Title

Evaluation of serum-based cancer biomarkers: a brief review from a clinical and computational viewpoint

Authors

Sohiya Yotsukura[1]* and Hiroshi Mamitsuka[1]

Affiliations

[1] Bioinformatics Center, Institute of Chemical Research, Kyoto University, Gokasho, Uji, 611-0011, Japan
*Corresponding Author: Sohiya Yotsukura
Bioinformatics Center, Institute of Chemical Research, Kyoto University, Gokasho, Uji, 611-0011, Japan
e-mail: yotsus@kuicr.kyoto-u.ac.jp
Phone: +81.774.38.3038 Fax: +81.774.38.3037

Abstract

Biomarkers are vital to detect diseases in various clinical stages. A variety of cancer serum biomarkers are already known, while for more accurate cancer-type detection, there required more rigorous evaluation manners, especially computational evaluation measures, for biomarkers. In this review, we first show three typical pitfalls in finding biomarkers and their examples, after briefly presenting standard five clinical biomarker screening phases by National Cancer Institute. We then introduce current computational biomarker evaluation measures, including current, standard methods with their intrinsic features. We further show an up-to-date list of existing cancer serum biomarkers, pointing out several issues, being caused by the limitations of current biomarker evaluation approaches. Finally we discuss the current attempts to develop new, statistically robust, computational serum-based biomarker measures in terms of specificity to each of various cancer types.

Keywords

Serum biomarkers, Quantitative assessment, Computational evaluation, Predictive values, likelihood ratios, Receiver Operating Characteristic Curve, Diagnostic odds ratio

Abbreviations

Prostate-specific antigen (PSA), true positive (TP), false positive (FP), true positive rate (TPR), false negative (FN), false negative rate (FNR), false positive rate (FPR), receiver operating characteristic (ROC), area under the curve (AUC), diagnostic odds ratio (DOR), Risk of Ovarian Cancer Malignancy Algorithm (ROMA), cell-free DNA (cfDNA), integrated discrimination improvement (IDI),

net reclassification improvement (NRI), Early prostate cancer antigen-2 (EPCA-2)

Vitae

Sohiya Yotsukura is a PhD candidate of the School of Pharmaceutical Sciences at Kyoto University. Her research interests include computational biology and machine learning techniques to analyze the functional impact of SNPs and tumorigenic somatic mutations on biological systems.

Hiroshi Mamitsuka is a professor at the Institute for Chemical Research in Kyoto University, jointly appointed as a professor at the School of Pharmaceutical Sciences of the same university. His research interests include machine learning, data mining and their applications in bioinformatics and chemoinformatics.

1. Introduction

Biomarkers are believed to increase the accuracy of diagnosis to precisely characterize the disease in a diagnostic or prognostic level. Biomarkers predict the response of the patient, helping to guide a more tailored treatment for the patient. Serum biomarkers are more appealing due to their simplicity of obtaining the blood samples. There are several serum cancer biomarkers, which are routinely used in clinical oncology, e.g. prostate specific antigen (PSA) for prostate cancer and cancer antigen (CA)-125 for ovarian cancer. However, their applications have significant limitations, because of low specificity, i.e. small probability of samples with no biomarkers in all non-diseased samples. In fact, the issue of specificity has become much more acute, since more than 30% or higher circulatory PSA level patients have to go for extensive testing and treatment, indicating its lack of specificity of prostate tumor detection [1]. In summary, lack of specific serum biomarkers has impeded the change in morbidity and mortality in cancer patients.

The traditional "a priori" approach for biomarker development needs a well-established biological procedure, being subjected to two-step clinical validation: 1) simple test with a high level of quality control, and 2) planned statistical prospective evaluations within the validation pilot studies to prove an established clinical impact [2]. Contrarily, more recently "a posteriori" approaches evaluate the clinical rationale of a "biological indicator" through a systematic discovery of various screening tools (e.g. microarray, bioinformatics, High-throughput DNA sequencing). These biological instruments are "black boxes", meaning that a clinical usage can be discovered through research pilot studies. Computational approaches give possible candidates for detecting certain diseases, by "sensitivity" and "specificity", within a patient population, but the proper quantification of a single biomarker in serum is limited to the evaluation technicalities. Therefore this review will focus on the latter approach since the recent technologies provide a plethora of potential candidates which are in proper need of evaluation.

Over the past twenty years, biomarkers have shown significant promise in the mechanism of how it will transform a patient's treatment. Therefore, biomarker research has been aimed towards the development of personalized targeted therapy. Despite the recent technological advancements, there are still relatively few biomarkers that are in routine clinical use today [3]. With a growing number of complex genomic tests for biomarker signatures becoming commercially available, the promise of personalized medicine is fast becoming a reality. Much attention has to be placed on the reason why the promising biomarkers and the biomarkers signatures entering the clinic is a long road ahead [4].

2. Biomarker Discovery Validation: Three Pitfalls

In this section, we first briefly show the most widely accepted guideline for evaluation and validation of biomarkers (Diagram 1): "Early Detection Research Network (EDRN)" developed by National Cancer Institute [5]. We then explain typical pitfalls and their examples of clinical biomarker evaluation failures, mainly caused by poor experimental design and inappropriate choice of the diagnostic assay.

Phase I of EDRN is the discovery of biomarkers through knowledge–based gene selection, gene expression profiling or protein profiling by setting the platform to rank and select the biomarkers via their characteristics. Most biomarker candidates are obtained from organized and characterized cohort studies, tissue banks or clinical trials with active follow-ups. Phase II establishes a clear indication of the biomarker's intended use in clinic by checking the validity, portability and reproducibility of these biomarker assays in various samples amongst various laboratories and clinics. The sensitivity and the specificity determined during this stage, which assess the quality of the biomarker, in the designated assay for clinical usage. Phase III evaluates the sensitivity and the specificity of the biomarker in various other diseases, to see its potential predictive value to ascertain the disease occurrence. Phase IV assesses the sensitivity and the specificity on prospective cohorts [6], identifying the false negative samples by evaluating the extent and characteristics of the disease at the time of detection. This process estimates the false referral rate and evaluates the diagnostic features of the biomarker, e.g. the definition, stage, grade of the tumor types. Phase V evaluates the overall benefits and the risks of performing the new biomarker diagnostic test in a controlled screened population.

