33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Evaluating Feature Selection Strategies for High Dimensional, Small Sample Size Datasets

Abhishek Golugula[1], George Lee[2], and Anant Madabhushi[2]
[1]Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854
[2]Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey 08854

***Abstract*— In this work, we analyze and evaluate different strategies for comparing Feature Selection (FS) schemes on High Dimensional (HD) biomedical datasets (e.g. gene and protein expression studies) with a small sample size (SSS). Additionally, we define a new feature, *Robustness*, specifically for comparing the ability of an FS scheme to be invariant to changes in its training data. While classifier accuracy has been the *de facto* method for evaluating FS schemes, on account of the *curse of dimensionality* problem, it might not always be the appropriate measure for HD/SSS datasets. SSS lends the dataset a higher probability of containing data that is not representative of the true distribution of the whole population. However, an ideal FS scheme must be robust enough to produce the same results each time there are changes to the training data. In this study, we employed the robustness performance measure in conjunction with classifier accuracy (measured via the K-Nearest Neighbor and Random Forest classifiers) to quantitatively compare five different FS schemes (T-test, F-test, Kolmogorov-Smirnov Test, Wilks Lambda Test and Wilcoxon Rand Sum Test) on 5 HD/SSS gene and protein expression datasets corresponding to ovarian cancer, lung cancer, bone lesions, celiac disease, and coronary heart disease. Of the five FS schemes compared, the Wilcoxon Rand Sum Test was found to outperform other FS schemes in terms of classification accuracy and robustness. Our results suggest that both classifier accuracy and robustness should be considered when deciding on the appropriate FS scheme for HD/SSS datasets.**

I. Introduction

High dimensional (HD) data streams including protein, gene-expression, and other molecular assays are being routinely acquired in the context of disease diagnosis and prognosis [1]. The objective then is to be able to build integrated classifiers to leverage this high dimensional data towards making diagnostic and/or prognostic decisions. However, building classifiers based off HD Small Sample Size (HD/SSS) data is typically difficult on account of the curse of dimensionality ($K >> n$ where K is the number of dimensions and n is the sample size) [2]. One recourse to building classifiers for a HD/SSS dataset (denoted hereafter by $\alpha \in \mathbb{R}^{n \times K}$) is to use a Dimensionality Reduction (DR) technique such as Principal Component Analysis (PCA) [3] to reduce α from size $n \times K$ to $n \times \hat{K}$, where $K >> \hat{K}$. However, the reduced featured space may contain noisy features that can affect the classifier's performance [2].

In such cases, studies have shown that performing Feature Selection (FS) before PCA can significantly improve classifier performance [3]. FS consists of selecting the most informative features that can best stratify the data based on its attribute profile into different categories [3]. FS methods reduce α initially to an intermediate subset $F \in \mathbb{R}^{n \times k}$, where $\hat{K} << k << K$, and where F only contains significant, class discriminating features as identified by a FS method. After the FS method chooses the most relevant data, PCA can further transform this data into a reduced subspace of $\hat{K}$ features which allows for representation of the original data into far fewer dimensions.

In a dataset with a large sample size and few features, the effects of outliers will be minimal and the training data will be representative of the population at large. The resultant F will have a high probability of containing the most relevant features [4]. However, in the HD/SSS case, the values of few outliers can drastically change the set of extracted features and this new set of potential noisy features may not adequately reflect or capture class-specific differences. An ideal FS scheme should be robust enough to overcome the effects of these outliers and still extract a set of discriminatory features. This raises two interesting questions:

(1) Which FS method yields the best F? The evaluation of FS schemes has been restricted to classification performance. However, this measure may not be appropriate in the context of HD/SSS, due to insufficient sampling, making it hard to establish the best F [5]. Moreover, similar classification accuracies do not imply that consistently similar sets of F will be produced by the same FS schemes [6]. This indicates that several sets of F may show similar classification performance. Thus, classifier accuracy alone is insufficient to evaluate an FS scheme in the HD/SSS context.

(2) Would this FS method obtain similar results with a SSS? A SSS makes the data unreliable as an adequate predictor of the population at large [7]. Therefore, there is a need to compare various FS schemes and the classification results they yield on a HD/SSS dataset. However, there have not been very many experimental studies to compare the performance of FS schemes, specifically in the context of biological data which is prone to the HD/SSS problem [8].

