Rangayyan, F. J. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs," *Journal of the Franklin Institute*, vol. 344, no. 3-4, pp. 312–348, 2007.

[32] P. Ramnarayan, R. R. Kapoor, M. Coren et al., "Measuring the impact of diagnostic decision support on the quality of clinical decision making: development of a reliable and valid composite score," *Journal of the American Medical Informatics Association*, vol. 10, no. 6, pp. 563–572, 2003.

[33] P. Ramnarayan and J. Britto, "Paediatric clinical decision support systems," *Archives of Disease in Childhood*, vol. 87, no. 5, pp. 361–362, 2002.

[34] K. Tan, P. R. Dear, and S. J. Newell, "Clinical decision support systems for neonatal care," *Cochrane Database of Systematic Reviews*, no. 2, Article ID CD004211, 2005.

[35] E. H. Mack, D. S. Wheeler, and P. J. Embi, "Clinical decision support systems in the pediatric intensive care unit," *Pediatric Critical Care Medicine*, vol. 10, no. 1, pp. 23–28, 2009.

[36] P. R. Innocent and R. I. John, "Computer aided fuzzy medical diagnosis," *Information Sciences*, vol. 162, no. 2, pp. 81–104, 2004.

[37] S. Shandilya, K. R. Ward, and K. Najarian, "A time-series approach for shock outcome prediction using machine learning," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '10)*, pp. 440–446, December 2010.

[38] P. Davuluri, Y. Tang, J. Wu et al., "A hybrid approach for hemorrhage segmentation in pelvic CT scans," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '11)*, pp. 548–554, 2011.

[39] J. Wu, P. Davuluri, K. R. Ward, C. Cockrell, R. Hobson, and K. Najarian, "Fracture detection in traumatic pelvic CT images," *Journal of Biomedical Imaging*, vol. 2012, Article ID 327198, 10 pages, 2012.

[40] S. Y. Ji, R. Smith, T. Huynh, and K. Najarian, "A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries," *BMC Medical Informatics and Decision Making*, vol. 9, no. 1, article 2, 2009.

[41] V. Maojo and C. A. Kulikowski, "Bioinformatics and medical informatics: collaborations on the road to genomic medicine?" *Journal of the American Medical Informatics Association*, vol. 10, no. 6, pp. 515–522, 2003.

[42] T. Huang, K. Tu, Y. Shyr, C. C. Wei, L. Xie, and Y. X. Li, "The prediction of interferon treatment effects based on time series microarray gene expression profiles," *Journal of Translational Medicine*, vol. 6, article 44, 2008.

[43] X. Zhang, C. Chen, M. Wu et al., "Plasma microRNA profile as a predictor of early virological response to interferon treatment in chronic hepatitis B patients," *Antiviral Therapy*, vol. 17, pp. 1243–1253, 2012.

[44] T. Huang, Z. Xu, L. Chen, Y. D. Cai, and X. Kong, "Computational analysis of HIV-1 resistance based on gene expression profiles and the virus-host interaction network," *PLoS ONE*, vol. 6, no. 3, Article ID e17291, 2011.

[45] T. Huang, J. Wang, Y. D. Cai, H. Yu, and K. C. Chou, "Hepatitis C virus network based classification of hepatocellular cirrhosis and carcinoma," *PloS One*, vol. 7, Article ID e34460, 2012.

[46] E. S. Berner, G. D. Webster, A. A. Shugerman et al., "Performance of four computer-based diagnostic systems," *The New England Journal of Medicine*, vol. 330, no. 25, pp. 1792–1796, 1994.

[47] G. O. Barnett, J. J. Cimino, J. A. Hupp, and E. P. Hoffer, "DXplain. An evolving diagnostic decision-support system," *Journal of the American Medical Association*, vol. 258, no. 1, pp. 67–74, 1987.

[48] R. Miller and F. Masarie Jr Jr., "The quick medical reference (QMR) relationships function: description and evaluation of a simple, efficient "multiple diagnoses" algorithm," in *Proceedings of the World Conference on Medical Informatics (Medinfo '92)*, pp. 512–518, 1992.

[49] C. Manotti, M. Moia, G. Palareti, V. Pengo, L. Ria, and A. G. Dettori, "Effect of computer-aided management on the quality of treatment in anticoagulated patients: a prospective, randomized, multicenter trial of APROAT (Automated Program for

Oral Anticoagulant Treatment)," *Haematologica*, vol. 86, no. 10, pp. 1060–1070, 2001.

[50] B. Kaplan, "Evaluating informatics applications—clinical decision support systems literature review," *International Journal of Medical Informatics*, vol. 64, pp. 15–37, 2001.

[51] X. Song, B. Hwong, G. Matos et al., "Understanding requirements for computer-aided healthcare workflows: experiences and challenges," in *Proceedings of the 28th International Conference on Software Engineering (ICSE '06)*, pp. 930–934, May 2006.

[52] J. Niès, I. Colombet, P. Degoulet, and P. Durieux, "Determinants of success for computerized clinical decision support systems integrated into CPOE systems: a systematic review," in *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA '06)*, p. 594, 2006.

The Scientific World Journal
Volume 2012, Article ID 989637, 9 pages
doi:10.1100/2012/989637

The ScientificWorldJOURNAL

Research Article

Robust Microarray Meta-Analysis Identifies Differentially Expressed Genes for Clinical Prediction

John H. Phan,[1] Andrew N. Young,[2] and May D. Wang[1]

[1] *Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 313 Ferst Drive, Atlanta, GA 30332, USA*
[2] *Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Grady Health System, Grady Memorial Hospital, Atlanta, GA 30303, USA*

Correspondence should be addressed to May D. Wang, maywang@bme.gatech.edu

Received 2 November 2012; Accepted 28 November 2012

Academic Editors: N. S. T. Hirata, M. A. Kon, and K. Najarian

Combining multiple microarray datasets increases sample size and leads to improved reproducibility in identification of informative genes and subsequent clinical prediction. Although microarrays have increased the rate of genomic data collection, sample size is still a major issue when identifying informative genetic biomarkers. Because of this, feature selection methods often suffer from false discoveries, resulting in poorly performing predictive models. We develop a simple meta-analysis-based feature selection method that captures the knowledge in each individual dataset and combines the results using a simple rank average. In a comprehensive study that measures robustness in terms of clinical application (i.e., breast, renal, and pancreatic cancer), microarray platform heterogeneity, and classifier (i.e., logistic regression, diagonal LDA, and linear SVM), we compare the rank average meta-analysis method to five other meta-analysis methods. Results indicate that rank average meta-analysis consistently performs well compared to five other meta-analysis methods.

1. Introduction

We develop a simple, yet robust meta-analysis-based feature selection (FS) method for microarrays that ranks genes by differential expression within several independent datasets,then combines the ranks using a simple average to produce a final list of rank-ordered genes. Such meta-analysis methods can increase the power of microarray data analysis by increasing sample size [1]. The subsequent improvement to differentially expressed gene (DEG) detection, or to FS is essential for downstream clinical applications. Many of these applications, such as disease diagnosis and disease subtyping, are predictive in nature and are important for guiding therapy. However, DEG detection can be difficult due to technical and biological noise or due to small sample sizes relative to large feature sizes [2]. These properties are typical of many microarray datasets. Despite small sample sizes, the number of gene expression datasets available to the research community has grown [3]. Thus, it is important to develop methods that can use all available knowledge by simultaneously analyzing several microarray datasets of similar clinical focus. However, combining high-throughput gene expression datasets can be difficult due to technological variability. Differences in microarray platform [4] or normalization and preprocessing methods [5] affect the comparability of gene expression values. Laboratory batch effects can also affect reproducibility [6]. Numerous studies have proposed novel strategies to remove batch effects [7]. However, in some cases, batch effect correction can have undesirable consequences [8]. In light of these challenges, several studies have proposed novel methods for meta-analysis of multiple microarray datasets.

Existing microarray meta-analysis methods either combine separate statistics for each gene expression dataset or aggregate samples into a single large dataset to estimate global gene expression. The study by Park et al. used analysis of variance to identify unwanted effects (e.g., the effect of different laboratories) and modeled these effects to detect DEGs [9]. Choi et al. used a similar approach to compute an "effect size" quantity, representing a measure of precision for

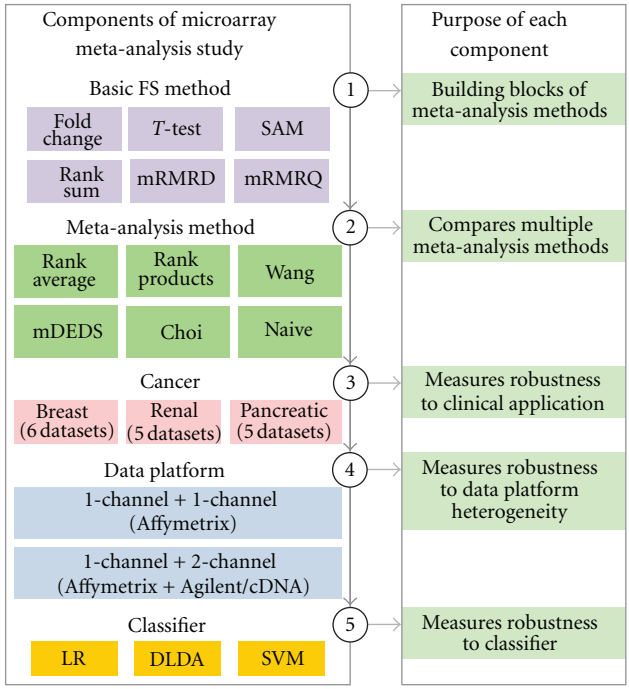


FIGURE 1: Study design diagram. We compare the predictive performance of meta-analysis-based feature selection (FS) methods by designing a study that considers five components: (1) basic FS methods that are the building blocks of some of the meta-analysis methods, (2) meta-analysis-based FS methods, (3) clinical application, (4) microarray data platform, and (5) classifier (logistic regression, diagonal LDA and linear SVM). Since the "best" meta-analysis-based FS method may be dataset- or application-specific, assessing performance over a wide variety of factors enables an evaluation of the method's robustness.

each study, and used this "effect size" to directly compare and combine microarray datasets [10]. Wang et al. combined the fold change of genes between classes from three microarray datasets and weighted each dataset by its variance such that datasets with higher variance contribute less to the final statistic [11]. Yoon et al. conducted a large-scale study of gene expression by examining the variation of genes across multiple microarray datasets, regardless of the clinical focus [12]. Breitling and Herzyk ranked fold changes between all interclass pairs of samples and computed the product of all ranks for each gene [13]. More recently, Campain and Yang reviewed several meta-analysis methods and assessed their performance using both classification accuracy and synthetic data [14]. Research has shifted towards methods that consider multiple FS methods, reflecting the fact that no single FS method performs well for all datasets [15]. Although several meta-analysis methods exist, except for the study by Campain and Yang, the literature rarely compares these methods in a comprehensive manner.

We develop the rank average method, a simple meta-analysis-based FS method, for identifying DEGs from multiple microarray datasets and design a study (Figure 1) to compare rank average to five other meta-analysis-based FS methods. We focus on the predictive ability of genes emerging from meta-analysis and show that rank average meta-analysis is robust with respect to three factors. These three factors are (1) clinical application (i.e., breast, renal, and pancreatic cancer diagnosis or subtyping), (2) data platform heterogeneity (i.e., combining different microarray platforms), and (3) classifier. Using a comprehensive factorial analysis, we rate each meta-analysis-based FS method relative to its peers. In terms of identifying genetic features with reproducible predictive performance and in terms of robustness to multiple factors, results indicate that rank average meta-analysis performs consistently well in comparison to five other meta-analysis-based FS methods.