In 2011, there were 7720 publications on biomarkers usage, but only 407 of these were actually patented [7]. Surprisingly, from these 407 patented biomarkers, none have obtained FDA approval. This fact reflects how many studies report the discovery of different potential biomarkers, but most of them do not meet the criteria of high sensitivity and specificity, necessary to enter into the clinical setting. Moreover, there is a shortage of quality specimens for the validation studies. This subsequently pushes the biomarker candidate into pitfalls preventing it from approval. In conclusion, the standardization of quantitation for the quality and the validation of the candidate biomarker need to be urgently improved to decrease the excessive economical resources placed on weak candidates. We here raise three typical pitfalls in biomarker validation

and their examples:

Pitfall 1: incorrect false positive control
A lack of external validation studies cannot justify the potential biomarker on its specified performance. For example, the broad sensitivity scale (0.40-0.73) can be easily obtained under a fixed specificity of 0.95, even though the specimens were not obtained at the time of diagnosis [8].

Example: Management of prostate cancer has long employed biomarkers, from the stage of detection to treatment monitoring. In 1987, serum PSA was suggested as a biomarker since its increased quantity as detected in prostate cancer can represent abnormalities in the architecture and the vasculature of the prostate [9]. Thus, it was believed that PSA quantity was directly correlated with the advancing clinical stage of the cancer and is useful for monitoring of disease curative therapy. However, PSA-based screening for prostate cancer is plagued by false positives [9], resulting in an only 25-40% positive predictive value (i.e. high probability of a diseased sample in all samples with a biomarker present). The major reason for such high false positive rates was derived from the fact that several non-cancerous events may elevate the level of PSA. For instance, inflammation, infection, and trauma are more common causes of elevated serum PSA than cancer.

Pitfall 2: sample population size
Inappropriate statistical results may result from the small number of patients and control subjects in published studies [8]. The outcomes from the small sample size would reflect logical fallacies in candidate markers due to the deficiencies in the inadequate study design. Biomarker development is a team process [8]. Since most tests and assays are outsourced, conceptual blind spots may arise which eventually result in unjustified conclusions. For serum biomarker studies, bodily fluids (blood, urine, semen, etc) and tissues are collected from a group of patients of different disease stages, and are compared with a group of healthy persons. Since sensitivity improves with increased disease stage, hence the lack of consistency through observational studies results in 'contradicting published reports likely [resulted] from studies testing different patient populations, using different methodologies, and applying different [cut-offs] for a positive test' [10].

Example: Xu et al. reported that lysophosphatidic acid is a potential effective biomarker for ovarian carcinoma [11, 12]. In the Xu study, elevated levels of plasma lysophosphatidic acid were detected in 90% of phase I patients and subsequently in 100% of phases II, III, or IV patients. It was also detected in 100% of recurrent ovarian cancer patients. Hence, it was demonstrated that Lysophosphatidic acid was a far more superior potential biomarker than CA-125, which only had a reported sensitivity of 22% and 60% in their stage I and advanced patients, respectively. However, Baker independently validated the lysophosphatidic acid test for ovarian cancer and concluded that lysophosphatidic acid levels are indistinguishable from ovarian case patients from controls subjects [13]. Hence, the Xu study was plagued by deficiencies that existed in the studied population. In other words, the small sample size of 149

patients demonstrated the "strawman argument" [8]. In conclusion, lysophosphatidic acid was found to be a nonspecific marker that is highly affected by small changes in sample collection practices, as well as processing practices.

Pitfall 3: inadequate experimental results

Choosing the optimal analysis method is critical in biomarker search by proteomics and metabolomics [14]. Preparations of the sample for proteomic and metabolic analyses are very important and can prevent errors that always will affect the final results. Unfortunately, most published studies only analyze the sample once, which does not validate the reproducibility of the results. Studies should permit the deviation from the mean (i.e. the error in the measurement) to avoid the pitfall of sample biases [8, 14-16].

Example: Leman et al. reported that early prostate cancer antigen-2 (EPCA-2) was a novel putative prostate cancer biomarker that performed better than PSA for diagnosis, prognosis, and disease maintenance [17]. The study comprised of 385 men, reported a 92% specificity, for overall prostate cancer, in contrast to 65% for PSA [6, 8, 12, 17, 18]. After a thorough examination of the methods was performed, the assays show technical inadequacies and discrepancies. Diamandis reported that the Leman assay design would not be a "sensitive or specific measure" to measure EPCA-2 in serum [1, 19]. More specifically, the investigation shows flaws included reporting values that were beyond the detection limit of the assay and improper reagents used to "capture" EPCA-2. Since the assay was technically inadequate, the study resulted in misleadingly "highly promising" from unreproducible data. This study typically addresses a failure phenomenon, where the inhibition of the assay verses direct changes in the analysis [8].

Current biomarkers have been plagued by many pitfalls including the above three typical cases. In particular, the example in the first pitfall can be rephrased such that many protein biomarkers are not cancer specific; that is, conditions other than the disease itself, may cause a protein biomarker to be present or be at an elevated level. In addition, the levels of some protein biomarkers are not elevated in all patients with a particular cancer [18]. Therefore, the computational assessment is an important tool to validate the quality of the biomarker.

3. Computational Measures for Evaluating Biomarkers

Diagram 2 is a 2 x 2 contingency table, showing cross-correlate disease status with biomarker presence. This table is obtained, under a pre-determined threshold for the biomarker test, to examine the usefulness of the candidate biomarker in the diagnostic, prognostic, predictive appraisal of the disease. This table has four cells with samples labeled by True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN), which are mainly used to compute the measures introduced below. Hereafter we use each of these four labels as the number of the corresponding samples.

3.1 Sensitivity and specificity

Sensitivity or True Positive Rate (TPR) is the probability that the biomarker will be detected positive in the disease samples. In Diagram 2, sensitivity=Probability(Biomarker present/Disease present), which can be estimated by (TP/(TP+FN)). Whereas, specificity, known as True Negative Rate (TNR), is the probability that the biomarker will be absent in a non-diseased case and can be estimated by (TN/(FP+TN)) [20]. Both are scaled such that values close to 1 to indicate high accurate diagnostic value and a value of zero is equivalent undefined diagnostic value. The relationship between sensitivity and specificity can be complemented via 1–specificity, which is False Positive Rate (FPR).