To address the problem of selecting consistent feature sets in the HD/SSS context, we define the robustness of a FS scheme as its ability to extract similar sets of F independent of the input training data. Denote F_N as the set of k features extracted in a dataset with a large sample size and $F_{HD/SSS}$ as the set of k features extracted in the presence of HD/SSS. Kupinski et al. [4] tries to investigate the probability that the features in F_N appear in $F_{HD/SSS}$. Canul-Reich et al. [6] compare the intersection of $F_{HD/SSS}$ produced by different FS schemes. However, to the best of our knowledge,

This work was made possible by grants by the Walter H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535, R01CA140772, and R03CA143991), Department of Defense Prostate Cancer Research Program, The Cancer Institute of New Jersey and the Society for Imaging Informatics in Medicine.*

a quantitative measure to evaluate the consistency of FS schemes has never been proposed.

The goal of our new FS based performance measure, Robustness (ϕ^R), is to quantitatively measure the deterioration of the performance of a FS scheme for HD/SSS data. A high ϕ^R does not imply that an optimal F, containing the most informative features of α, was extracted from the dataset. It only implies that the FS scheme has the ability to consistently pick the same features in F across different training sets. Similarly, a high classification accuracy (ϕ^A) does not imply that an optimal FS strategy was used. Clearly both measures, ϕ^A and ϕ^R, must be used together to identify an appropriate FS scheme. In this paper, we establish such a measure to determine which FS schemes extract the best F in terms of ϕ^R and ϕ^A for the HD/SSS case.

In this work, we examine the results of 5 different FS schemes (*t*-test, F-test, Kolmogorov-Smirnov Test, Wilks Lambda Test and Wilcoxon Rand Sum Test) on publicly available gene and protein expression datasets. After using these FS methods and PCA to reduce α to size $n \times \hat{K}$, we calculate classification accuracies, using K-Nearest Neighbor (KNN) and Random Forest (RF) classifiers, and the Robustness measure for HD/SSS data. Using these two measures, we attempt to determine the optimal FS scheme when dealing with HD/SSS gene and protein expression studies.

Our main contributions in this paper are to:

- Establish a novel measure, Robustness (ϕ^R), that measures the ability of an FS scheme to extract similar sets of features independent of changes to the input training data for the HD/SSS case.
- Find the optimal FS scheme in the case of a bioinformatics application based on Robustness (ϕ^R) and classifier accuracy (ϕ^A).

II. Robustness for evaluating an FS scheme

A. Establishing Variance ($\phi_{\mathcal{N}}$) of an FS scheme

We define $S = \{x_1, x_2, x_3 ... x_n\}$ as the set of samples and $\psi : S^n \rightarrow \mathbb{R}^{1\times k}$ to be a function that extracts the k most significant features as ranked by a specific FS scheme based on a subset $D \subset S$. Assume that the cardinality of $|D| = mn$ and that D can be created by randomly sampling $m\%$ of the n samples in S. This process can be repeated Q times such that $D_1, D_2, D_3 ... D_Q \subset S$ are created. The resultant set of extracted k features from each D_u is $F_u = \psi(D_u)$, where $u \in \{1, 2, ...Q\}$ and ψ may be a FS scheme. Because each D_u will be different, each resultant F_u may be different. Thus, we define a measure $\phi_{\mathcal{N}}$ associated with ψ to determine the variance in feature subsets as a function of D_u.

$$\phi_{\mathcal{N}} = \left| \bigcup_{u=1}^{Q} F_u \right| \tag{1}$$

where $\bigcup$ refers to the union of sets. Despite different training sets of D_u, an ideal FS scheme should produce similar sets of F_u and thus a low $\phi_{\mathcal{N}}$. Failure to produce similar sets of F_u would result in a higher $\phi_{\mathcal{N}}$. $\phi_{\mathcal{N}}$ thus measures the robustness of an FS scheme to changes in the training data. We now establish the range of $\phi_{\mathcal{N}}$ in Propostion II.1.