2. Methods

2.1. Microarray Datasets. We use six breast cancer, five renal cancer, and five pancreatic cancer gene expression datasets (Table 1) to compare meta-analysis-based FS methods. Each renal cancer dataset examines patient samples from several subtypes of tumors: clear cell (CC), oncocytoma (ONC), chromophobe (CHR), and papillary (PAP). We are interested in identifying genes differentially expressed between the CC subtype and all other subtypes, that is, CC versus ONC/CHR/PAP. These renal cancer datasets share a similar clinical focus. However, they are heterogeneous in terms of microarray platform [16–21]. Similarly, the breast cancer datasets are heterogeneous in both platform and clinical focus [22–26]. Although patient samples from each dataset have undergone different treatment for breast cancer and have been extracted at different stages of the disease, each sample is labeled as either estrogen receptor positive (ER+) or negative (ER−). Thus, we assess the performance of classifiers that predict the estrogen receptor status. The pancreatic cancer datasets also include a variety of platforms and clinical focuses [27–31]. We identify genes to discriminate pancreatic cancer versus noncancer patient samples. These datasets contain different numbers of probes (or probesets in the case of Affymetrix datasets) due to differences in microarray platform. Within each dataset group, we reduce the number of probes in each dataset to a common shared set based on probe sequence similarity.

2.2. Rank Average Meta-Analysis. The meta-analysis-based FS method proposed in this paper ranks genes individually in each dataset and computes the average rank of each gene. Gene rank order is determined by a measure of differential expression (which can be any of a number of basic FS methods such as fold change or t-test) and we assume that this rank order is invariant to batch effects. Using the average rank of a gene across several datasets to obtain the final multidataset rank order, we can infer (1) the relative strength of that gene in differentiating the patient samples of interest and (2) the consistency of the gene's differential expression across multiple studies.

The remainder of this section uses the following mathematical notation. K is the total number of datasets, M is the total number of genes in each dataset, and N_k is the number of samples in dataset k, where $k = 1 \cdots K$ and N is the total

Table 1: Microarray datasets.

(a) Breast cancer estrogen receptor status

Dataset	ER+	ER−	Platform	No. of probes
MDACC Train	80	50	Affy HG-U133A	22283
MDACC Test	60	40	Affy HG-U133A	22283
Miller	213	34	Affy HG-U133A	22283
Sotiriou	72	24	Affy HG-U133A	22283
Minn	57	42	Affy HG-U133A	22283
Van't Veer	226	69	Agilent 2-Color	24496

Common probes: 8953.

(b) Renal cancer subtype

Dataset	CC	Other	Platform	No. of probes
Schuetz	13	12	Affy HG-Focus	8793
Jones	32	29	Affy HG-U133A	22283
Kort	10	30	Affy HG-U133+2.0	54675
Yusenko	26	27	Affy HG-U133+2.0	54675
Higgins	26	9	cDNA 2-Color	22689

Common probes: 946.

(c) Pancreatic cancer diagnosis

Dataset	Normal	Cancer	Platform	No. of probes
Badea	39	39	Affy HG-U133+2.0	54675
Ishikawa	25	24	Affy HG-U133A/B	44928
Pei	16	36	Affy HG-U133+2.0	54675
Pilarsky	18	27	Affy HG-U133A/B	44928
Iacobuzio-Donahue	5	17	cDNA 2-Color	43910

Common probes: 4530.

number of samples in all datasets. We denote a gene i in dataset k as a vector

$$\overline{g}_{i,k} = \left(x_1^{i,k}, x_2^{i,k}, \ldots, x_{N_k}^{i,k}\right), \tag{1}$$

where $x_j^{i,k}$ is the expression value of gene i of sample j in dataset k. In the case of sample aggregation (i.e., the naive method of meta-analysis), we denote a gene i across all datasets with

$$\overline{g}_{i,\bullet} = \left(\left(x_1^{i,1}, x_2^{i,1}, \ldots, x_{N_1}^{i,1}\right), \left(x_1^{i,2}, x_2^{i,2}, \ldots, x_{N_2}^{i,2}\right), \ldots, \left(x_1^{i,K}, x_2^{i,K}, \ldots, x_{N_K}^{i,K}\right)\right). \tag{2}$$

Using this notation, we can define a function, $r_{i,k,\vartheta} = R_\vartheta(\overline{g}_{i,k})$, to compute the rank, $r_{i,k,\vartheta}$, of a gene, $\overline{g}_{i,k}$, using a ranking algorithm denoted by ϑ. A smaller rank indicates a greater degree of differential expression. In the case of sample aggregation, the ranking function takes the form $r_{i,\bullet,\vartheta} = R_\vartheta(\overline{g}_{i,\bullet})$. The average rank, $\overline{r}_i$, of a gene i across all datasets, weighted by number of samples in each dataset, N_k, is

$$\overline{r}_i = \frac{1}{N}\sum_{k=1}^{K} N_k R_{\vartheta_k}\left(\overline{g}_{i,k}\right). \tag{3}$$

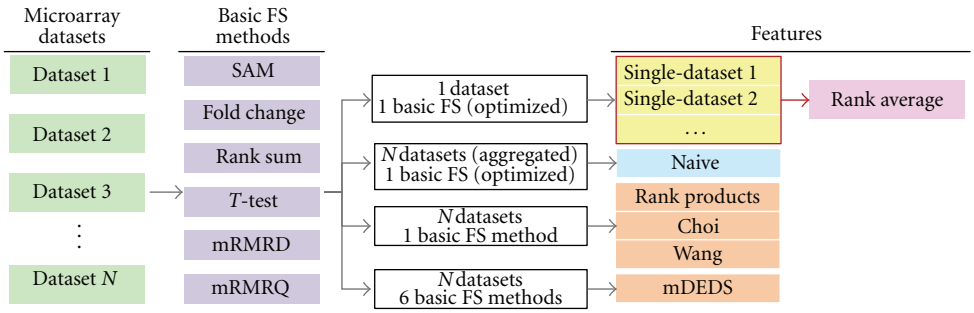
Weighting gives preference to ranks from datasets with larger sample sizes.

We consider several basic FS, or gene ranking, methods as follows: fold change (FC), t-test (T), significance analysis of microarrays (SAM) [32], rank-sum (RS), minimum redundancy maximum relevance using the difference formulation (mRMRD), and mRMR using the quotient formulation (mRMRQ) [33]. We explicitly define the rank algorithm for the kth dataset as

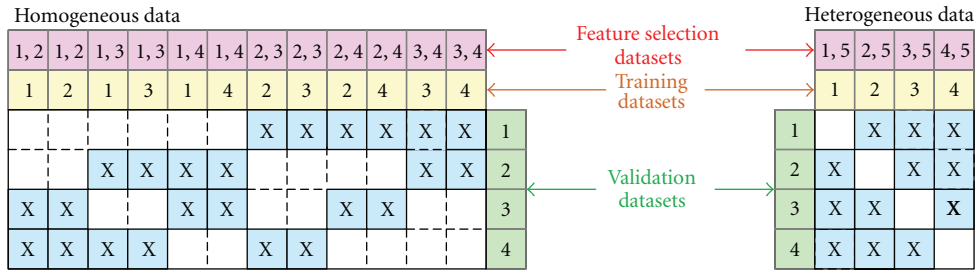
$$\vartheta_k \in \{\text{FC}, T, \text{SAM}, \text{RS}, \text{mRMRD}, \text{mRMRQ}\}. \tag{4}$$

For each dataset and each basic FS method, we use three-fold cross-validation to compute an estimate of classification performance (measured using AUC) averaged over 20 feature sizes (ranging from the top single feature to the top twenty features). We then choose the basic FS method, ϑ_k, with highest estimated classification performance for each dataset. Because each basic FS method makes different assumptions about DEGs and the correctness of these assumptions varies from dataset to dataset, allowing a different basic FS method for each dataset can improve performance.

2.3. Predictive Performance. We use classification performance to assess meta-analysis-based FS methods with the assumption that improved FS leads to higher prediction performance when classifying samples from an independent



(a) Selecting features from multiple microarray datasets using six meta-analysis-based methods



(b) Example of dataset permutations for evaluating meta-analysis predictive performance

Figure 2: Procedure for comparing the predictive performance of six microarray meta-analysis-based FS methods. (a) Features are selected from microarray datasets using the rank average meta-analysis method (pink box), several other meta-analysis methods (orange boxes: mDEDS, rank products, Choi, and Wang), and a naive method (blue box) that aggregates samples into a larger dataset. Rank average meta-analysis chooses a single feature selection (FS) method from among several basic FS methods (SAM, fold change, rank sum, *t*-test, mRMRD, and mRMRQ) for each individual dataset that optimizes prediction performance (via cross-validation) over the top 20 features. A simple weighted average of gene ranks from all individual datasets produces the final set of rank average meta-analysis features. The rank products, Choi, and Wang methods use one basic FS method to select features from multiple datasets while the mDEDS method uses all six basic FS methods. (b) Features are selected from two or more datasets from each group to build a classifier (pink boxes), which is trained with samples from only one dataset (yellow boxes). The performance of the classifier is assessed using independent datasets (datasets not used for training or feature selection, green boxes). The predictive performance of a microarray meta-analysis-based FS method is an average over all permutations of training and validation datasets (blue boxes). In the example, datasets 1–4 consist of one-channel Affymetrix arrays while dataset 5 (in the case of heterogeneous data) consists of two-channel arrays.

dataset. We assess prediction performance using independent training and testing datasets because of the small sample size of some of the datasets and because we want to reflect clinical scenarios in which predictive models would likely be derived from data collected from a separate batch of patients. We compare our proposed rank average meta-analysis method to other meta-analysis methods including: (1) the rank products method [13], (2) the mDEDS method [14], (3) Choi et al.'s method of interstudy variability [10], (4) Wang et al.'s method of weighting differential expression by variance [11], and (5) a naive method that aggregates samples from multiple datasets. The rank products, mDEDS, Choi, and Wang methods can be applied to multiple datasets as well as to single datasets. For each method and each dataset group, we compute single-dataset performance, combined homogeneous-dataset performance (from two to four datasets combined), and combined heterogeneous-dataset performance (Figure 2(a)).

Classification performance depends on both feature selection and number of samples available for training. We are interested in performance gains due to meta-analysis-based FS alone. We isolate this performance gain by training classifiers with samples from a single dataset only, while allowing the features used for training to come from multiple datasets. Thus, any improvement (or degradation) in classification performance of a meta-analysis-based FS method in comparison to the baseline single-dataset FS is due to features selected rather than to increases in training sample size. We assess classification performance using a separate validation dataset and permute the datasets such that each individual dataset in each dataset group—renal, breast, and pancreatic cancer—is used at least once for validation. Moreover, for each permutation, we use 100 iterations of bootstrap sampling from the training datasets to estimate classification performance. Figure 2(b) is an example of the permutations possible with a five-dataset group (datasets 1–4 are the same platform while dataset 5 is a different platform), in which the prediction performance of two-dataset combination is assessed. This procedure can be expanded to handle three-dataset, four-dataset, or higher combinations for FS.