A highly sensitive test will be positive in nearly all patients with the disease, but there may also be positive for many patients without the disease. Ideally, most patients without the disease should present negative test results. So specificity is defined to check the sensitivity over patients without the disease. The ideal biomarker to be used as a diagnostic, prognostic, predictive or pharmacogenic tool should be both sensitive and specific. Therefore, sensitivity and specificity combined assess the quality performance of the biomarker. One measure for this purpose is Youden index, defined as (sensitivity + specificity -1) [20].

3.2 Predictive values

Predictive values are conditional proportional measures of the total cohort with positive and negative index (Diagram 2). In other words, a positive predictive value is probability that a case with a biomarker present is diseased and vise-versa for the negative predictive value. The positive predictive value decreases proportionally with the prevalence of the disease, indicating the biomarker assessment being not reliable for rare conditions, which will result in many FP results, compared to the TP results.

3.3 Likelihood ratios

The likelihood ratio is a diagnostic test to assess of the biomarker's sensitivity and specificity of detecting the disease. If the test is perfect, all patients tested positive would have the disease. Since this is not the case and there are patients that test positive without the disease, the discrepancy must be corrected for the TPR by the FPR, resulting in the positive likelihood ratio, being defined as (sensitivity)/(1-specificity) or by TPR/FPR [6]. Similarly, the validity of a negative result is still in question due to the presence of FN patients that exhibit false results with the disease present. The negative likelihood ratio is then defined as (1-sensitivity)/(specificity) or FNR/TNR. The likelihood ratios have two advantages: 1) it is the ratio of sensitivity and specificity alone, by which it is independent of population settings and can be used at the individual patient level, and 2) it allows a quantification of the probability of disease for a particular patient [21].

3.4 Receiver Operating Characteristic (ROC) Curve and Area under the ROC curve (AUC)

ROC is a graphical representation, which uses the values of sensitivity and specificity in varying amount of possible thresholds (Note: Diagram 2 is obtained under some certain threshold). The dependence between sensitivity and specificity induces a trade-off between the two quantities: one value increasing while the other decreases as the threshold for positivity is moved. The ROC curve is obtained by plotting sensitivity (TPR) against 1–specificity (FPR). Sensitivity and specificity calculated at various cut-off points generate the ROC curve, which is the main model used to assess the candidate biomarker correlation the disease phenotype [6].

As the positivity threshold is varied, the upper left hand corner of the ROC curve where sensitivity is 1 and specificity is 1. Similarly, the test would be uninformative if the ROC curve would be the upward diagonal line. If the distribution of the diseased and non-diseased groups are not similar, an asymmetrical curve will result indicating the limitation of this quantification assessment [22]. Therefore, the ROC curve is an indicator of the candidate's performance, since it shows a range of specificity of the candidate. Moreover, the sensitivity value of highly specific candidate is more important than that of low specific candidate.

The area under the curve of an ROC curve (AUC) is a way to reduce ROC performance to a single value representing expected performance. AUC can be interpreted as an average of sensitivity over all FPRs or as the average FPR over all sensitivities [23]. An AUC value of 0.5 (50%) indicates no association between true and false positives, and a value of 1.0 (100%) indicates perfect association [24] between the true and predicted outcome. AUC is a systematic summary over sensitivity over all possible cases (cutoff value), indicating a very sophisticated standard to measure the significance of binary classification. AUC considers all values of specificities without any weights on high specificity, which is a disadvantage of AUC.

3.5 Time-dependent ROC Curves and Time-dependent AUC

AUC has become the "gold standard" for assessing performance of models for binary outcomes. Kern highlights the "time" factor [25, 26]. The dimensionality of time incorporates the time-varying nature of the clinical onset time of the disease. Hence, the predictive accuracy for the marker can be pronounced since disease-occurrence can be precisely measured via calculations of time-specific ROC curves and further time-specific AUC. In fact time-specific ROC curves achieve optimal accuracy in predicting the future disease status, due to the advantage of capturing the time-varying nature of markers [25, 27].

3.6 Diagnostic Odds Ratio (DOR)

The Diagnostic Odds Ratio (DOR) is diagnostically defined as the positive odds of subjects with disease relative to the odds of subjects without disease. DOR can

range from zero to infinity. The higher values of DOR indicate better test performance [15]. A DOR of less than one indicates that the test can be improved simply by inverting the outcome of the test. A DOR value of exactly one means that the test is highly likely to predict a positive outcome, whatever the true condition. The definition of DOR can be written in a variety of ways, mainly by the following three:

1) (TP/FN)/(FP/TN),

2) a ratio of the positive likelihood ratio over the negative likelihood ratio, i.e. (TPR/FPR)/(FNR/TNR),

3) ((sensitivity)/(1-specificity))/((1-sensitivity)/(specificity)).

DOR depends significantly on the sensitivity and specificity of a test. A test with both high specificity and sensitivity with low rate of false positives and false negatives has high DOR. Maintaining same sensitivity of the test, DOR increases with the increase of the test specificity. For example, a test with sensitivity > 90% and specificity of 99% has a DOR greater than 500, with an approximate 95% confidence interval for the population. The p-value of DOR is <0.05, indicating studies exhibited great heterogeneity [20, 22, 28].

The contingency table of DOR is generated at some cutoff value, not over the average of specificity value like in AUC, while as DOR can be shown by sensitivity and specificity, so it can be connected to AUC by averaging it over all specificities by changing cutoff values. DOR summarizes study accuracy in a single number, focusing on some cutoff value. This point makes DOR a relatively easy measurement even for meta-analysis (which combines independent studies for a single biomarker to get effect in the general population). This is possible because DOR can be combined in terms of ratios of odds [28].