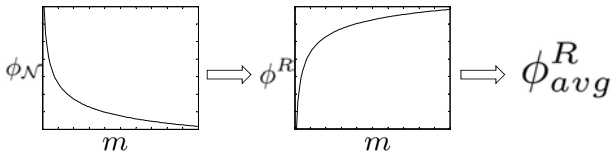

Fig. 1. Derivation of Robustness term, ϕ^R_{avg}. $\phi_{\mathcal{N}}$ is first calculated across a range of m where a lower m will result in a higher number of noisy features being identified. Next, ϕ^R is calculated from $\phi_{\mathcal{N}}$ where a higher m indicates higher robustness. Finally, ϕ^R_{avg} is calculated as the average across m.

Proposition II.1. *For $k \in \{1, 2, ...K\}$, $m \in (0, 1]$, $D_u \subset S$, where $u \in \{1, 2, ...Q\}$, it follows that $k \leq \phi_{\mathcal{N}} \leq Qk$.*

Define a general β where $\beta \in \{1, 2, ...Q\}$:

(i) $\phi_{\mathcal{N}} = \left| \bigcup_{u=1}^{Q} F_u \right| \geq |F_\beta| = k$

(ii) $\phi_{\mathcal{N}} = \left| \bigcup_{u=1}^{Q} F_u \right| \leq Q\,|F_\beta| = Qk$

B. Extending Variance ($\phi_{\mathcal{N}}$), to Robustness (ϕ^R)

A high $\phi_{\mathcal{N}}$ for a FS scheme indicates a high tendency to pick up different features in the sets of $\hat{D}_1, \hat{D}_2, \hat{D}_3, ..., \hat{D}_Q \subset S$, where $u \in \{1, 2, ...Q\}$. Using the range of $\phi_{\mathcal{N}}$, from Proposition II.1, a new normalized measure called Robustness (ϕ^R), can be derived in the following manner:

$$\phi^R = \frac{Q}{Q-1}\left(1 - \frac{\phi_{\mathcal{N}}}{Qk}\right) \tag{2}$$

Since, $k \leq \phi_{\mathcal{N}} \leq Qk$, the resultant ϕ^R will also be bounded such that $0 \leq \phi^R \leq 1$. For a FS scheme, $\phi^R = 1$ will signify the most robustness and $\phi^R = 0$ will signify the least robustness (orthogonal sets of F picked).

A singular value, ϕ^R_{avg}, for a FS scheme can be obtained on a particular dataset by averaging the values of ϕ^R over a range of m, allowing for ϕ^R to be employed seamlessly over different types of datasets. We can define $m_1, m_2 \in \mathbb{R}$ such that $0 < m_1 < m_2 \leq 1$.

$$\phi^R_{avg} = \frac{1}{(m_2 - m_1)} \int_{m_1}^{m_2} \phi^R dm \tag{3}$$

C. Determination of the most optimal FS scheme

As discussed in Section 1, both ϕ^R_{avg} and ϕ^A have to be considered equally when deciding the optimal FS scheme. Since ϕ^R_{avg} and ϕ^A measure fundamentally different properties, a FS scheme that optimizes one measure may not necessarily also optimize the other measure. Since both measures are similarly scaled, $\phi^A, \phi^R_{avg} \in [0, 1]$, the optimal normalized score, ϕ^{Sc}, can be computed as:

$$\phi^{Sc} = (\phi^A)^{\omega_A} (\phi^R_{avg})^{\omega_R} \tag{4}$$

where $\omega_A \in [1, \infty)$, $\omega_R \in [1, \infty)$ and $(\omega_A, \omega_R) \in \{(1, [1, \infty)), ([1, \infty), 1)\}$ are variable weights for each measure adjusted depending on the needs of the application. In addition, $\min\{\omega_A, \omega_R\} = 1$ so that the measure with the lower weight is always 1.

D. Feature Selection (FS) Methods used in this study

We denote $\mathbb{W}_1$ and $\mathbb{W}_2$ as class 1 and 2, and μ_1 and μ_2 as the means, σ_1^2 and σ_2^2 as the variances, and n_1 and n_2 as the sample sizes of $\mathbb{W}_1$ and $\mathbb{W}_2$. we further denote σ^2 as the total variance of all the samples in both classes, F_1 and F_2 as the empirical cumulative distribution functions of $\mathbb{W}_1$ and $\mathbb{W}_2$, and b_i as the rank of the sample $x_i \in \mathbb{W}_2$ with respect to the rest of the samples. Table I summarizes the 5 FS schemes (*t*-test, F-test, Kolmogorov-Smirnov Test (KST), Wilks Lambda Test (WLT), and Wilcoxon Rand Sum Test (WRST) [9]) used in this study.