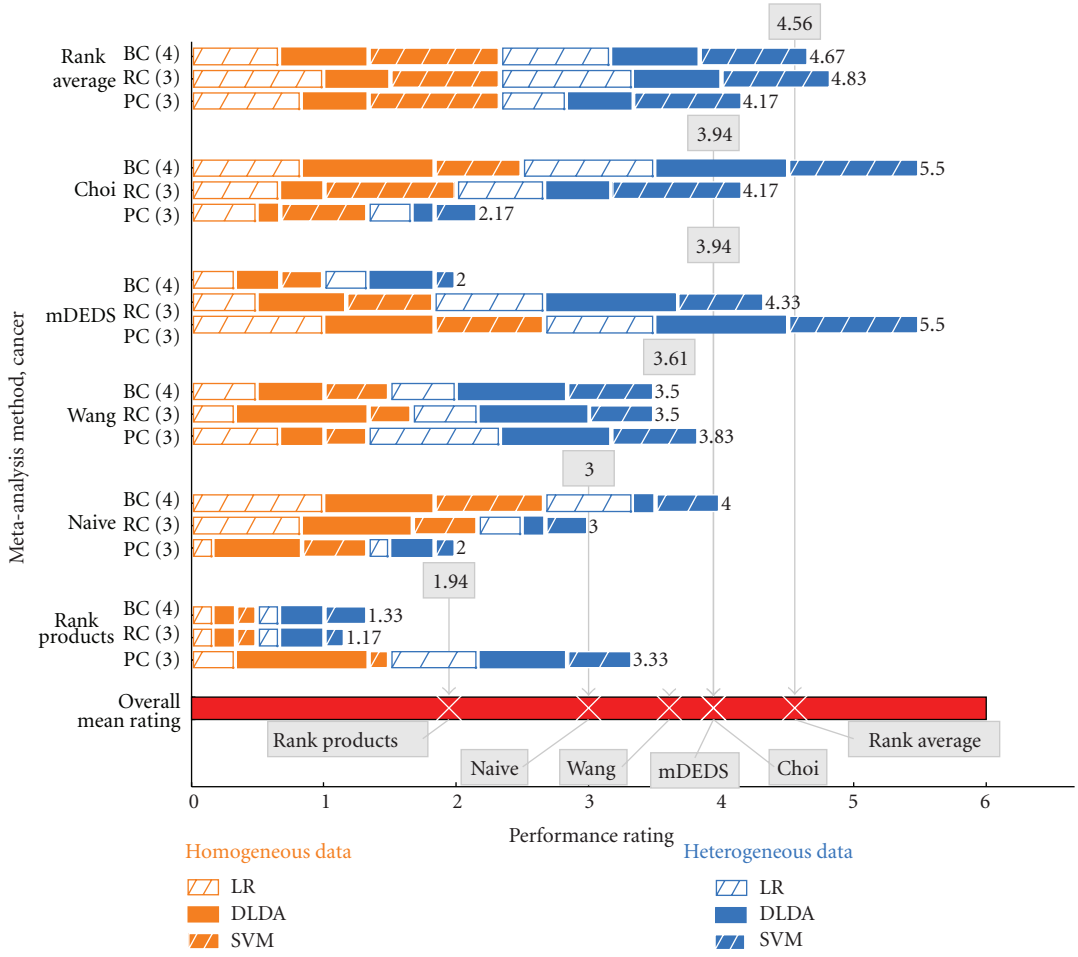


Figure 3: Rating meta-analysis methods by prediction performance when combining all available datasets. Each meta-analysis method (rank average, rank products, Wang, mDEDS, Choi, and naive) is rated relative to its peers. We assess performance rating across three factors: (1) clinical application (breast cancer: BC, renal cancer: RC, and pancreatic cancer: PC), (2) data platform heterogeneity (homogeneous: orange, heterogeneous: blue), and (3) classifier (logistic regression: LR, diagonal LDA: DLDA and linear SVM). For each combination of factors, the rating of each meta-analysis method is represented by an additive bar. Methods with higher absolute prediction performance receive higher ratings (and longer bars). When considering absolute prediction performance, rank average, with a mean overall rating of 4.56, performs consistently well compared to its peers.

The procedure for measuring predictive performance of heterogeneous-dataset combination is slightly different. Each dataset group contains several one-channel Affymetrix datasets and one two-channel dataset (either cDNA or Agilent). Gene expression values of the two-channel datasets are computed as log ratios, resulting in different dynamic ranges compared to the one-channel datasets. We assess the robustness of each meta-analysis-based FS method to heterogeneous data platforms by first determining the performance of the method when combining only Affymetrix data (Figure 2(b), homogeneous data), then comparing to results obtained when combining a mixture of Affymetrix and two-channel arrays (Figure 2(b), Heterogeneous Data). For example, we compute heterogeneous combination performance by combining one or more Affymetrix datasets to the two-channel dataset, then training a classifier using one of the Affymetrix datasets, and testing samples from an independent dataset (again Affymetrix). Thus, not only should a good meta-analysis-based FS method perform well with respect to single dataset FS, but also the method should exhibit minimal performance degradation, if any, when combining heterogeneous data platforms.

3. Results

3.1. Robustness of Rank Average Meta-Analysis. We rate each meta-analysis method by absolute prediction performance (Figure 3). Based on this criterion, we find that rank average meta-analysis, with the highest overall mean rating of 4.56, performs consistently well compared to five other meta-analysis methods including the mDEDS, rank products, Choi, Wang, and naive methods. This analysis answers the question: which meta-analysis-based FS method consistently exhibits the largest prediction performance when combining

Table 2: Differentially expressed genes identified from rank average meta-analysis of multiple microarray datasets.

Breast cancer			Renal cancer			Pancreatic cancer		
Gene symbol	Weighted average rank	Top 20 in # of datasets	Gene symbol	Weighted average rank	Top 20 in # of datasets	Gene symbol	Weighted average rank	Top 20 in # of datasets
ESR1	0.20	6	LOX	13.65	4	S100P	31.42	2
NAT1	33.99	3	COL5A2	16.86	3	LAMC2	51.44	2
DNALI1	48.46	1	ADFP	19.08	4	PHLDA2	201.93	1
SCUBE2	69.27	1	SCNN1A	19.25	2	S100A2	233.07	0
TFF1	76.74	1	LOXL2	21.37	3	MSLN	234.39	1
MYB	82.17	0	ELTD1	27.17	4	WFDC2	236.00	1
CYP2B7P1	86.93	1	PPARGC1A	30.73	1	ITGB6	238.13	0
PDZK1	98.81	0	IFITM1	31.19	2	HK2	239.87	2
PADI2	114.44	0	RALGPS1	37.17	2	R88990*	244.34	0
DNAJC12	123.83	0	VWF	37.85	2	ANO1	252.57	1
TSPAN1	126.87	0	CD70	41.65	0	MXRA5	261.28	0
CDH3	127.46	1	ARHGDIB	42.60	1	PLEK2	264.09	0
XBP1	134.70	0	P4HA1	48.91	2	CDC2	279.79	2
KRT18	136.35	0	BST2	50.56	2	VCAN	285.59	0
EEF1A2	138.25	0	F2R	52.22	1	FERMT1	286.92	1
SLC16A6	140.73	1	SPARC	52.86	1	MCOLN3	309.32	0
ACADSB	142.55	1	LDB2	56.29	2	TNFRSF21	315.68	1
SRD5A1	159.99	1	GJA1	58.54	0	KYNU	324.78	0
CHAD	164.19	0	PLAG1	60.29	1	TACC3	333.27	0
P4HTM	165.08	1	DSG2	68.03	1	TMC5	336.72	0

* Gene symbol not available, using accession number instead.

all available datasets? We assign a rating to each meta-analysis method for every combination of three factors that include (1) clinical application or dataset group, (2) data platform heterogeneity (combining similar or different microarray platforms), and (3) classifiers (logistic regression: LR, diagonal linear discriminant: DLDA, and linear SVM). Ratings for each meta-analysis method are relative to its peers, with higher ratings indicating better prediction performance under the same combination of factors. In Figure 3, bars are proportional to performance ratings. Using pancreatic cancer (PC) as an example, the rank average meta-analysis method has a rating of five (corresponding to a predictive performance AUC of 81.5, See Supplemental Table S1 available online at doi:10.1100/2012/989637) when analyzing homogeneous datasets and when using the logistic regression classifier. This means that its absolute prediction performance is higher than that of four other meta-analysis methods compared under the same conditions (i.e., homogeneous data, logistic regression classifier). The results illustrated in Figure 3 and obtained through a comprehensive analysis of three factors suggest that, relative to its peers, rank average meta-analysis is robust when considering absolute prediction performance.

3.2. Rank Average Identifies Biologically Sensible Genes. For each dataset group, we combine all available microarray datasets and use the rank average meta-analysis method to identify DEGs. Assessing DEG detection performance by examining the genes is difficult unless we know, via validation, whether or not these genes are truly differentially expressed. However, because of the sheer number of genes in high-throughput datasets, the validation process is often time and resource intensive. Despite this, we examine the top ranked genes from each dataset group to verify that the rank average meta-analysis method is identifying genes that are biologically sensible.

Table 2 lists the top 20 genes selected from meta-analysis of each of the three dataset groups: six breast cancer, five renal cancer, and five pancreatic cancer datasets. We optimize the FS method for each individual dataset using three-fold cross validation and the diagonal LDA classifier. The optimal FS method for each dataset differs. We compare ER+ and ER− samples for each breast cancer dataset and find, not surprisingly, that the ESR1 gene (estrogen receptor) is the top ranked gene for all but one dataset. Accordingly, the weighted average rank of ESR1 places it at the top of the combined list. Among the other genes in the list, NAT1 [34], DNALI1, SCUBE2 [35], and TFF1 [36] have been implicated in breast cancer. Although the individual dataset ranks of these genes vary from low to relatively high ranks (e.g., 200 to 300), it is the consistency of selecting these genes from multiple datasets that places them at the top of the combined list. In Table 2, we include the number of individual datasets in which the gene is ranked in the top 20.

Table 3: Properties of six microarray meta-analysis methods.

	Rank average	mDEDS	Rank products	Choi	Wang	Naive (control)
Basic FS methods considered	FC, *T*, SAM, RS, mRMRQ/D	FC, *T*, SAM, RS, mRMRQ/D	FC[1]	*T*[2]	FC[3]	FC, *T*, SAM, RS, mRMRQ/D
Chooses data-specific basic FS method(s)	*Yes*	No	No	No	No	*Yes*
Rank-Based	*Yes*	No	*Yes*	No	No	No

[1] Fold change between all interclass pairs of samples. [2] Most similar to a *t*-statistic, but includes an estimate of interstudy variation. [3] Computes a variance-weighted average of fold change. FC: Fold Change, $T = t$-statistic (t-test), SAM: significance analysis of microarrays, RS: rank sum test, mRMRQ/D: minimum redundancy, maximum relevance with quotient/difference.

We compare the renal cancer clear cell subtype to three other subtypes (i.e., chromophobe, oncocytoma, and papillary) to identify DEGs. The top gene we identify is LOX, which is an oncogene implicated in clear cell renal cancer [37]. The ADFP gene, ranked at #3 in the combined list, is especially interesting because it may be a potential urinary biomarker for detecting renal cancer [38]. ADFP is ranked favorably in all but the Higgins dataset, in which it is ranked at number 75.

The rank average meta-analysis method identifies S100P as the top pancreatic cancer gene, which has been implicated in several studies [39, 40]. The S100P gene has a relatively favorable ranking in the Pilarsky and Pei datasets and moderate to un-favorable rankings in the other datasets, indicating that analysis of individual datasets may not readily identify the gene. Another example, LAMC2, is ranked favorably in the Ishikawa and Pei datasets, but relatively higher in the other datasets. Overall, LAMC2 is ranked second in the combined results and is, according the to literature, a purported pancreatic cancer gene [41]. Weighted average ranks for the pancreatic cancer results increase quickly compared to the breast and renal cancer results, indicating increased heterogeneity among the ranks of the individual datasets. One explanation for this is the slight difference in dataset subtype comparisons. For example, one of the datasets, Ishikawa, extracted RNA samples from pancreatic juice rather than from solid tumors.

The degree of differential expression (and consequently, the rank) of a gene can vary significantly from dataset to dataset. Combining DEG detection results by averaging ranks across datasets reduces variability and improves statistical confidence. Analysis of a single microarray dataset may result in errors during DEG detection—for example, false positives and false negatives (genes that should be differentially expressed, but not favorably ranked). In general, these errors can be reduced by increasing sample size. Combining microarray datasets by averaging ranks effectively increases sample size while enabling robust analysis of heterogeneous data.

4. Discussion

In order to understand the differences in performance among the six meta-analysis-based FS methods, we identify and list the differences and similarities in Table 3. We focus on three properties: (a) basic FS methods forming the basis of meta-analysis, (b) the manner in which these basic FS methods are chosen and applied to individual microarray datasets, and (c) the use of ranks.

Among the five meta-analysis methods (not including the naive control method) rank average and mDEDS are the only methods that consider multiple basic FS methods—for example, fold change, *t*-statistic, SAM, and rank sum—for detecting DEGs (Table 3, row 1). The rank products, Choi and Wang methods use modified forms of basic FS methods. Moreover, rank average is the only method that chooses one basic FS method for each dataset to maximize prediction performance (Table 3, row 2). In contrast, mDEDS uses all of the available basic FS methods for each dataset. Finally, rank average and rank products are the only meta-analysis methods that are rank-based (Table 3, row 3).