3.7 Issues with DOR

DOR is recently increasingly used, while the cut-off value needs to be chosen optimally. This is often by Youden index, while indexing a cut-off value to maximize some function such as specificity has become a common practice. For example, Lotrakul *et al.* recommended the cut-off value at 9, demonstrating that this cut-off value can be a powerful screening tool for patients with depression [29]. Bohning *et al.* proposed a solution to this issue by plotting a non-parametric estimate of the log-DOR against the cut-off value [20]. Another procedure provides a prospective statistical test for the hypotheses, to determine if the test treatment is beneficial for the entire patient population or only for that subgroup defined by the biomarker [30]. In other words, it provides an estimate of the optimal biomarker cut-off point. Glas *et al.* suggested the DOR as a single indicator of test performance to facilitate the formal meta-analysis of studies on diagnostic test performance [31], while still the precision of DOR has to be optimized by choosing the cut-off value [32]. We can thus say that it is still unclear on how to determine the cutoff value for DOR. Fischer and colleagues developed a recommended guide to enhance the appraisal of diagnostic tests to combine several measures of test accuracy, such as sensitivity, specificity, ROC, LR and DOR [33]. Also Pepe et al. showed the relation between DOR and AUC, and show that an OR as large as 3.0 may have little impact on AUC [5, 34, 35]. In conclusion, despite the increased use of DOR in the medical literature, it has

many flaws in its measurements and might still remain poorly understood by clinicians [28].

3.8 Latest Measures

Along with the recent trend on computational methods for evaluating biomarkers, we here raise new and rather under-developing measures on two cases: 1) two or more studies for a single biomarker, and 2) one study for multiple biomarkers.

Now various studies for a single biomarker can be generated [8, 15]. For example, even by the same research group, studies can be repeated at a certain time interval. We can analyze them to get its widened effect in the general population. Two (or more) studies can be summarized into so-called reclassification tables, from which sensitivity and specificity of each study can be derived. The reclassification table-based approaches have recently been gaining popularity, where a typical measure is "net reclassification improvement (NRI)" (which is slightly modified into IDI (integrated discrimination improvement) later) [26, 35]. NRI or IDI is a function, which sums sensitivity and specificity of one study. For example, for two diagnostic tests, NRI is defined as the difference between two tests, each being the sum of specificity and sensitivity:

$$\text{NRI} = (\text{Sensitivity} + \text{Specificity})_{\text{second test}} - (\text{Sensitivity} + \text{Specificity})_{\text{first test}}$$

NRI sums two rates (sensitivity and specificity) rather than a weighted average of the two rates based on the ratio of patients with disease to without disease. The sum of sensitivity and specificity is equivalent to Youden index, meaning that NRI compares the performance of two tests. Two performance results might be similar even if the performance is low. This implies that NRI checks the stability of two studies rather than examining the significance of the biomarker itself. Thus NRI would be more appropriate for evaluating the similarity of two different models rather than evaluating a single biomarker by two studies, where one model means a certain variable set or a particular condition such as a time point. So one possible case is that one model has a set of biomarkers and one different biomarker is added to this set to generate the other model.

The combination of multiple biomarkers would be one direction to achieve higher discriminative performance. For example, OVA1, a FDA approved test for pre-surgical evaluation of a woman's ovarian mass for cancer, examines five biomarkers: transthyretin, apolipoprotein, A-1, 2-microglobulin, transferrin, and CA-125 [36]. OVA1 demonstrated 90% sensitivity for ovarian cancers stages I compared with 61% for CA-125 [37]. Surprisingly, however, the specificity and positive predictive value of OVA1 with physician assessment was worse than CA-125 alone [38, 39]. One possible explanation for this result is the menopausal status that skews the results for this multivariate index assay. After OVA1, researchers have developed computational tools for diagnosis procedures with multiple biomarkers. ROMA (Risk of Ovarian Malignancy Algorithm) focuses on two biomarkers for ovarian cancer: HE4 and CA-125 [38, 39]. ROMA calculates "predictive index" (the final score given by the logit of the predictive index) by linear regression over the serum concentration of two biomarkers. An important point of ROMA is that those linear equations are changed by the patient's

menopausal status at the time of testing, to solve the low specificity problem of OVA1 [38]. ROMA is currently promoted for two major reasons, compared to OVA1. First, ROMA reports a higher specificity [38, 39]. Second, ROMA is more cost-effective, because it has less markers.

4. Current Biomarkers

Table 1 provides a comprehensive comparison of current potential and approved biomarkers, including 26 biomolecules, 2 metabolic biomarkers and 4 cell biomarkers. A unique feature of this table is positive and negative predictive values and DOR, which are extracted from corresponding references and attached to all biomarkers. Furthermore, one marker can be used for evaluating more than one cancer type. For example, heat shock proteins (HSPs) are used for five types: gastric, prostate, osteosarcoma, uterine and bladder, meaning five DOR values for all five types in HSPs. The DOR in this table takes a very wide range of values, while all values are reasonably larger than 1.0, except only seven cases. The largest DOR value is 218.5 for breast cancer by cfDNA, and the next largest is 104.816 for colorectal cancer by DR-70. Again we stress that this table is a thorough summary on current all possible cancer biomarkers.

One important issue of cancer biomarkers is to identify cancer-specific biomarkers, by which the cancer type of a patient would be detected more accurately. For example, DOR of cfDNA in a breast cancer patient is 218.5. The odds of having cfDNA due to a metastatic carcinoma is known as 2.29 [40], and even the odds ratio of having cfDNA in the blood from breast cancer can be 218.5/2.29=95.41. This will give researchers the conclusion that cfDNA will be 95 times more likely to be present in the blood of a breast cancer patient than in a patient with another type of carcinoma, meaning that cfDNA can be a breast cancer-specific biomarker. On the other hand, CA19-9 can be detected in three cancers: pancreatic, ovarian and bladder, where three DOR values are 15.637, 23.7 and 20.16, respectively, all being relatively high and similar. HSPs also has relatively similar values for five types, particularly two tumors, i.e. osteocarcinoma and prostate, having the same value, 2.25. In these cases, DOR would not be an appropriate measure to identify a cancer-type specific biomarker.

5. Discussion

As explained in Section 3.8, OVA1 (approved in 2009) for ovarian carcinoma with five biomarkers including CA-125 had to show weaker performance in terms of specificity and positive predictive value than CA-125 alone (approved in 1997). This implies the need of improved statistical quantification methods to decrease the lack of specificity in the current instrumentation methods for low abundance proteins. This point brought a mentality shift focusing on a more robust statistical predictive tool, resulting in the Risk of Ovarian Malignancy Algorithm (ROMA), more generally a predictive model, i.e. logistic regression over two variables [38].

The development of ROMA is, on evaluating biomarkers, solving many important

problems, which in general could not be overcome by previous methods. These problems would be summarized into three issues: 1) interpretation issue, 2) statistical error, and 3) technical (experimental) error.