TABLE I
THE 5 FS SCHEMES USED IN THIS STUDY AND THE CORRESPONDING FORMULATIONS.

FS scheme	Score
t-test	$t = \frac{\|\mu_1 - \mu_2\|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
F-test	$F = \max\{\frac{\sigma_1^2}{\sigma_2^2}, \frac{\sigma_2^2}{\sigma_1^2}\}$
KST	$L = \max\{F_1(x) - F_2(x), F_2(x) - F_1(x)\}$
WLT	$\Lambda = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{(n_1+n_2)\sigma^2}$
WRST	$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_2} b_i$

III. EXPERIMENTAL DESIGN AND RESULTS

A. Dataset Description

5 publicly available binary class gene and protein expression datasets denoted respectively as α^1, α^2, α^3, α^4, and α^5, summarized in Table II, for Ovarian Cancer [10], Lung Cancer [11], Bone Lesions [12], Celiac Disease [13] and Coronary Heart Disease [14] were used in this study. The sample size of the data ranged from 132 to 253 patients and the dimensionality ranged from 12625 to 22125 features.

TABLE II
THE 5 HD/SSS DATASETS USED IN THIS STUDY

	Dataset	Samples	Features	Description
α^1	Ovarian Cancer	253	15154 proteins	162 Tumor, 91 Normal
α^2	Lung Cancer	187	22125 genes	97 Cancer, 90 Control
α^3	Bone Lesions	173	12625 genes	137 Lesions, 36 Normal
α^4	Celiac Disease	132	18981 genes	110 Celiac, 22 Healthy
α^5	Coronary Heart Disease	153	20589 genes	87 Atherosclerotic, 66 Control

B. Experimental Data and Results

The ϕ_{avg}^R and ϕ^A of the 5 datasets (α^1, α^2, α^3, α^4, and α^5) are first calculated using the 5 FS schemes (*t*-test, F-test, KST, WLT, and WRST). The ϕ^{Sc} values are then computed using the values of ϕ_{avg}^R and ϕ^A across the five datasets for a FS scheme to determine which FS scheme is most optimal.

TABLE III
Experiment 1: ϕ_{avg}^R ACROSS ALL 5 DATASETS USING 5 DIFFERENT FS SCHEMES

Tests	α^1	α^2	α^3	α^4	α^5	mean
T-test	0.978	0.897	0.810	0.738	0.742	0.833± 0.104
F-test	0.973	0.915	**0.952**	0.730	**0.942**	**0.902± 0.098**
KST	0.968	0.870	0.787	0.733	0.705	0.813± 0.107
WLT	**0.985**	0.904	0.822	0.788	0.754	0.851± 0.093
WRST	**0.985**	**0.925**	0.838	**0.798**	0.782	0.866± 0.087

1) Experiment 1 - Evaluation of FS schemes via Robustness measures ϕ^R and ϕ_{avg}^R:

Using $Q = 100$, $k = 100$, and $.2 \leq m \leq .99$, ϕ_{avg}^R and the mean value of ϕ_{avg}^R across the 5 datasets were calculated for all five FS schemes. In addition, ϕ^R (which measures robustness at each value of m) is also shown for dataset α^1.

As m decreases, the data would lose the ability to properly represent the overall population because of the SSS. Consequently, ϕ^R of the FS schemes would deteriorate. This is confirmed in Figure 2 that as m decreases, ϕ^R also decreases. In Table III, we see the overall ϕ_{avg}^R calculated for each FS scheme on each dataset. Overall, F-test has the highest ϕ_{avg}^R value of all the schemes (mean of $\phi_{avg}^R = .902$ across all 5 datasets) and WRST had the best ϕ_{avg}^R in 3 of the 5 datasets. The *t*-test, KST, WLT had lower values with mean of $\phi_{avg}^R = .833$, $\phi_{avg}^R = .813$, and $\phi_{avg}^R = .851$ across the 5 datasets respectively.

2) Experiment 2 - Evaluation of FS schemes via classification accuracies ϕ^A:

Subsets of F were extracted with $k = 100$ as ranked by the 5 FS schemes. PCA was then used to reduce each dataset to $\hat{K} = 3$. The samples were randomly split using 10-fold cross validation and classification accuracies were calculated using classifiers KNN, with k = 1, and RF, with 250 Trees.