Among the basic FS methods, no method can be considered the best because of the data-dependent nature of microarray analysis. Thus, rank average and mDEDS benefit by considering multiple basic FS methods. However, some basic FS methods can produce erroneous results when inappropriately applied (e.g., using a *t*-statistic with gene expression data that is not normally distributed). Rank average meta-analysis further benefits from selecting a single basic FS that optimizes prediction performance. On the other hand, the performance of mDEDS meta-analysis can degrade if it includes a basic FS method that is incompatible with the data. Likewise, the performance of rank products can degrade when the fold change FS method is not appropriate for the data. The Choi and Wang methods may also suffer from this problem. However, they seem to perform fairly well when applied to the datasets in this study (see Figure 3). Finally, rank-based meta-analysis methods that consider multiple basic FS methods allow a fair comparison among the basic FS methods. In light of these results, for microarray meta-analysis, we recommend (1) to use rank-based methods, (2) to consider a wide variety of basic FS methods, and (3) to optimize the FS method for each individual dataset based on application-specific criteria (e.g., prediction performance for diagnostic applications).

Despite the benefits summarized in Table 3, rank average meta-analysis and the evaluation criteria presented in this study are not without limitations. The limitations of this study include (1) the scope of data and classifiers considered, (2) the criterion for measuring performance of a meta-analysis method, and (3) normalization and pre-processing of gene expression data. First, the results of this study may be dataset-specific. Although we have strived to provide a wide range of scenarios to allow adequate assessment of

these meta-analysis methods, results may differ when applied to other dataset groups. Second, we use prediction AUC as the performance criterion. However, microarray-based clinical prediction is only one possible application. Other applications may need to identify genes based on biological relevance [15]. It is unclear which meta-analysis methods would perform well in such applications. The rank average meta-analysis method benefits from choosing a basic FS method for each individual dataset that optimizes (via cross-validation) prediction performance. Thus, there is a potential bias in the performance of rank average meta-analysis. On the other hand, the ability to choose basic FS methods that perform well for a particular application, such as prediction, could be considered a favorable property of rank average meta-analysis. Finally, it is possible that normalization of gene expression datasets (e.g., quantile normalization) can improve the performance of meta-analysis by reducing batch effect. Specifically, removal of batch effects (1) can improve prediction performance when training and testing are applied to independent, heterogeneous datasets and (2) can improve the performance of simple meta-analysis methods that aggregate samples from multiple heterogeneous datasets. However, we do not consider any batch-effect normalization procedures in this study.

5. Conclusions

In order to address the sample-size problem in gene expression analysis as well as the need for accurate solutions for clinical prediction problems, we proposed the rank average meta-analysis-based FS method. Rank average meta-analysis identifies differentially expressed genes from multiple microarray datasets. We used a comprehensive study of multiple factors and found that rank average performs consistently well compared to five other meta-analysis methods in terms of prediction performance. This comprehensive study enabled us to measure the robustness of rank average to three factors that are often encountered in clinical prediction applications. These factors include clinical application (e.g., breast, renal, and pancreatic cancer), microarray data platform heterogeneity, and classifier model (logistic regression, diagonal LDA, and SVM). Rank average meta-analysis, performs well because it selects dataset-specific basic FS methods and then averages the ranks across all individual datasets to produce a final robust gene ranking. In comparison to five other meta-analysis methods the rank average method is not always the best method for some factor combinations. However, it is consistently among the best performing in terms of its ability to identify predictive genes. Although we presented results from analysis of microarray gene expression data, the proposed methods may be generalized for other bioinformatics problems that require feature selection.

Acknowledgments

This work was supported in part by Grants from National Institutes of Health (Bioengineering Research Partnership R01CA108468, Center for Cancer Nanotechnology Excellence U54CA119338); Georgia Cancer Coalition (Distinguished Cancer Scholar Award to M. D. Wang); Hewlett Packard; and Microsoft Research. The funding sources listed here have supported this multiyear investigation of microarray meta-analysis for clinical prediction, including covering the stipends and salaries of multiple coauthors, computing hardware and software licenses, travel expenses to technical meetings to present this work, and publication expenses.

References

[1] K. M. Lin, J. Kang, H. Shin, and J. Lee, "A cube framework for incorporating inter-gene information into biological data mining," *International Journal of Data Mining and Bioinformatics*, vol. 3, no. 1, pp. 3–22, 2009.

[2] J. H. Phan, R. A. Moffitt, T. H. Stokes et al., "Convergence of biomarkers, bioinformatics and nanotechnology for individualized cancer treatment," *Trends in Biotechnology*, vol. 27, no. 6, pp. 350–358, 2009.

[3] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, "Microarray data analysis: from disarray to consolidation and consensus," *Nature Reviews Genetics*, vol. 7, no. 1, pp. 55–65, 2006.

[4] L. Shi, W. Tong, H. Fang et al., "Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential," *BMC Bioinformatics*, vol. 6, no. 2, article S12, 2005.

[5] B. M. Bolstad, R. A. Irizarry, M. Åstrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias," *Bioinformatics*, vol. 19, no. 2, pp. 185–193, 2003.

[6] P. Stafford and M. Brun, "Three methods for optimization of cross-laboratory and cross-platform microarray expression data," *Nucleic Acids Research*, vol. 35, no. 10, article e72, 2007.

[7] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical Bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.

[8] L. Lusa, L. M. McShane, J. F. Reid et al., "Challenges in projecting clustering results across gene expression-profiling datasets," *Journal of the National Cancer Institute*, vol. 99, no. 22, pp. 1715–1723, 2007.

[9] T. Park, S. G. Yi, Y. K. Shin, and S. Y. Lee, "Combining multiple microarrays in the presence of controlling variables," *Bioinformatics*, vol. 22, no. 14, pp. 1682–1689, 2006.

[10] J. K. Choi, U. Yu, S. Kim, and O. J. Yoo, "Combining multiple microarray studies and modeling interstudy variation," *Bioinformatics*, vol. 19, no. 1, pp. i84–i90, 2003.

[11] J. Wang, K. R. Coombes, W. E. Highsmith, M. J. Keating, and L. V. Abruzzo, "Differences in gene expression between B-cell chronic lymphocytic leukemia and normal B cells: a meta-analysis of three microarray studies," *Bioinformatics*, vol. 20, no. 17, pp. 3166–3178, 2004.

[12] S. Yoon, Y. Yang, J. Choi, and J. Seong, "Large scale data mining approach for gene-specific standardization of microarray gene expression data," *Bioinformatics*, vol. 22, no. 23, pp. 2898–2904, 2006.

[13] R. Breitling and P. Herzyk, "Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 5, pp. 1171–1189, 2005.

[14] A. Campain and Y. H. Yang, "Comparison study of microarray meta-analysis methods," *BMC Bioinformatics*, vol. 11, article 408, 2010.

[15] J. H. Phan, Q. Yin-Goen, A. N. Young, and M. D. Wang, "Improving the efficiency of biomarker identification using biological knowledge," *Pacific Symposium on Biocomputing*, pp. 427–438, 2009.

[16] A. N. Schuetz, Q. Yin-Goen, M. B. Amin et al., "Molecular classification of renal tumors by gene expession profiling," *Journal of Molecular Diagnostics*, vol. 7, no. 2, pp. 206–218, 2005.

[17] J. Jones, H. Otu, D. Spentzos et al., "Gene signatures of progression and metastasis in renal cell cancer," *Clinical Cancer Research*, vol. 11, no. 16, pp. 5730–5739, 2005.

[18] E. J. Kort, L. Farber, M. Tretiakova et al., "The E2F3-oncomir-1 axis is activated in Wilms' tumor," *Cancer Research*, vol. 68, no. 11, pp. 4034–4038, 2008.

[19] M. V. Yusenko, R. P. Kuiper, T. Boethe et al., "High-resolution DNA copy number and gene expression analyses distinguish chromophobe renal cell carcinomas and renal oncocytomas," *BMC Cancer*, vol. 9, article 152, 2009.

[20] M. V. Yusenko, D. Zubakov, and G. Kovacs, "Gene expression profiling of chromophobe renal cell carcinomas and renal oncocytomas by Affymetrix GeneChip using pooled and individual tumours," *International Journal of Biological Sciences*, vol. 5, no. 6, pp. 517–527, 2009.

[21] J. P. T. Higgins, R. Shinghal, H. Gill et al., "Gene expression patterns in renal cell carcinoma assessed by complementary DNA microarray," *American Journal of Pathology*, vol. 162, no. 3, pp. 925–932, 2003.

[22] L. Shi, G. Campbell, W. D. Jones et al., "The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models," *Nature Biotechnology*, vol. 28, no. 8, pp. 827–838, 2010.

[23] L. D. Miller, J. Smeds, J. George et al., "An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 38, pp. 13550–13555, 2005.

[24] C. Sotiriou, P. Wirapati, S. Loi et al., "Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis," *Journal of the National Cancer Institute*, vol. 98, no. 4, pp. 262–272, 2006.

[25] A. J. Minn, G. P. Gupta, P. M. Siegel et al., "Genes that mediate breast cancer metastasis to lung," *Nature*, vol. 436, no. 7050, pp. 518–524, 2005.

[26] M. J. van de Vijver, Y. D. He, L. J. van 't Veer et al., "A gene-expression signature as a predictor of survival in breast cancer," *The New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.

[27] L. Badea, V. Herlea, S. O. Dima, T. Dumitrascu, and I. Popescu, "Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia," *Hepato-Gastroenterology*, vol. 55, no. 88, pp. 2016–2027, 2008.

[28] H. Pei, L. Li, B. L. Fridley et al., "FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt," *Cancer Cell*, vol. 16, no. 3, pp. 259–266, 2009.

[29] M. Ishikawa, K. Yoshida, Y. Yamashita et al., "Experimental trial for diagnosis of pancreatic ductal carcinoma based on gene expression profiles of pancreatic ductal cells," *Cancer Science*, vol. 96, no. 7, pp. 387–393, 2005.

[30] C. Pilarsky, O. Ammerpohl, B. Sipos et al., "Activation of Wnt signalling in stroma from pancreatic cancer identified by gene expression profiling," *Journal of Cellular and Molecular Medicine*, vol. 12, no. 6B, pp. 2823–2835, 2008.

[31] C. A. Iacobuzio-Donahue, A. Maitra, M. Olsen et al., "Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays," *American Journal of Pathology*, vol. 162, no. 4, pp. 1151–1162, 2003.

[32] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 9, pp. 5116–5121, 2001.

[33] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[34] I. Bièche, I. Girault, E. Urbain, S. Tozlu, and R. Lidereau, "Relationship between intratumoral expression of genes coding for xenobiotic-metabolizing enzymes and benefit from adjuvant tamoxifen in estrogen receptor alpha-positive postmenopausal breast carcinoma," *Breast Cancer Research*, vol. 6, no. 3, pp. R252–R263, 2004.

[35] T. Z. Parris, A. Danielsson, S. Nemes et al., "Clinical implications of gene dosage and gene expression patterns in diploid breast carcinoma," *Clinical Cancer Research*, vol. 16, no. 15, pp. 3860–3874, 2010.

[36] S. J. Prest, F. E. May, and B. R. Westley, "The estrogen-regulated protein, TFF1, stimulates migration of human breast cancer cells," *The FASEB Journal*, vol. 16, no. 6, pp. 592–594, 2002.

[37] H. Liu, A. R. Brannon, A. R. Reddy et al., "Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma," *BMC Systems Biology*, vol. 4, article 51, 2010.

[38] J. J. Morrissey, A. N. London, J. Luo, and E. D. Kharasch, "Urinary biomarkers for the early diagnosis of kidney cancer," *Mayo Clinic Proceedings*, vol. 85, no. 5, pp. 413–421, 2010.

[39] T. Arumugam, D. M. Simeone, K. Van Golen, and C. D. Logsdon, "S100P promotes pancreatic cancer growth, survival, and invasion," *Clinical Cancer Research*, vol. 11, no. 15, pp. 5356–5364, 2005.

[40] H. J. Whiteman, M. E. Weeks, S. E. Dowen et al., "The role of S100P in the invasion of pancreatic cancer cells is mediated through cytoskeletal changes and regulation of cathepsin D," *Cancer Research*, vol. 67, no. 18, pp. 8633–8642, 2007.