1) Interpretation issue: previous measures like DOR have rather simple definitions and can be easily computed, while it might be hard for clinicians to understand them accurately and use them properly for diagnosis and prognosis purposes. On the other hand, ROMA, a predictive model, gives a score or a so-called risk, i.e. the probability that a patient will be with the disease (in the future). The score can be used for diagnosis directly, meaning that though the model behind the measure might be still a blackbox for clinicians, the model can show the score for any disease category like stage I to V of cancer progress more clearly. This would be more accepted by researchers/practitioners.

2) Statistical errors: one study has an unavoidable limitation on its sample size, sometimes causing a significant bias. To overcome this issue, a popular approach is meta-analysis, where independent studies are gathered and their statistical features, such as DOR, can be combined. Meta-analysis is a sophisticated manner to compensate many issues of a single study, particularly bias, while it is impossible to remove these issues completely, because usually meta-analysis uses only the statistics derived from the original studies and these statistics eventually have to keep the original issue. On the other hand, ROMA uses a predictive model, in which parameters, such as coefficients of the linear equations, should be estimated from the original instances, where each single assay can be one instance. This implies that a predictive model is more robust, being rather released from the bias generated in each original study, by which the predictive model makes meta-study easier than ever.

3) Technical errors: The current technological innovation has not achieved the sufficient sensitivity to properly identify the low abundance proteins. Until the instrumentation is drastically improved, highly predictive statistical quantification methods are needed to decrease the lack of specificity in the current instrumentation methods. One possible solution is definitely meta-analysis, and as explained above, predictive models, such as ROMA, would be a further solution for this problem.

Overall developing predictive models would be a highly promising direction to achieve higher diagnostic, prognostic and predictive performance for evaluating biomarkers. ROMA uses logistic regression, which is a widely used basic and powerful predictive model. Higher performance models than logistic regression are already developed in machine learning and are currently well-used in many applications. For example, regression trees and support vector regression were utilized for validating biomarkers already [41]. Further high-performance predictive methods might be used in the future, and might change the quality of cancer biomarker evaluation. Another possible direction would combine predictive models with previous measures like DOR. For example, ROMA with DOR might be a powerful technique that statistics can play to compensate the lack of precision and accuracy used to validate the candidate biomarker.

Disclosure of Potential Conflicts of Interest

The authors declare that they have no potential conflict of interest disclosed.

Tables

Table 1
A comprehensive list of currently popular or approved biomarkers.
The following table details the characteristics, applications and a quantitative measure of its quality as a clinical diagnostic tool in its respective cancer [3, 36, 44-109]

Figure Legends & Figures

Diagram 1
EDRN: Phases of Biomarker Development.
The large number of biomolecules that are initially considered as potential candidates pass through the other stages (Phase II-V) to validate its specificity and sensitivity in diagnosis, prognosis, and for treatment prediction. The long, costly and stringent road through the five-phase process filters the candidates to the 'certain few' that are FDA approved [5, 7].

Diagram 2
Biomarker Performance Characteristics: Quantification Assessment of Biomarker.

This contingency table distinguishes the population pool into the various subgroups (TP, FP, FN, TN), depending on the absence and the presence of the biomarker. This subsequently quantifies the sensitivity and the specificity of the biomarker correlated to the disease.

References:

[1] Heidenreich A BM, Joniau S, Mason MD, Mottet N, Schmid HP et al. Guidelines on prostate cancer. Eur Urol 2008;53:68-80.
[2] McGuire WL CG. Prognostic factors and treatment decisions in axillary-node-negative breast cancer. N Engl J Med. 1992;25:1756-61.
[3] Fuzery AK LJ, Chan MM, Chan DW. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clinical Proteomics. Clin Proteomics. 2013;10:13.
[4] EP D. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? BMC Med. 2012.
[5] Pepe MS ER, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of Biomarker Development for Early Detection of Cancer. J Natl Cancer Inst. 2001;93:1054-61.
[6] Kumar M SS. Biomarkers of diseases in medicine. Current Trends of Science. 2009:403-17.
[7] Drunker E KK. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. EPMA J. 2013;4:4-7.
[8] Kern. Why Your New Cancer Biomarker May Never Work: Recurrent Patterns and Remarkable Diversity in Biomarker Failures. Cancer Res. 2012;72:6097-101.
[9] Schröder FH CH, Wolters T, van den Bergh RC, Gosselaar C, Bangma CH et al. Early detection of prostate cancer in 2007. Part 1: PSA and PSA kinetics. Eur Urol. 2008;53:468-77.
[10] HB G. Are biomarkers for bladder cancer beneficial? J Urol. 2010;183:11-2.
[11] Xu Y SZ, Wiper DW, Wu M, Morton RE, Elson P et al. Lysophosphatidic acid as a potential biomarker for ovarian and other gynecologic cancers. JAMA. 1998;280:719-23.
[12] EP D. Cancer biomarkers: can we turn recent failures into success? J Natl Cancer Inst. 2010;102:1462-7.
[13] Baker DL MP, Miller B, et al. Plasma lysophosphatidic acid concentration and ovarian cancer. JAMA. 2002;287:3081-2.
[14] Drucker E KK. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive
and personalised medicine. EPMA Journal. 2013;4.
[15] Lewington S BF, Clarke R. A review on meta-analysis of biomarkers: promises and pitfalls. Clin Chem. 2012;58:1192-204.
[16] Buckler AJ PD, Ouelette M, Danagoulian J, Wernsing G, Suzek BE A novel knowledge representation Framework for the statistical validation of quantitiative imaging biomarkers. J Digit Imaging 2013;26:614-29.
[17] Leman ES CG, Trock BJ, Sokoll LJ, Chan DW, Mangold L et al. . EPCA-2: a highly specific serum marker for prostate cancer. J Urol. 2007;69:714-20.

[18] Li Y SZ, Wu Y, Babovic-Vuksanovic D, Li Y, Cunningham JM, Pankratz VS, Yang P. Cystic fibrosis transmembrane conductance regulator gene mutation and lung cancer risk. Lung Cancer. Lung Cancer. 2010;70:14-21.

[19] EP D. Early prostate cancer antigen-2 (EPCA-2): a controversial prostate cancer biomarker? Clin Chem. 2010;56:542-4.