The four tests, *t*-test, KST, WLT, and WRST, all have a similar ϕ^A (with the WRST being marginally higher than the rest) across both classifiers in Table IV and V. Only the F-test compares the variance of the distributions without considering the means. Because it only considers variance, it seems to underperform the other tests in terms of classification accuracy. However, it is also important to note that there is no one FS scheme that consistently yields the highest classification accuracy across all datasets.

3) Experiment 3 - Determination of the optimal FS scheme in the presence of HD/SSS:

The means of ϕ_{avg}^R and ϕ^A across the 5 datasets in Experiments 1 and 2 were used to determine the ϕ^{Sc} values for all 5 FS schemes. $(\omega_A, \omega_R) = (1, 1)$ is used to allocate equal weight to both measures.

Even though the F-test has the highest ϕ_{avg}^R value, it has a very low ϕ^A value across both classifiers. On the other hand, *t*-test, KST, WLT, and WRST have similar ϕ^A, but WRST has a much higher ϕ_{avg}^R. With $(\omega_A, \omega_R) = (1, 1)$, these results indicate that WRST is the most optimal FS scheme. Furthermore, WRST ceases to be the most optimal FS scheme only for $\omega_R > 2.8$ for KNN and $\omega_R > 1.8$ for RF. At these values, the F-test, which has the highest ϕ_{avg}^R, becomes the most optimal FS scheme. Regardless, because WRST is the most optimal for most of the domain

TABLE IV

Experiment 2: KNN ϕ^A (%) WITH ± VARIANCE (%) ACROSS ALL 5 DATASETS USING 5 DIFFERENT FS SCHEMES

Tests	α^1	α^2	α^3	α^4	α^5	mean
T-test	92.32±0.33	**61.37±1.42**	70.83±1.11	80.36±1.29	59.06±1.90	72.99
F-test	78.42±0.61	53.11±1.25	**72.44±0.92**	70.93±1.18	54.50±1.43	65.88
KST	92.98±0.11	60.53±1.48	70.83±1.04	77.71±1.24	**59.75±1.62**	72.36
WLT	**95.09±0.05**	59.74±1.19	71.44±1.27	79.59±0.95	56.50±1.59	72.47
WRST	94.81±0.07	61.11±1.12	70.94±1.26	**83.07±0.81**	58.94±1.44	**73.77**

TABLE V

Experiment 2: RF ϕ^A (%) WITH ± VARIANCE (%) ACROSS ALL 5 DATASETS USING 5 DIFFERENT FS SCHEMES

Tests	α^1	α^2	α^3	α^4	α^5	mean
T-test	90.10±0.33	**64.42±1.01**	72.11±0.98	**84.50±1.28**	57.94±1.56	73.81
F-test	79.58±0.94	52.26±0.98	**79.28±0.90**	79.53±1.09	54.50±2.04	69.03
KST	90.11±0.15	62.79±1.18	73.89±1.15	81.36±0.81	**60.62±1.21**	73.75
WLT	90.91±0.22	63.79±1.10	76.67±0.90	83.71±1.09	55.31±1.50	74.08
WRST	**91.17±0.52**	63.21±1.05	75.33±0.79	82.29±0.99	59.19±1.81	**74.24**

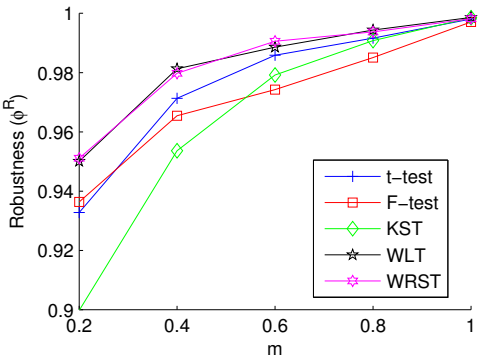

Fig. 2. *Experiment 1:* Comparison of ϕ^R accross the 5 FS schemes with respect to m, the percentage of sample size, for dataset α^1

TABLE VI

Experiment 3: VALUE OF ϕ^{Sc} FOR ALL 5 FS SCHEMES WITH THE TWO CLASSIFIERS KNN AND RF

Tests	KNN	RF
T-test	0.608	0.615
F-test	0.595	0.623
KST	0.588	0.599
WLT	0.616	0.630
WRST	**0.639**	**0.643**

of (ω_A, ω_R), we may conclude that it obtains the best F that solves the issue of both HD and the problem of a SSS.