[41] M. Katayama, A. Funakoshi, T. Sumii, N. Sanzen, and K. Sekiguchi, "Laminin $\gamma 2$-chain fragment circulating level increases in patients with metastatic pancreatic ductal cell adenocarcinomas," *Cancer Letters*, vol. 225, no. 1, pp. 167–176, 2005.

The Scientific World Journal
Volume 2012, Article ID 380495, 11 pages
doi:10.1100/2012/380495

The ScientificWorldJOURNAL

Research Article

Gene Expression Profiles for Predicting Metastasis in Breast Cancer: A Cross-Study Comparison of Classification Methods

Mark Burton,[1,2] Mads Thomassen,[1,2] Qihua Tan,[1,2,3] and Torben A. Kruse[1,2]

[1] *Research Unit of Human Genetics, Institute of Clinical Research, University of Southern Denmark, Sdr. Boulevard 29, 5000 Odense C, Denmark*
[2] *Department of Clinical Genetics, Odense University Hospital, Sdr. Boulevard 29, 5000 Odense C, Denmark*
[3] *Institute of Public Health, University of Southern Denmark, J. B. Winsløws Vej 9B, 5000 Odense C, Denmark*

Correspondence should be addressed to Mark Burton, mark.burton@ouh.regionsyddanmark.dk

Received 25 August 2012; Accepted 2 October 2012

Academic Editors: M. A. Kon and K. Najarian

Machine learning has increasingly been used with microarray gene expression data and for the development of classifiers using a variety of methods. However, method comparisons in cross-study datasets are very scarce. This study compares the performance of seven classification methods and the effect of voting for predicting metastasis outcome in breast cancer patients, in three situations: within the same dataset or across datasets on similar or dissimilar microarray platforms. Combining classification results from seven classifiers into one voting decision performed significantly better during internal validation as well as external validation in similar microarray platforms than the underlying classification methods. When validating between different microarray platforms, random forest, another voting-based method, proved to be the best performing method. We conclude that voting based classifiers provided an advantage with respect to classifying metastasis outcome in breast cancer patients.

1. Introduction

The analysis of high-dimensional gene expression datasets has posed new computational challenges. These datasets have, for example, in breast cancer research, been applied to develop classifiers predicting metastasis outcome, disease recurrence, or breast cancer survival. Some of the classification methods most frequently applied to microarray data are logistic regression [1, 2], support vector machines (SVM) [3–12], neural networks (NNET) [1, 13], random forest (RF) [1, 12], and classifiers based on voting [1]. However, few studies have systematically compared the predictive performance of such methods using microarray gene expression datasets on breast cancer. In their studies, method comparisons have been done within the same datasets by, for example, 10-fold cross-validation, leave-one-out cross-validation, or hold-out procedures [14–18], addressing prediction of relapse within a 5-year period [14, 16, 19], or molecular subtype classification [15]. Furthermore, even fewer studies have compared cross-study validation between classification methods within the field of breast cancer research. Two studies addressed ER-positivity and molecular subtype classification [20, 21], while another tested prediction of relapse within a 5-year period in a small group of 19 independent patients [22].

This study compares the performance of seven classification methods belonging to four different categories for predicting metastatic outcome in lymph negative breast cancer patients, which have not been treated with adjuvant systemic therapy. The classification methods used included an ensemble decision tree model (random forest), regression (logistic regression), four support vector machines and a neural network. To address various degrees of variation for such tasks, the comparisons were done either within the same dataset (internal) or between different datasets (external). Within the same dataset model building and classification were performed using 10-fold cross-validation. Across datasets the comparisons were done in two ways. The first is in which the validations are conducted between studies using the same microarray platform (classifiers developed from an Affymetrix dataset and validated on an independent Affymetrix dataset), while the second encompasses validations across studies with different platforms (classifiers developed from an Agilent dataset and validated on an independent Affymetrix dataset). Furthermore, we

Table 1: Overview of datasets used.

Dataset	Chip	Probes (K)	Patients	Outcome	Treatment	Define genes	Internal CV	External validation train	External validation test	Reference
Amsterdam	Agilent/Rosetta	25	295, N^+, N^-	DM	None, et, ct	√				[31]
Amsterdam (AM) (subset of the above)	Agilent/Rosetta	25	151, N^-	DM	None	√	√	√		[31]
Rotterdam (RO)	Affymetrix HG-133A	22	286, N^-	DM	None	√	√	√		[28]
HUMAC	Spotted oligonucleotides	29	60, N^-	ME	None	√				[4]
Huang	Affymetrix 95av2	12	52, N^+	RE	ct	√				[27]
Sotiriou 2003	Spotted cDNA	7.6	99, N^+/N^-	RE	et, ct	√				[24]
Sotiriou 2006	Affymetrix HG-133A	22	179, N^+/N^-	DM	et	√				[25]
Uppsala	Affymetrix HG-133A+B	44	236, N^+/N^-	DF	None, ct, et	√				[6]
Stockholm	Affymetrix HG-133A+B	44	159, N^+/N^-	RE	None, ct, et	√				[23]
TRANSBIG (TR)	Affymetrix HG-133A	22	147, N^-	DM	None				√	[32]
Mainz (MA)	Affymetrix HG-133A	22	200, N^-	DM	None				√	[33]

The columns show the following: "dataset": the individual names for the eight datasets; "chip": microarray chip used; "probes": number of probes on the chip measured in thousands (K = 1000); patients': number of patients in the study and their nodal status (N^+ and N^- indicates number of node-positive and -negative patients; "outcome" covers the clinical outcome being DM: distant metastasis, ME: metastasis, RE: relapse, and DF: death from breast cancer; "treatment" shows patient treatments abbreviated by et: endocrine therapy, ct: chemo therapy, and none: no adjuvant therapy.

examined the effect of combining the classification results on each sample by the seven methods into one final classification determined by majority voting, and performances compared by internal and external validation as well.

2. Materials and Methods

2.1. Datasets Used in This Study. The following eight datasets were used for either defining the gene features and or training purposes in the further study: samples from the studies [23–27] and samples from the Gene Expression Omnibus-(GEO-) series GSE2034 [28], GSE4796 [4], and GSE3494 [6] (Table 1). A subset of 151 node-negative samples from the dataset by van de Vijver (AM) and the entire GSE2034 dataset [28] (abbreviated RO) were used for classifier development in the further study (Table 1). The following datasets were used as independent testing sets: the node-negative samples from GSE7390 [29] (abbreviated TR) and the GSE11121 dataset [30] (abbreviated MA) (Table 1).

2.2. Dataset Processing. The eight datasets above were downloaded and directly used for identification of rank-significant genes. Following this identification, the four datasets: AM, RO, TR, and MA were all standardized to have mean zero and standard deviation one. Calculations and classification were all conducted using the R free package. For random forest, logistic regression, support vector machines, and neural network we used the *randomForest*, *glm*, *e1071,* and *nnet* packages, respectively.

2.3. Identification of Cross-Study Rank-Significant Features. To determine which genes should be used to build gene expression classifiers, we used the eight publicly available datasets mentioned above, which were used in our two previous studies [34, 35]. This was done by applying the microarray meta-analysis described in [34], upon the individual gene expression values of each individual probe/gene in the eight datasets. This method ranks each individual gene in each dataset according to its signal-to-noise ratio, calculates the

gene's mean rank across datasets, and determines if this mean rank is significantly high or low, according to a significance cutoff at FDR ≤ 0.05.

2.4. Classifier Building. The features within each training dataset were ranked according to their random forest variable importance measure. For each feature, this value reports the standardized drop in prediction accuracy when the classlabels are permuted [36]. For each feature, this rank was used for model building by subsequently adding one feature at a time in a "top-down" manner. To avoid creating bias, during gene selection and training of the final classifier, and on classification performance, we used ten-times repeated 10-fold cross-validation accuracies as a performance measure, as this metric has previously been shown to give an excellent bias-variance balance [37]. In this study, the models were developed to achieve the best mean sensitivity and specificity thus forcing the overall accuracy to give a balanced sensitivity and specificity. Seven different classification methods were used for model building which included: random forest (RF) [36], logistic regression, SVM with a radial- (R-SVM), a linear (L-SVM), polynomial (P-SVM), or a sigmoid-based kernel (S-SVM) [38], and a neural network with a single hidden layer (NNET). The voting approach is described in detail below. As all classification methods have hyperparameters, we optimized these parameters during model building using a grid-like search of parameter combinations. In random forest, we optimized the number of trees in the forest ($n$tree) from settings of 2000, 3000, 4000, and 5000 trees, and the number of subselected predictors for node splitting (mtry) with settings of: $1, 0.5 \cdot \sqrt{(\text{number of features})}$, $1 \cdot \sqrt{(\text{number of features})}$, $2 \cdot \sqrt{(\text{number of features})}$, and total number of features. In all support vector machines, the slack variable penalizing cost parameter (C) was optimized using settings of 0.01, 0.1, 1, and 10, and the γ-parameter, controlling the spreading of samples in feature space, with the settings of 0.001, 0.01, 0.1, and 1, and for P-SVM also the polynomial degree using degrees of 2, 3, and 4.

2.5. Voting. The voting procedure can be regarded as a metamodel, where a sample is first fed to be predicted by each of the respective classification methods. These predictions are next fed to the final metamodel, combining each of these predictions into a final classification determined by majority voting.

The voting procedure, at the level of internal dataset prediction, consisted of two steps. In the first step, each sample is classified ten times during 10×10 CV, meaning that each sample is given 10 votes for classification within each classification method. To prevent ties, the nine first votes were used for class decision. In the second step the final votes from each of the seven classification methods are combined into one vote, thus creating the cross-classification voting result.

During external validation, every sample is classified once by each of the seven classification methods. The voting classification for each sample is determined by the winning class assigned by the seven classification methods (voters).

2.6. Classification Performance Assessment. We compared the bAcc, defined as the mean of sensitivity and specificity, of classifiers at two levels either internal or external. Internal performance was determined by the 10-times repeated 10-fold cross-validation classification accuracies. External performances obtained through transferring the trained classifier from the training sets to classify each of the independent samples are reported. In external validation, two different situations were examined: (1) between similar (RO on TR or MA) and (2) different microarray platforms (AM on TR or MA), covering Affymetrix-based classifiers validated on an Affymetrix dataset and Agilent-based classifiers validated on another Affymetrix dataset, respectively.

2.7. Endpoint/Outcome Definition. The outcome is defined as metastasis after time of diagnosis. As this study addresses outcome classification, we did not consider the time-to-event component or censoring, due to the fact that survival analysis sometimes can be misleading when considering classification, and because transformation of time-to-event into a binary outcome can blur prediction of the classes [39].

2.8. Comparison of External Validation Performance. There is to our knowledge no standard statistics for comparing classifier performance on unbalanced datasets using the balanced accuracy as a performance measure. Therefore, in order to test the significance of the performance difference between the classification methods (defined as a significant difference between correct predictions using method A versus using method B), we used a repeated downsampled binomial test approach consisting of five steps. (1) The classifiers classification results upon the entire test data were initially converted into a balanced test result by downsampling. Downsampling obtains a class-balanced dataset from an imbalanced dataset by removing a subset of randomly selected samples from the majority class, where the number of samples removed equals the difference in sample size between the major and the minor class. In this study the majority class is the nonmetastasis class; (2) the number of samples correctly classified by one classification method but incorrectly by the other classification method and vice versa is counted; (3) the significance of the difference in these counts is determined using a binomial χ^2-test; (4) the P-value of this test is stored. The steps 2 to 4 are repeated 1000 times; (5) from the 1000 tests, the median P value is reported as the statistical significance impact between the two compared methods.

3. Results

3.1. Features and Classifiers/Models. In this study the classifiers were developed to predict metastasis outcome using full follow-up time. To make the classifiers globally applicable and robust, we identified genes being significantly associated with outcomes across eight different studies using different microarray platforms and originating from different populations. These eight datasets are referred to as the "feature definers" (FD) (Table 1). In the further analysis, two of the

FD datasets were used as training sets. The first, Rotterdam (RO), is an Affymetrix-based dataset containing 286 samples, and the second, Amsterdam (AM), is a node-negative subset of 151 samples from the entire FD-Amsterdam dataset (Table 1). Two independent datasets, not used for feature selection or classifier development, were used as test sets. These comprise the TRANSBIG (TR) and Mainz (MA) datasets, which are based on the Affymetrix platform and consist of 147 and 200 samples, respectively.