[20] Bohning D HH, Patilea V. A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. Stat Meth Med Res. 2011;20:541-50.

[21] J A. Moving Beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. Aust Prescr. 2003;26:111-3.

[22] Grund B SS. Analysis of biomarker data: logs, odds ratios and ROC curves. Curr Opin HIV AIDS. 2010;5:473-9.

[23] P S. ROC analysis: applications to the classification of biological sequences and 3D structures. Brief Bioinform. 2008;9:198-209.

[24] Warnock DG PC. A roadmap for biomarker qualification. Nat Biotech. 2010;28:444-5.

[25] Zheng Y CT, Feng Z. Application of the time-Dependent ROC curves for prognostic accuracy with multiple biomarkers. Biometrics. 2005;62:279-87.

[26] Pencina MJ DARS, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27:157-72.

[27] Hanley JA MB. The meaning and use of the area under a receiver operating characteristic (ROC) curve. J Radiol. 1982;143:29-36.

[28] Grimes DA SK. Making sense of odds and odds ratios. Obstet Gynecol. 2008;111:423-6.

[29] Lotrakul M SS, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. BMC Psychiatry. 1998;8:46.

[30] Jiang W FB, Simon R. Biomarker adaptive threshold design: A procedure for evaluating treatment with possible biomarker-defined subset effect. J Natl Cancer Inst. 2007;99:1036-43.

[31] Glas AS LJ, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003;56:1129-35.

[32] Magder LS FA. Optimal choice of a cut point for a quantitative diagnostic test performed for research purposes. J Clin Epidemiol. 2003;56:956-62.

[33] Fischer JE BL, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. J Intensive Care Med. 2003;29:1043-51.

[34] Pepe MS FZ, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrative the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2007;167:362-8.

[35] Pepe MS JH, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159:82-890.

[36] J A. OVA1 test for preoperative assessment of ovarian cancer. Community Oncol. 2010;7:249-51.

[37] Ueland FR DC, Seamon LG, et al. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. Obstet Gynecol. 2011;117:1289-97.

[38] Chudeka-Glaz A C-PA, Menkiszak J, Sompolska-Rzechula A, Strojna A, Byra E et al. Preoperative diagnostic performance of ROMA (Risk of Ovarian Malignancy Algorithm) in relation to etiopathogenesis of epithelial ovarian tumors. J Mol Biomark Diagn. 2013.

[39] Montagnana M DE, Ruzzenente O, Bresciana V, Nuzzo T, Gelati M et al. The ROMA (Risk of Ovarian Malignancy Algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: is it really useful? Clin Chem Lab Med. 2011;49:521-5.

[40] Xu E SQ, Gu J, Chow WH, Ajani JA, Wu X. Association of mitochondrial DNA copy number in peripheral blood leukocytes with risk of esophageal adenocarcinoma. Carcinogenesis. 2013.

[41] Chen L XJ, Riggins RB, Clarke R, Wang Y. Identifying cancer biomarkers by network-constrained support vector machines. BMC Syst Biol. 2011;5:161-81.

[42] NR C. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115:928-35.

[43] Janes H PM, Gu W. Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med. 2008;149:751-60.

[44] T G. The new data on prostate cancer screening: What should we do now? Clev Clin J Med 2009;76:446-8.

[45] Gann PH MJ, Catalona WJ, Stampfer MJ. Strategies combining total and percent free prostate specific antigen for detecting prostate cancer: a prospective evaluation. J Urol. 2002;167:2427-34.

[46] Sterling RK JL, Gordon F, Sherman M, Venook AP, Reddy KR. Clinical utility of AFP-L3% measurement in North American patients with HCV-related cirrhosis. Am J Gastroenterol. 2007;102:2196-205.

[47] Yuen MF LC. Screening for hepatocellular carcinoma: survival benefit and cost-effectiveness. Ann Oncol. 2003;14:1463-7.

[48] Zinkin NT GF, Bhaskar K, Out HH, Spentzos D, Kalmowitz B, et al. Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. Clin Cancer Res. 2008.

[49] Partheen K KB, and Sundfeldt K. Evaluation of ovarian cancer biomarkers HE4 and CA-125 in women presenting with a suspicious cystic ovarian mass. J Gynecol Oncol. 2011;22:244-52.

[50] Bozkurt M YA, Aral I. Evaluation of the importance of the serum levels of CA-125, CA15-3, CA-19-9, carcinoembryonic antigen and alpha fetoprotein for distinguishing benign and malignant adnexal masses and contribution of different test combinations to diagnostic accuracy. Eur J Gynaecol Oncol. 2013;34:540-4.

[51] Okamoto T OT, Izuo M, Ito Y, Yamashita T, Tanaka R et al. . Decision making using postoperative CEA and CA 15-3 for detection of breast cancer recurrence. Breast Cancer. 1995;2:127-31.

[52] Stieber P ND, Heinemann V. Tumor markers in metastatic breast cancer: High tumor specificity within the reference range. J Clin Oncol. 2006;24.

[53] Passerini R CM, Boveri S, Salvatici M, Radice D, Zorzino L, et al. The Pitfalls of CA19-9: Routine Testing and Comparison of Two Automated Immunoassays in a Reference Oncology Center. Am J Clin Pathol. 2012;138:281-7.

[54] Gui JC YW, and Liu XD. CA19-9 and CA242 as tumor markers for the diagnosis of pancreatic cancer: a meta-analysis. CLin Exp Med. 2013.

[55] Bellehaninna UK CR. Serum CA 19-9 as a Biomarker for panceatic cancer - a comprehensive review. Indian J Surg Oncol. 2011;2:88-100.

[56] Cho HY KK, Jeon T, Kim YB, and No JH. CA19-9 elevation in ovarian mature cystic teratoma: Discrimination from ovarian cancer CA 19-9 level in teratoma. Med Sci Moni. 2013;19:230-5.

[57] Roy S DA, and Kar K. Comparison of urinary and serum CA 19-9 as markers of early stage urothelial carcinoma. Int Braz J Urol. 2013;39:631-8.

[58] Lüftner D MJ, Akrivakis C, Geppert R, Petrides PE, Wernecke KD et al. . Tumor type M2 pyruvate kinase expression in advanced breast cancer. Anticancer Res. 2000;20:5077-82.