IV. CONCLUDING REMARKS

High Dimensional (HD) datasets combined with a Small Sample Size (SSS) are common in the field of bioinformatics. Feature Selection (FS) schemes are a common way to reduce the dimensionality of biomedical datasets but analysis of the performance of FS schemes has been limited to classification accuracy (ϕ^A). Such a measure has a high error rate in the context of HD/SSS data. In this paper, we establish a new measure, Robustness (ϕ^R), to gauge the performance of an FS scheme specifically in the context of HD/SSS. ϕ^R measures the ability of an FS scheme to extract a consistent set of features and evaluates the robustness of the scheme to changes in training data. However, this measure does not reflect the ability of an FS scheme to extract informative features from a dataset. Both ϕ^R and ϕ^A must be used in conjunction to truly gauge the performance of an FS scheme. We analyze 5 different FS schemes (*t*-test, F-test, KST, WLT and WRST) and compare them using two classifiers (KNN and RF) and our new measure ϕ^R. Of the 5 FS schemes, WRST seems to be the better FS scheme with a relatively higher ϕ^R and ϕ^A compared to other schemes ($\phi^{Sc} = .639$ and $\phi^{Sc} = .643$ with KNN and RF respectively, using $(\omega_A, \omega_R) = (1, 1)$). Finally, we acknowledge that the 5 FS schemes used in this study are based on different assumptions about the data, such as the parametric assumptions made by *t*-test and F-test. However, our experimental framework did not incorporate any prior knowledge about the underlying distributions of the datasets (since we did not want to bias any of the FS schemes). In future work, we intend to explore the validity of our conclusions by including additional FS schemes on a larger number of datasets.

REFERENCES

[1] A. Madabhushi *et al.*, "Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data." *CMIG*, Feb 2011.

[2] Y. Wu *et al.*, "Optimal multimodal fusion for multimedia data analysis," *ACM Conf. on Multimedia*, pp. 572 – 579, 2004.

[3] G. Lee *et al.*, "Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies," *IEEE Trans. Comp. Biol. and Bioinf.*, vol. 5(3), pp. 368 – 384, 2008.

[4] M. A. Kupinski and M. L. Giger, "Feature selection with limited datasets," *Medical Physics*, vol. 26, p. 21762182, 1999.

[5] S. Raudys and A. Jain, "Small sample size effects in statistical pattern recognition: recommendations for practitioners," *PAMI*, vol. 13, no. 3, pp. 252 –264, Mar. 1991.

[6] J. Canul-Reich *et al.*, "Iterative feature perturbation method as a gene selector for microarray data," Intelligent Systems Labratory, Tech. Rep., January 2009.

[7] T. Way *et al.*, "Effect of finite sample size on feature selection and classification: A simulation study," *Medical Physics*, vol. 37, pp. 907–920, 2010.

[8] Q. Liu *et al.*, "Comparison of feature selection and classification for maldi-ms data," *IEEE Trans. on Comp. Biol. and Bioinf.*, July 2009.

[9] Y. Saeys *et al.*, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, pp. 2507–2517, 2007.

[10] E. F. Petricoin *et al.*, "Use of proteomic patterns in serum to identify ovarian cancer," *THE LANCET*, vol. 359, pp. 572–577, February 2002.

[11] A. Gustafson, R. Soldi, C. Anderlind, M. Scholand *et al.*, "Airway pi3k pathway activation is an early and reversible event in lung cancer development," *Sci Transl Med*, vol. 2, p. 26ra25, 21.

[12] E. Tian *et al.*, "The role of the wnt-signaling antagonist dkk1 in the development of osteolytic lesions in multiple myeloma," *N Engl J Med*, vol. 349(26), pp. 2483–94, 2003.

[13] G. Heap, G. Trynka, R. Jansen, M. Bruinenberg *et al.*, "Complex nature of snp genotype effects on gene expression in primary human leucocytes." *BMC Med Genomics*, vol. 2:1, 2009.

[14] S. H. Schirmer *et al.*, "Suppression of inflammatory signaling in monocytes from patients with coronary artery disease," *J Mol Cell Cardiol*, vol. 46, pp. 177–185, 2009.