As a preliminary feature selection step, we identified genes being significantly associated with outcomes across eight different studies using a rank-based method (as described in Section 2). This method led to identification of 519 rank-significant genes. By matching the 519 rank-significant genes and those present in AM, RO, TR, and MA, these genes were reduced to 283 (Figure 1) and were thus used for classifier building. The list of 283 genes is shown in Supplementary Table 1 in Supplementary Material available online at doi:10.1100/2012/380495.

In order to build the models, the 283 features within each training dataset were ranked according to their random forest variable importance measure (Figure 2). For a given feature, this measure reports the standardized drop in prediction accuracy when the class labels are permuted [36]. This rank was then used for model building by subsequently adding one feature at a time in a "top-down" forward manner (Figure 2).

3.2. Comparison of Classification Methods: Internal Validation Performance. To reduce variability, and complexity and to keep validation parameters as constant as possible, the performance of the classifiers was tested within the same dataset by a ten-times repeated 10-fold cross-validation (Figure 2). This validation scheme partitions the training data into 10 nearly equal-sized folds. Subsequently, 10 iterations of training and validation are performed. During each of these iterations a different fold of the training data is left out for validation and the remaining are used for learning. The mean accuracy of all 10-folds validated is thus the 10-fold cross-validated (10×10 CV) accuracy of the model. By repeating this process 10 times a more robust and unbiased estimation of the performance is obtained. In our study, the balanced accuracy (bAcc), defined as the mean of sensitivity and specificity, was used as a performance measure. It should be noted that the individual classification performances are artificially elevated due to information leakages, caused by AM and RO being used for primary feature selection, and that the entire AM and RO datasets are used for importance ranking prior to cross-validation. However, the differences between the individual classification method performances are assumed, unaffected by these leakages.

The performances within the AM and RO by each classification method were combined, and the mean performance calculated. The classifiers based on NNET had the best performance achieving a mean 10×10 CV bAcc of 78%, followed by S-SVM, L-SVM, R-SVM, P-SVM, RF, and LR achieving mean 10×10 CV bAcc of 74.1%, 72.1%, 71.6%, 70.9%, 69.4%, and 68.0%, respectively (Figure 3 and Supplementary Table 2). The significance of these differences was tested using the down-sampling statistical test described in Section 2, showing that NNET significantly outperformed RF ($P = 0.011$), LR ($P = 1.2e^{-5}$), L-SVM ($P = 0.027$), and P-SVM ($P = 6.3e^{-4}$). NNET only borderline significantly outperformed R-SVM ($P = 0.07$) and S-SVM ($P = 0.10$). Furthermore, S-SVM also performed significantly better than RF ($P = 0.049$), LR ($P = 9.0e^{-5}$), and P-SVM ($P = 0.049$), and R-SVM outperformed RF ($P = 0.018$). No significant performance difference was found when comparing the other classification methods.

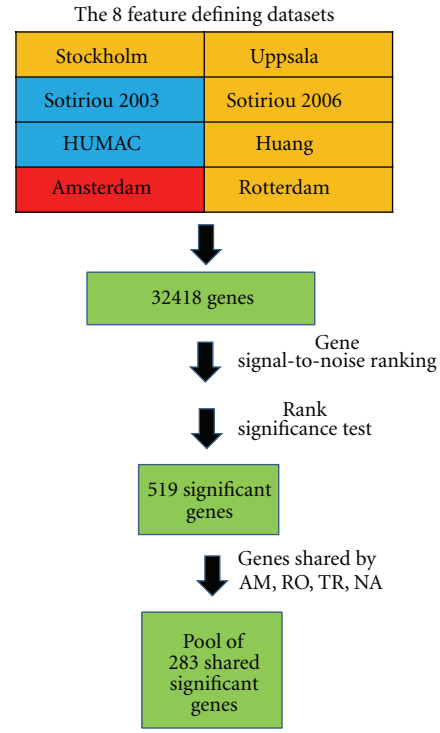


Figure 1: Feature selection. Eight breast cancer gene expression datasets (feature defining datasets), covering 32418 genes, were used to define a list of rank significant genes. Datasets using the Affymetrix platform, spotted oligonucleotides, and the Agilent platform are colored orange, blue, and red, respectively. These genes were first ranked within each of the eight datasets according to their signal-to-noise ratio, and their across dataset mean rank calculated. This mean rank was significance tested as described in Section 2, resulting in a list of 519 rank significant genes. These 519 genes were reduced to a pool of 283 genes shared by the two training sets (AM and RO) and the testing sets (TR and MA), used in the further study.

We next combined the cross-validated results by the seven methods into a voting procedure. This led to a mean 10×10 CV bAcc of 86.9%, which significantly outperformed all the seven underlying classification methods: RF ($P = 1.1e^{-19}$), LR ($P = 1.7e^{-18}$), R-SVM ($P = 4.2e^{-12}$), L-SVM ($P = 1.4e^{-14}$), P-SVM ($P = 8.9e^{-16}$), S-SVM ($P = 5.8e^{-11}$), and NNET ($P = 2.6e^{-4}$) (Figure 3 and Supplementary Table 2).

3.3. External Validation Performance between Similar Microarray Platforms. The performance of the classifiers was validated in independent datasets based on the same

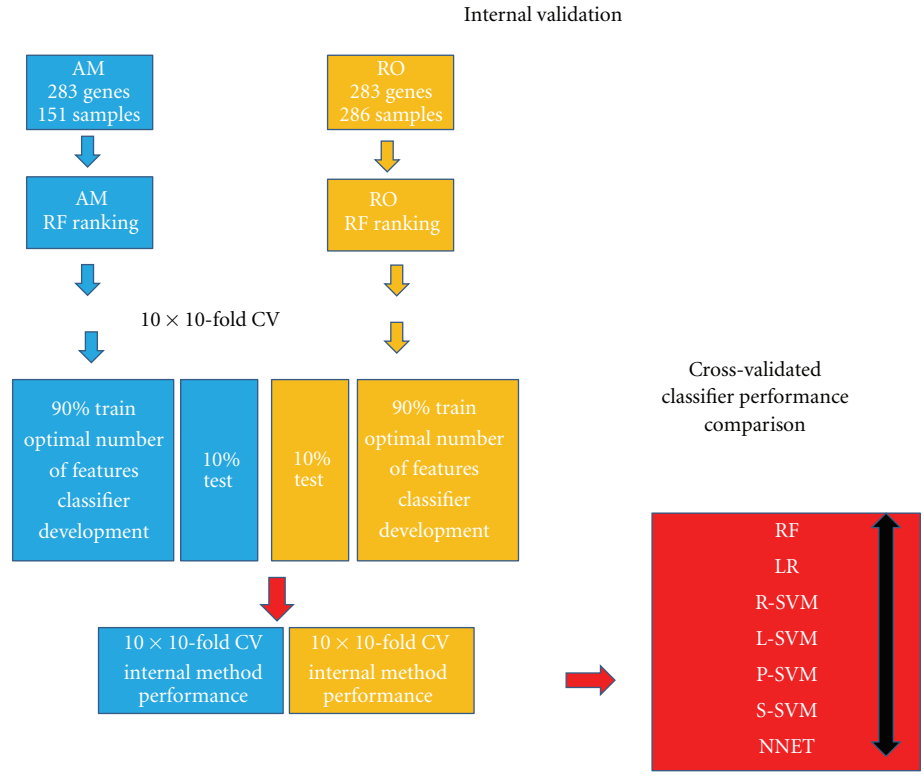


Figure 2: Internal validation procedure. The two datasets, AM (blue) and RO (orange) composed of the 283 rank-significant genes and 151 or 286 samples, respectively, were used for internal performance evaluation. These datasets were first individually used to rank each feature by their random forest variable importance value (RF ranking). These ranks were separately used for selecting the optimal number of features by adding one feature using the same classification method, using a 10-times repeated 10-fold cross-validation procedure. The AM and RO 10-times cross-validation results using the same classification method were combined, and the mean classification performance of each method was compared.

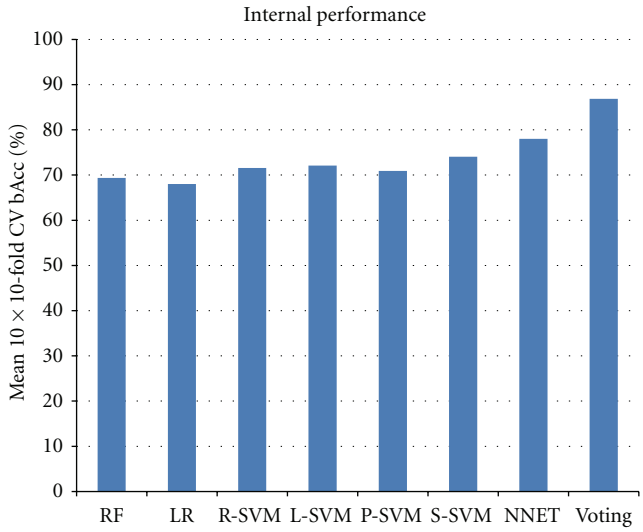


Figure 3: Internal validation performance. Shown in blue histograms are the mean 10-times repeated 10-fold cross-validation balanced accuracy performance (bAcc) within the two training datasets: AM and RO. Methods used are random forest (RF), logistic regression (LR), support vector machines with a radial (R-SVM), linear (L-SVM), polynomial (P-SVM), sigmoid kernel (S-SVM), a neural network with a single hidden layer (NNET), or cross-method voting (Voting).

Table 2: Number of features in the models.

Dataset	AM	RO
Method	(n = 151)	(n = 286)
RF	21	21
LR	5	11
R-SVM	20	25
L-SVM	8	11
P-SVM	4	7
S-SVM	17	35
NNET	21	16

AM and RO are the Amsterdam and Rotterdam training sets and n shows the number of samples in the respective datasets. Methods used are as follows: RF: random forest, LR: logistic regression, R-, L-, P-, and S-SVM: support vector machine with a radial basis function, linear, polynomial, or sigmoid kernel, and NNET: neural network with a single hidden layer.

microarray platform (Affymetrix), which covers the validation of RO-based classifiers on the TR and MA test data (Figure 4), which contained between 7 to 35 features (Table 2). In this setting, the entire classifiers developed in the training set, using the features and rules associated with the classifiers, were used to classify the independent samples in the entire test sets, and the performance is defined as the mean test accuracy in TR and MA.

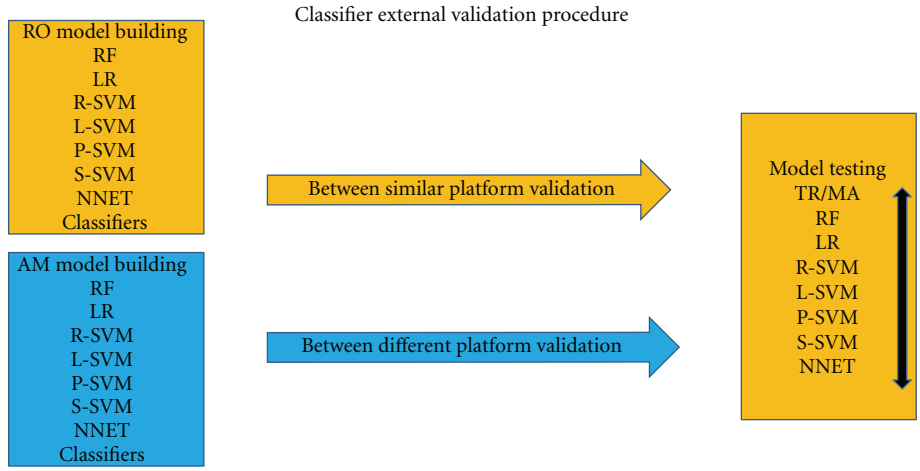


Figure 4: The procedure for external validation of classifiers. External classifier validation. Two datasets were used for training (AM and RO), and two others for testing (TR and MA). Datasets based on the Affymetrix and Agilent platforms are shown in orange and blue, respectively. RO and AM classifiers were used for evaluating external validation of classifiers developed from datasets using similar platform and using different microarray platforms, respectively. The models built in RO were tested in TR and MA and their mean performance calculated. This was done for all classification methods and compared. The same was done for testing AM models.