[59] Barillari P BA, Chirletti P, et al. Role of CEA, TPA, and CA 19-9 in the early detection of localized and diffuse recurrent rectal cancer. Dis Colon Rectum. 1992;435:471-6.

[60] Palmqvist R EB, Lindmark G, Hallmans G, Tavelin B, Nilsson O et al. Prediagnostic levels of carcinoembryonic antigen and CA 242 in colorectal cancer: a matched case-control study. Dis Colon Rectum. 2003;46:1538-44.

[61] Mahdavi A PT, Nezhat F. Induction of ovulation and ovarian cancer: a critical review of the literature. ASRM. 2006.

[62] Gabrielli S RR, Pilu G, Pavani A, Capelli M, Milano V et al. . Accuracy of transvaginal ultrasound and serum hCG in the diagnosis of ectopic pregnancy ultrasound. Obstet Gynecol. 1992;2:110-5.

[63] Lempiainen A HK, Blomqvist C, Alfthan H, Stenman UH. Hyperglycosylated human chorionic gonadotropin in serum of testicular cancer patients. Clin Chem. 2012.

[64] Soyluk O BH, Aral F, Alagol F, Özbey NC. Papillary thyroid carcinoma patients assessed to be at low or intermediary risk after primary treatment are at greater risk of long term recurrence if they are thyroglobulin antibody positive or do not have distinctly low thyroglobulin at initial assessment. Thyroid. 2011.

[65] Baudin E DCC, Cailleux AF, Leboulleux S, Travagli JP, Schlumberger M. Positive predictive value of serum thyroglobulin levels, measured during the first year of follow-up after thyroid hormone withdrawal, in thyroid cancer patients. J Clin Endocrinol Metab. 2003;88:1107-10.

[66] Tan H XY, Xu J, Wang F, Nie S, Yang M et al. . Association of increased heat shock protein 70 levels in the lymphocyte with high risk of adverse pregnancy outcomes in early pregnancy: a nested case-control study. Cell Stress Chaperon. 2007;12:230-6.

[67] Shibata T AT, Tahara T, Yoshioka D, Maruyama N, Fujita H, Kamiya Y, Nakamura M et al. Protective role of genetic polymorphism of heat shock protein 70-2 for gastric cancer risk. Dig Dis Sci. 2009.

[68] Uozaki H IT, Kakiuchi C, Horiuchi H, Gotoh T, Iijima T et al. . Expressiom of heat shock proteins in osteosarcoma and its relationship to prognosis. Pathol Res Pract. 2000;196:665-73.

[69] Margel D P-FM, Baniel J, Yossepowitch O , Cohen IR. Stress proteins and cytokines are urinary biomarkers for diagnosis and staging of bladder cancer. Eur J Uro. 2011.

[70] King KL LA, Chau GY, Chi CW, Wu CW, Huang CL, Lui WY. Prognostic significance of heat shock protein-27 expression in hepatocellular carcinoma and its relation to histologic grading and survival. Cancer. 2000.

[71] Ulivi P ML, Casoni GL, Scarpi E, Bucchi L, Silvestrini R et al. Multiple marker detection in peripheral blood for NSCLC diagnosis. PloS One. 2013;8:e57401.

[72] Chen CL TH, Liu JC, Kashiwabara C, Feldman D, Sher L et al. . Reciprocal regulation by TLR4 and TGF-β in tumor-initiating stem-like cells. J Clin Invest. 2013;123:2832-49.

[73] González-Santiago AE M-TL, Sánchez-Llamas F, Troyo-Sanromán R, Gurrola-Díaz CM. TGF-β 1 serum concentration as a complementary diagnostic biomarker of lung cancer: establishment of a cut-point value. J Clin Lab Anal. 2011.

[74] Slattery ML HJ, Lundgreen A, Wolff RK. Genetic variation in the TGF-β signaling pathway and colon and rectal cancer risk. Cancer Epidemiol Biomarkers Prev. 2011.

[75] Mose F NL, Njunguna R, Tamooh H, John-Stewart G, Farquhar C, Kiarie J. Biomarker evaluation of self-reported condom use among women in HIV-discordant couples. Int J STD AIDS. 2013;24:537-40.

[76] Lazzeri M HA, Abrate A, de la Taille A, Redorta JP, McNicholas T et al. . Clinical performance of serum prostate-specific antigen isoform [-2]proPSA (p2PSA) and its derivatives, %p2PSA and the prostate health index (PHI), in men with a family history of prostate cancer: results from a multicentre European study, the PROMEtheuS project. BJU Intl. 2013;112:313-21.

[77] Yang Z LZ, Zhao B, Zhang W, Zhang J, Gi Z et al. . Diagnosis and preoperative predictive value of serum HE4 concentrations for optimal debulking in epithelial ovarian cance. Oncol Letters. 2013;6:28-34.

[78] Yesil A BG, Colak Y, Paker N and Gonen C. Prognostic significance of DR-70 levels in dysplastic colorectal polyps. Gastroenterol Res Pract. 2013.

[79] Shabaik A LG, Peterson M, Hasteh F, Tipps A, Datnow B, Weidner N. Reliability of Her2/neu, estrogen receptor, and progesterone receptor testing by immunohistochemistry on cell block of FNA and serous effusions from patients with primary and metastatic breast carcinoma. Diagn Cytopathol. 2011;39:328-32.

[80] Deandrea S TR, Foschi R, Montella M, Dal Maso L, Falcini F et al. Alcohol and breast cancer risk defined by estrogen and progesterone receptor status: a case-control study. Cancer Epidemiol Biomarkers Prev. 2008;17:2025-8.

[81] Huang HJ NP, Drijkoingen M, Paridaens R, Wildiers H, Van Kimbergen E etal. Association between tumour characteristics and HER-2/neu by immunohistochemistry in 1362 women with primary operable breast cancer. J CLin Pathol. 2005;58:611-6.

[82] Sekijima Y US, Tojo K, Sano K, Shimizu Y, Imaeda T et al. High prevalence of wild-type transthyretin deposition in patients with idiopathic carpal tunnel syndrome: a common cause of carpal tunnel syndrome in the elderly. Human Pathol. 2011;42:1785-91.