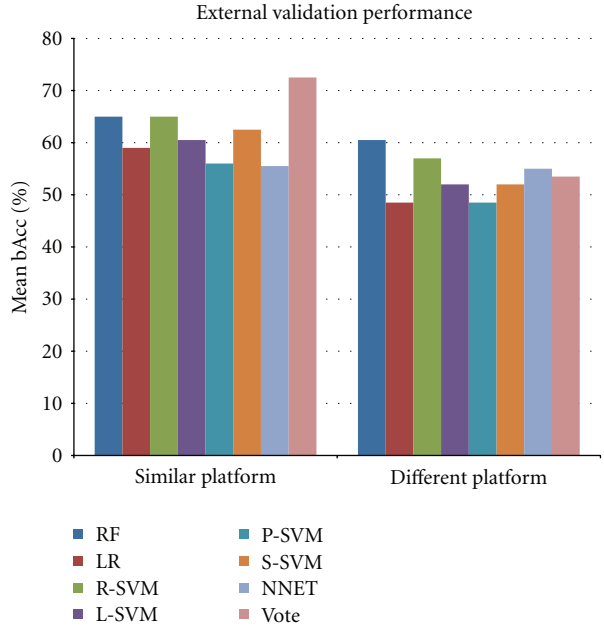


Figure 5: External validation performance. External classifier performance. The histogram shows the mean testing performance when classifiers are validated in test sets based on similar platforms or different platforms as from which they were developed. Each bar represents the mean balanced accuracy by random forest (RF), logistic regression (LR), support vector machines with a radial (R-SVM), linear (L-SVM), polynomial (P-SVM), sigmoid kernel (S-SVM), a neural network with a single hidden layer (NNET), or cross-method voting (VOTE), respectively.

RF and R-SVM had the best external classification performance achieving a mean bAcc of 65%, while NNET had the poorest performance (55.5% mean bAcc) (Figure 5). RF performed significantly better than LR ($P = 0.0085$), L-SVM ($P = 0.026$), P-SVM ($P = 0.0057$), and NNET ($P = 0.0012$), and R-SVM also performed significantly better than LR ($P = 0.0064$), L-SVM ($P = 0.025$), P-SVM ($P = 0.013$) and NNET ($P = 0.0025$) (Figure 5).

The voting procedure increased the performance to a 72.5% mean bAcc and significantly outperformed the seven underlying methods: RF ($P = 6.1e^{-5}$), LR ($P = 1.5e^{-8}$), R-SVM ($P = 0.00026$), L-SVM ($P = 6.0e^{-8}$), P-SVM

($P = 2.1e^{-7}$), S-SVM ($P = 1.9e^{-6}$), and NNET ($P = 4.7e^{-7}$) (Figure 5). Detailed overview of the individual validation results is shown in Supplementary Table 3.

3.4. External Validation Performance between Different Microarray Platforms. The performance of the classifiers was finally validated in independent datasets (Affymetrix) based on a different microarray platform from the one used by training data (Agilent) and covers the validation of AM-based classifiers on the TR and MA test data. These classifiers contained 4 to 21 features (Table 2). As in the case of the between-similar-platform validation, the entire classifier developed in the training set, using the features and rules associated with the classifiers, was used to classify the independent samples in the entire test sets (Figure 4), and the performance is defined as the mean test accuracy in TR and MA.

Comparison of classifiers developed on an Agilent dataset and validated on an Affymetrix dataset revealed that the mean classification performances based on RF had the best performance amongst the seven methods, achieving a mean bAcc of 60.5%, while the poorest performances were achieved by the LR- and P-SVM classifiers, which obtained only 48.5% bAcc (Figure 5). RF performed significantly better than the other six methods: LR ($P = 3e^{-4}$), R-SVM ($P = 0.038$), L-SVM ($P = 0.0012$), P-SVM ($P = 9.8e^{-5}$), S-SVM ($P = 0.0015$), and NNET ($P = 0.032$) (Figure 5). In contrast to the between-similar-platforms validations, the voting procedure only obtained a mean bAcc performance of 53.5%, which was a borderline significantly inferior to RF ($P = 0.059$) (Figure 5). Detailed description of the individual between different platforms validation results is shown in Supplementary Table 4.

4. Conclusion and Discussion

This study compared seven classification methods and a voting procedure ability to predict metastasis outcome in lymph node-negative breast cancer patients. The results showed that during internal assessment and external validation—methods based on voting had the best performance.

Our study first compared the internal performance within a single dataset and showed that NNET had the best performance followed by the support vector machines, while RF and LR had the worst performances. This implies that at least for prediction of metastasis outcome within the same dataset—NNET and support vector machines displays superiority. This finding agrees well with other studies using cross- or hold-out procedures for performance comparisons. For example, one study comparing the performance of eight different classification methods showed that NNET and SVMs in general perform better than the other six methods for predicting outcome in eight different cancer microarray datasets [15]. Several studies confirm our finding of RF inferiority when using cross-validation [17, 18, 40]. Interestingly, a study conducting algorithm comparison on microarray gene expression based drug signatures showed that NNET and R-SVM had the best performance when tested in the most heterogeneous datasets [41]. As the datasets used in our study are likely to be very heterogeneous, due to the nature and etiology of breast cancer, the superior performances of NNET and support vector machines could reflect the ability of these particular methods to distinguish outcome in such complex datasets.

Combining the classification results by each method into classification based on voting significantly increased the internal performance. The finding of voting superiority in the internal validations suggests that voting would be valuable when applied to datasets having a combination of limited technical variation (due to using same protocols and platforms) and biological heterogeneity. Although the patients in our study are limited to being node negative, they may still be very heterogeneous due to the existence of various breast cancer molecular subgroups and the disease etiology. Voting may therefore reduce the variation associated with this biological heterogeneity. This is in line with the above-mentioned study, showing that some classification methods are more suitable for prediction tasks in complex datasets [41].

Our finding of voting superiority agrees with four other studies: one using multiple different feature extraction methods in combination with SVM for gene microarray classification showed that using a voting-based method across all the examined combinations achieved a better 10-fold cross-validated classification performance compared to any single combination [33]; a second study showed that an SVM-based ensemble outperformed single SVM for microarray data classification [42]; a third study comparing the performance of principal component discriminant classifiers either with or without voting using cross-validation applied on a simulated dataset a leukemia microarray gene expression dataset, a Gaucher serum proteomics dataset and a grape extract metabolomics dataset, also showed that voting had a better performance than the nonvoting method [43]; a fourth study comparing the performance of single models to combined models in thirteen diverse microarray datasets, which included predictions of estrogen receptor positivity and complete pathological response to chemotherapy in breast cancer, found that the majority of combined multiple models had a better classification performance than the single models [21]. Furthermore, our findings also agree with a study by Taylor and Kim who, by splitting their original datasets into training and test parts, showed that voting based on nearest mean voters was a top performing method with respect to classification on lung or prostate cancer data. In contrast to our results, RF was found to perform equally well as the mean voter [44]. However, this discrepancy is likely caused by the difference in classification tasks. In contrast to our results, Statnikov and coworkers found that SVM-based ensemble/voting methods perform similar or worse compared to SVM nonensemble/voting methods, when tested on ten different human gene expression datasets by 10-fold cross-validation. However, these comparisons were primarily based on multicategory classification [45].

In the second experiment, we investigated the classifiers performance when tested in an external dataset based on a similar microarray platform. In this setup, RF and

R-SVM achieved the best performances, both significantly outperforming four of the five remaining classification methods. Furthermore, the voting procedure significantly outperformed the seven underlying classification methods. This suggests that every classifier in the voting committee agrees on most of the samples that are predicted correctly and that the majority of voters do not make the same misclassifications. The finding that voting and RF have the best performances could be explained by these methods' ability to reduce the cross-study prediction variance, without simultaneously increasing prediction bias [37]. Only a limited number of studies have compared the cross-study performance of multiple classification methods. A study by Tan and Gilbert compared the performance of single C4.5 classifiers with the voting-like C4.5 bagging and boosting classifiers on gene expression data cancer classification. In four of the experiments, an independent dataset measured on similar platforms was used for testing. Interestingly, one of the results found that bagging and boosting performed better than single C4.5 classifiers when predicting relapse within a 5-year period in a small group of 19 independent breast cancer patients, achieving 88.7%, 88.7%, and 75% bAcc, respectively [46]. These voting results are higher than the mean voting performance achieved in our two test sets (72.5% bAcc). This could be due to three factors: (1) the training and testing sets used by Tan and Gilbert originate from the same population (The Netherlands), and the sample preparations and gene expression measurements were performed using the same protocols; (2) the classification task also differs. It might be easier to predict relapse within a 5-year period than predicting if a patient would ever metastasize; (3) the voting methods used also differ. Another study deployed a committee of neural networks for gene expression based leukemia subclassification using three gene expression datasets measured on the same microarray platform. The study used a first dataset for feature selection, a second for network training and committee development, and a third independent test set for validating the committee. When compared to the performance by each of the underlying classifiers in the committee, the committee neural networks proved to perform better or equally well in the final testing set [47].

In the third experiment, the trained classifiers were externally validated on datasets based on a different microarray platform. With this setup, the performances dropped dramatically. This suggests that the data distributions in the Agilent and Affymetrix datasets are dissimilar. This is likely caused by biological and technical variation. The fact that the training and test samples originate from two different patient populations could make the data distributions dissimilar. The technical variation may originate from several sources, for example, the size of the oligonucleotides used, probe coverage, labeling, cross-hybridization, and detection limits by the scanner. Furthermore, the two platforms use different strategies for measuring the same RNA quantity. On the Agilent platform, this quantity is measured as the ratio of fluorescence intensities between a sample and a reference at each spot on the array, while the Affymetrix platform uses single channel measurements for a collection of probe sets covering one gene, which are therefore not comparable. To circumvent this obstacle, we standardized the datasets. However, this standardization seemed not to be sufficient for avoiding a drop in performance by all the classification methods used. Therefore, it is likely that the data distributions of the training and test sets are very heterogeneous thus hampering the external application of the classifiers.

Although all classification methods experience a drop in performance when validating between datasets measured on different platforms, the results showed that RF remained the strongest method and significantly outperformed the six other methods. Surprisingly, the voting procedure performed poorly when validated on data measured on a different microarray platform and was a borderline significantly outperformed by RF. This is probably due to randomness by each method/voter. The finding of RF performing better than the cross-method voting procedure suggests that when tested on datasets using a different microarray technology for gene expression measurement, voting procedures based on the same classification algorithm, in the case of random forest being a collection of decision trees, are more advantageous than voting procedures based on diverse classification methods. This implies that RF compared to voting is more capable of reducing the prediction variance associated with validation across studies and platforms. Therefore, in a situation when validating between different microarray platforms and where voting is outperformed, an approach called bagging might prove advantageous. Bagging uses voters consisting of multiple classifiers developed by bootstrap resamplings from the same dataset and based on the same classification method (decisions trees in the case of random forest) [48]. Thus, bagged SVM, LR, and NNET might be considered ideal for cross-study-cross-platform validations. RF may also be powerful, as the method is based on multiple decision rules, which might be better at segregating a complex data structure. This situation is in line with a study showing that molecular classification of cancer achieves better or similar performance as other classification algorithms, when using decision rules based on a single gene or a gene pair [49].

In the literature, there has been a limited number of studies comparing the performance of multiple classification methods, applied to across the dataset and microarray platform validations. One study by Yoshida compared a nearest template prediction method (NTP) with CART (single decision tree method), weighted voting, SVM, and k-nearest neighbor classification (k-NN) across datasets using Agilent datasets for training and Affymetrix datasets for testing [20]. For prediction of estrogen receptor positivity in breast cancer, NTP had the best performance, while SVM had the worst performance. For predictions of breast cancer molecular subtypes, SVM had the best performance in two of three testing sets used for this purpose, while NTP had the best performance in the third dataset. The worst performing methods were achieved by CART and k-NN [20]. In our study, SVM was not a top performing method for cross-platform testing. These differences are likely due to two factors: first, the study by Hoshida did not apply the entire classifier to the test sets, but only the list of genes defined

by the training datasets. This list was used to train and test a classifier in the validation dataset; second, the study addresses completely different classification tasks compared to our study.

Our results showed that when validation is applied between two datasets of similar or different microarray platforms, LR and NNET were among the poorest performing methods.

The general poor performance of LR could be due to several factors. First, a strong LR model is frequently composed of predictors being highly univariate significant and remains significant in the multivariate model. The fact that the list of 519 rank-significant genes defined by the eight feature definer datasets was reduced to a pool of 283 genes could have led to the exclusion of some highly significant genes, due to the only reason that they were not shared by the AM, RO, TR, and MA datasets, thus impairing the possibility for development of a stronger model. This could explain the poor performance by LR classifiers developed by the individual AM and RO datasets; second, an LR model requires a large sample size for providing robust maximum likelihood parameter estimation. Although the training datasets contain 151 and 286 samples, these sample sizes may not be sufficient for developing a strong model if some of the highly discriminative genes are absent; third, LR models rely on the assumption that there is no colinearity between the variables, meaning that the variables/features should be independent from each other. This assumption may be violated if predictors in a logistic model consist of, for example, gene expression features, some of which could be coregulated, thus leading to colinearity and weakening the model; finally, LR is sensitive to outliers. As we have not removed any samples from our datasets, and the possibility of outlier presence thus could be evident, this could also hamper the predictive power of the LR models.

The finding of NNET had a high internal 10-fold cross validation performance but a weak external validation performance could suggest that the NNET classifiers are not very generable. Another explanation could be that the transfer function used by the neural network was a sigmoid function, which is identical to that used in logistic regression, thus leading to some of the weaknesses observed in logistic regression, although the parameter estimation in neural networks is not conducted by maximum-likelihood but by a gradient descent algorithm. Interestingly, a study has compared the classification performance of four different single hidden layer feedforward neural networks on three microarray gene expression cancer dataset, showing that an SVD-neural classifier based on a *tansig* activation function and using single value decomposition for parameter estimation had a better performance compared to the three other methods and that this classifier outperformed support vector machines, principle component analysis classifiers, and Fisher discriminant analysis classifiers [50]. This implies that using another neural network type could achieve a better performance when applied for external validation in datasets based on similar or different microarray platforms.

In conclusion, voting-based classifiers provided an advantage with respect to classifying metastasis outcome in breast cancer patients. When testing was performed within the same dataset or between datasets using similar microarray platforms, combining class decisions by multiple classification methods significantly increased the classification performance. Random forest, a voting-like method, proved to be the strongest method when testing was performed in datasets based on a different microarray platform.

Conflict of Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was funded by the Danish Ministry of Interior, the Danish Strategically Research Council and DBCG-TIBCAT, the Clinical Institute at the University of Southern Denmark and the Human Microarray Center associated with the Department of Clinical Genetics at the Odense University Hospital. Furthermore, the authors acknowledge authors and people making the microarray datasets used in this study publically available.

References

[1] E. Karlsson, U. Delle, A. Danielsson et al., "Gene expression variation to predict 10-year survival in lymph-node-negative breast cancer," *BMC Cancer*, vol. 8, article 254, 2008.

[2] H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis," *Molecular Systems Biology*, vol. 3, article 140, 2007.

[3] Q. Tan, M. Thomassen, and T. A. Kruse, "Feature selection for predicting tumor metastases in microarray experiments using paired design," *Cancer Informatics*, vol. 3, pp. 133–138, 2007.

[4] M. Thomassen, Q. Tan, F. Eiriksdottir, M. Bak, S. Cold, and T. A. Kruse, "Prediction of metastasis from low-malignant breast cancer by gene expression profiling," *International Journal of Cancer*, vol. 120, no. 5, pp. 1070–1075, 2007.

[5] R. Sabatier, P. Finetti, N. Cervera et al., "A gene expression signature identifies two prognostic subgroups of basal breast cancer," *Breast Cancer Research and Treatment*, vol. 126, no. 2, pp. 407–420, 2011.

[6] L. D. Miller, J. Smeds, J. George et al., "An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 38, pp. 13550–13555, 2005.

[7] M. H. van Vliet, R. Fabien, H. M. Horlings, M. J. van de Vijver, M. J. T. Reinders, and L. F. A. Wessels, "Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability," *BMC Genomics*, vol. 9, article 375, 2008.

[8] M. Thomassen, Q. Tan, F. Eiriksdottir, M. Bak, S. Cold, and T. A. Kruse, "Comparison of gene sets for expression profiling: prediction of metastasis from low-malignant breast cancer," *Clinical Cancer Research*, vol. 13, no. 18, part 1, pp. 5355–5360, 2007.

[9] N. Servant, M. A. Bollet, H. Halfwerk et al., "Search for a gene expression signature of breast cancer local recurrence in young women," *Clinical Cancer Research*, vol. 18, no. 6, pp. 1704–1715, 2012.

[10] T. Zeng and J. Liu, "Mixture classification model based on clinical markers for breast cancer prognosis," *Artificial Intelligence in Medicine*, vol. 48, no. 2-3, pp. 129–137, 2010.

[11] M. Garcia, R. Millat-carus, F. Bertucci, P. Finetti, D. Birnbaum, and G. Bidaut, "Interactome-transcriptome integration for predicting distant metastasis in breast cancer," *Bioinformatics*, vol. 28, no. 5, pp. 672–678, 2012.

[12] R. Díaz-Uriarte and A. S. de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, article 3, 2006.

[13] L. J. Lancashire, D. G. Powe, J. S. Reis-Filho et al., "A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks," *Breast Cancer Research and Treatment*, vol. 120, no. 1, pp. 83–93, 2010.

[14] A. Statnikov and C. F. Aliferis, "Are random forests better than support vector machines for microarray-based cancer classification?" *AMIA Annual Symposium Proceedings*, pp. 686–690, 2007.

[15] M. Pirooznia, J. Y. Yang, M. Q. Qu, and Y. Deng, "A comparative study of different machine learning methods on microarray gene expression data," *BMC Genomics*, vol. 9, supplement 1, article S13, 2008.

[16] M. Zucknick, S. Richardson, and E. A. Stronach, "Comparing the characteristics of gene expression profiles derived by univariate and multivariate classification methods," *Statistical Applications in Genetics and Molecular Biology*, vol. 7, no. 1, article 7, 2008.

[17] A. Statnikov, L. Wang, and C. F. Aliferis, "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," *BMC Bioinformatics*, vol. 9, article 319, 2008.

[18] J. Önskog, E. Freyhult, M. Landfors, P. Rydén, and T. R. Hvidsten, "Classification of microarrays; synergistic effects between normalization, gene selection and machine learning," *BMC Bioinformatics*, vol. 12, article 390, 2011.

[19] S. Y. Kim, "Effects of sample size on robustness and prediction accuracy of a prognostic gene signature," *BMC Bioinformatics*, vol. 10, article 147, 2009.

[20] Y. Hoshida, "Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment," *PLoS One*, vol. 5, no. 11, Article ID e15543, 2010.

[21] M. Chen, L. Shi, R. Kelly, R. Perkins, H. Fang, and W. Tong, "Selecting a single model or combining multiple models for microarray-based classifier development?—a comparative analysis based on large and diverse datasets generated from the MAQC-II project," *BMC Bioinformatics*, vol. 12, supplement 10, Article ID S3, 2011.

[22] M. Zervakis, M. E. Blazadonakis, G. Tsiliki, V. Danilatou, M. Tsiknakis, and D. Kafetzopoulos, "Outcome prediction based on microarray analysis: a critical perspective on methods," *BMC Bioinformatics*, vol. 10, article 53, 2009.

[23] S. Calza, P. Hall, G. Auer et al., "Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients," *Breast Cancer Research*, vol. 8, no. 4, article R34, 2006.

[24] C. Sotiriou, S. Y. Neo, L. M. McShane et al., "Breast cancer classification and prognosis based on gene expression profiles from a population-based study," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 18, pp. 10393–10398, 2003.

[25] C. Sotiriou, P. Wirapati, S. Loi et al., "Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis," *Journal of the National Cancer Institute*, vol. 98, no. 4, pp. 262–272, 2006.

[26] M. J. van de Vijver, Y. D. He, L. J. van't Veer et al., "A gene-expression signature as a predictor of survival in breast cancer," *New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.

[27] E. Huang, S. H. Cheng, H. Dressman et al., "Gene expression predictors of breast cancer outcomes," *The Lancet*, vol. 361, no. 9369, pp. 1590–1596, 2003.

[28] Y. Wang, J. G. M. Klijn, Y. Zhang et al., "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer," *The Lancet*, vol. 365, no. 9460, pp. 671–679, 2005.

[29] M. Buyse, S. Loi, L. van't Veer et al., "Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer," *Journal of the National Cancer Institute*, vol. 98, no. 17, pp. 1183–1192, 2006.

[30] M. Schmidt, D. Böhm, C. von Törne et al., "The humoral immune system has a key prognostic impact in node-negative breast cancer," *Cancer Research*, vol. 68, no. 13, pp. 5405–5413, 2008.

[31] L. J. van't Veer, H. Dai, M. J. van de Vijver et al., "Gene expression profiling predicts clinical outcome of breast cancer," *Nature*, vol. 415, no. 6871, pp. 530–536, 2002.

[32] C. Desmedt, F. Piette, S. Loi et al., "Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series," *Clinical Cancer Research*, vol. 13, no. 11, pp. 3207–3214, 2007.

[33] L. Nanni, S. Brahnam, and A. Lumini, "Combining multiple approaches for gene microarray classification," *Bioinformatics*, vol. 28, no. 8, pp. 1151–1157, 2012.

[34] M. Thomassen, Q. Tan, and T. A. Kruse, "Gene expression meta-analysis identifies metastatic pathways and transcription factors in breast cancer," *BMC Cancer*, vol. 8, article 394, 2008.

[35] M. Thomassen, Q. Tan, and T. A. Kruse, "Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis," *Breast Cancer Research and Treatment*, vol. 113, no. 2, pp. 239–249, 2009.

[36] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining Inference and Prediction*, Springer, New York, NY, USA, 2nd edition, 2009.

[38] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, Ny, USA, 1995.

[39] A. Dupuy and R. M. Simon, "Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting," *Journal of the National Cancer Institute*, vol. 99, no. 2, pp. 147–157, 2007.

[40] R. Hewett and P. Kijsanayothin, "Tumor classification ranking from microarray data," *BMC Genomics*, vol. 9, supplement 2, article S21, 2008.

[41] G. Natsoulis, L. El Ghaoui, G. R. G. Lanckriet et al., "Classification of a large microarray data set: algorithm comparison and analysis of drug signatures," *Genome Research*, vol. 15, no. 5, pp. 724–736, 2005.

[42] Y. Peng, "A novel ensemble machine learning for robust microarray data classification," *Computers in Biology and Medicine*, vol. 36, no. 6, pp. 553–573, 2006.

[43] C. J. Xu, H. C. J. Hoefsloot, and A. K. Smilde, "To aggregate or not to aggregate high-dimensional classifiers," *BMC Bioinformatics*, vol. 12, article 153, 2011.

[44] S. L. Taylor and K. Kim, "A jackknife and voting classifier approach to feature selection and classification," *Cancer Informatics*, vol. 10, pp. 133–147, 2011.

[45] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.
[46] A. C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," *Appl Bioinformatics*, vol. 2, supplement 3, pp. S75–S83, 2003.
[47] M. S. Sewak, N. P. Reddy, and Z. H. Duan, "Gene expression based leukemia sub-classification using committee neural networks," *Bioinformatics and Biology Insights*, vol. 3, pp. 89–98, 2009.
[48] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
[49] X. Wang and O. Gotoh, "Accurate molecular classification of cancer using simple rules," *BMC Medical Genomics*, vol. 2, article 64, 2009.
[50] H. T. Huynh, J. J. Kim, and Y. Won, "Performance comparison of SLFN training algorithms for DNA microarray classification," *Advances in Experimental Medicine and Biology*, vol. 696, pp. 135–143, 2011.