[83] Zhang J FP, Liu H, Bai H, Wang Y, Zhang F. Apolipoprotein A-I and B levels, dyslipidemia and metabolic syndrome in south-west Chinese women with PCOS. Human Reprod. 2012.

[84] Barroso-Sousa R LR, Patricia I, Mendonca R, Memoria RR, Spiller F et al. Decreased levels of alpha-1-acid glycoprotein are related to the mortality of septic patients in the emergency department. Clinics (Sao Paulo). 2013;68:1134-9.

[85] Mainous AG 3rd DV, Knoll ME, Hulihan MM, Grant AM, Wright RU. Transferrin saturation and hospital length of stay and mortality in Medicare beneficiaries. J Am Geriatr Soc. 2013;61:132-6.

[86] Badoux XC KM, Wang X, O'Biren SM, Ferrajoli A, Daferl S et al. Cyclophosphamide, fludarabine, alemtuzumab, and rituximab as salvage therapy for heavily pretreated patients with chronic lymphocytic leukemia. Blood. 2011;118:2085-93.

[87] Kammerer S RR, Hoyal CR, Reneland R, Marnellos G, Kiechle M et al. Association of the NuMA region on chromosome 11q13 with breast cancer susceptibility. Proc Natl Acad Sci U S A. 2005;102:2004-9.

[88] Soto ME BN, Chineya ACP, Rizo H, Telich-Tarriba JE, Juarez-Orozco LE, Melendez G et al. Predictive value of antinuclear antibodies in autoimmune diseases classified by clinical criteria: Analytical study in a specialized health institute, 1 year follow-up. Results Immunology. 2013.

[89] Joseph B HP, Aziz H, Snyder K, Wynne J, Kulvatunyou N et al. . Continuous noninvasive hemoglobin monitor from pulse ox: ready for prime time? World J Surg. 2013;37:525-9.

[90] Laudisio A BS, Gemma A, Ferrucci L, Antonelli Incalzi R. Metabolic syndrome and hemoglobin levels in elderly adults: the Invecchiare in Chianti Study. J Am Geriatr Soc. 2013;61:963-8.

[91] Takahashi Y NT, Takeno S, Kimura Y, Okubo M, Kawahara K. Reduced expression of p63 has prognostic implications for patients with esophageal squamous cell carcinoma. Oncol Reports. 2006;15:323-8.

[92] Kim TW LH, Kang YK, Choe MS, Ryu MH, Chang HM. Prognostic significance of c-kit mutation in localized gastrointestinal stromal tumors. CLin Cancer Res. 2004;10:3076-81.

[93] Meigs JB WK, Sullivan LM, Hunt KJ, Haffner SM, Stern MP. Using metabolic syndrome traits for efficient detection of impaired glucose tolerance. Diabetes Care. 2004;27:1417-26.

[94] Seliger VI RD, Van Goor F, Schmelz A, Mueller P. The predictive potential of the sweat chloride test in cystic fibrosis patients with the G551D mutation. J Cyst Fibros. 2013;12:706-13.

[95] McWilliams RR PG, Rabe KG, Holtegaard LM, Lynch PJ, Bishop MD, Highsmith EW Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutations and risk for pancreatic adenocarcinoma. Cancer. 2010.

[96] Auewarakul CU LD, Promsuwicha O, ;Munkhetvit C. C kit receptor tyrosine kinase (CD117) expression and its positive predictive value for the diagnosis of Thai adult acute myeloid leukemia. Ann Hematol. 2006;85:108-12.

[97] Nakagawa T MS, Goto Y, Koyanagi K, Kitago M, Shingai T et al. . Detection of circulating tumor cells in early-stage breast cancer metastasis to axillary lymph nodes. Clin Cancer Res. 2007.

[98] Daga H MH, Yamamoto N, Shibata T, Endo M, Watanabe H et al. Small cell lung cancer and other thoracic malignancies: 1519PD Phase II study of Amrubicin in Patients with Refractory or Resistant Rleapsed Small-Cell Lung Cancer: Japan Clinical Oncology Group Study. Ann Oncol. 2012;23: ix492-ix8.

[99] Tang L ZS, Wei Liu W, Parchim NF, Huang J, Tang Y, Gan P, Zhong M. Diagnostic accuracy of circulating tumor cells detection in gastric cancer: systematic review and meta-analysis. BMC Cancer. 2013;13.

[100] Jin T, Peng, H, Wu, H. Clinical value of circulating liver cancer cells for the diagnosis of hepatocellular carcinoma: A meta-analysis. Biomed Rep. 2013;1:731-6.

[101] March S G-PJ, Masaguer A, Pizcueta P, Engel P, Bosch J. 549 P-selectin mediates leukocyte rolling in concanavalin-A induced hepatitis. J Hepatol. 2004;40.

[102] Msaouel P KM. Diagnostic value of circulating tumor cell detection in bladder and urothelial cancer: systematic review and meta-analysis. BMC Cancer. 2011.

[103] Walter RB AF, Estey EH, Bernstein ID. Acute myeloid leukemia stem cells and CD33-targeted immunotherapy. Blood. 2011.

[104] Kyrgidis A TT, Triaridis S. Melanoma: Stem cells, sun exposure and hallmarks for carcinogenesis, molecular concepts and future clinical implications. J Carcinog. 2010.

[105] Saad AA RS, Kanate A, Sebhai A, Merano G, Hobbs G. Correlation among [18-F] FDG-PET/CT, tumor marker CA 27.29, and circulating tumor cells in metastatic breast cancer. J Clin Oncol. 2007;25:suppl 10533.

[106] Murakami Y TH, Tanahashi T, Tanaka J, Kumada T, et al. Comprehensive miRNA expression analysis in peripheral blood Can diagnose liver disease. PloS One. 2012;7.

[107] Sozzi G CD, Leon M, Cirincione R, Roz L, Ratcliffe C, Roz E et al. Quantification of free cirulating DNA As a diagnostic marker in lung cancer. J Clin Oncol. 2003;21:3902-8.

[108] Douillard JY OG, Cobo M, Ciuleanu T, McCormack R, Webster A et al. . First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. Brit J Cancer. 2014;110:55-62.

[109] Medeiros LR RD, da Rosa MI, Bozzetti MC. Accuracy of CA125 in the diagnosis of ovarian tumors: A quantitative systematic review. Int J Gynecol Obstet. 2009;142:99-105.