samples increased the detection rate to 100%. Analysis of a heterogeneous sample set is therefore improved by personal evaluation in some circumstances, whereas analysis of good-quality DNA can be performed solely by the software.

In addition to polymorphism A896G, samples containing DNA from whole blood (n = 120) were examined for a second *TLR4* polymorphism, C1196T. Our aim was to look for linkage of these 2 polymorphisms. Genotyping revealed 90.83% homozygous variant C, 9.17% heterozygous, and no homozygous variants for T. The mean (SD) T_m for C was 58.18 (0.45) °C, and for T was 48.31 (0.41) °C [ΔT_m = 10.01 (0.15) °C]. With the exception of the 1 of 120 individuals being homozygous for 1196C and heterozygous for A896G, we confirmed the linkage (99.17%) of these 2 variants for mid-Europeans *(12)*.

Using a total of 497 heterogeneous DNA samples, we demonstrated that the LightTyper instrument is highly suitable for the analysis of 2 polymorphisms in the *TLR4* gene; in contrast to the similar setup using the LightCycler, we had no nonanalyzable samples in the subset containing DNA from whole blood (n = 120). In summary, we simplified and improved our previous detection method for variant A896G *(7)* by several modifications. (*a*) We replaced the anchor and sensor with a single probe, thereby reducing chances for mistakes. (*b*) We increased the sample size from 32 to 96 samples by use of the LightTyper instrument. (*c*) We improved the T_m separation by introducing an LNA base into the SimpleProbe oligomer, allowing for reliable computerized genotype detection. The assay is robust, and results were highly reproducible based on repeated analysis of 10% of the samples. Even for heterogeneous sample sets varying widely in quality and DNA amounts, personal evaluation of the results was needed only in some cases (4.4%). Although 2 separate analyses showed T_m fluctuations of ~1 °C, the system was very accurate, showing almost no variances within one run, and the resulting temperature differences were very stable. Day-to-day-variability and DNA quality were the most likely factors causing T_m variations.

The single-probe format is advantageous because the molecular neighborhood can be neglected. Hybridization is a complex process, however, and probe selection requires some attention. In this context, LNA bases are a helpful new tool to establish robust assays. We highly recommend this method for *TLR4* genotyping and for high-throughput analysis using computerized genotype detection in routine clinical settings.

References

1. Muzio M, Bosisio D, Polentarutti N, D'Amico G, Stoppacciaro A, Mancinelli R, et al. Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: selective expression of TLR3 in dendritic cells. J Immunol 2000;164:5998–6004.
2. Franchimont D, Vermeire S, El Housni H, Pierik M, Van Steen K, Gustot T, et al. Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299Gly polymorphism is associated with Crohn's disease and ulcerative colitis. Gut 2004;53:987–92.
3. Palsson-McDermott EM, O'Neill LA. Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4. Immunology 2004;113:153–62.
4. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, et al. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat Genet 2000;25:187–91.
5. Lorenz E, Frees KL, Schwartz DA. Determination of the TLR4 genotype using allele-specific PCR. Biotechniques 2001;31:22–4.
6. Schmitt C, Humeny A, Becker CM, Brune K, Pahl A. Polymorphisms of TLR4: rapid genotyping and reduced response to lipopolysaccharide of TLR4 mutant alleles. Clin Chem 2002;48:1661–7.
7. Heesen M, Wessiepe M, Kunz D, Vasickova K, Blomeke B. Rapid and reliable genotyping for the Toll-like receptor 4 A896G polymorphism using fluorescence-labeled hybridization probes in a real-time polymerase chain reaction assay. Clin Chim Acta 2003;333:47–9.
8. Arjomand-Nahad F, Diefenbach K, Landt O, Gaikovitch E, Roots I. Genotyping of the triallelic variant G2677T/A in MDR1 using LightCycler with locked-nucleic-acid-modified hybridization probes. Anal Biochem 2004;334:201–3.
9. Mouritzen P, Nielsen AT, Pfundheller HM, Choleva Y, Kongsbak L, Moller S. Single nucleotide polymorphism genotyping using locked nucleic acid (LNA). Expert Rev Mol Diagn 2003;3:27–38.
10. Hamann L, Hamprecht A, Gomma A, Schumann RR. Rapid and inexpensive real-time PCR for genotyping functional polymorphisms within the Toll-like receptor-2, -4, and -9 genes. J Immunol Methods 2004;285:281–91.
11. Okayama N, Fujimura K, Suehiro Y, Hamanaka Y, Fujiwara M, Matsubara T, et al. Simple genotype analysis of the Asp299Gly polymorphism of the Toll-like receptor-4 gene that is associated with lipopolysaccharide hyporesponsiveness. J Clin Lab Anal 2002;16:56–8.
12. Michel O, LeVan TD, Stern D, Dentener M, Thorn J, Gnat D, et al. Systemic responsiveness to lipopolysaccharide and polymorphisms in the toll-like receptor 4 gene in human beings. J Allergy Clin Immunol 2003;112:923–9.

DOI: 10.1373/clinchem.2005.049197

Analytical and Preanalytical Biases in Serum Proteomic Pattern Analysis for Breast Cancer Diagnosis, *Aly Karsan,[1*] Bernhard J. Eigl,[2] Stephane Flibotte,[3] Karen Gelmon,[2] Philip Switzer,[4] Patricia Hassell,[5] Dorothy Harrison,[5] Jennifer Law,[1] Malcolm Hayes,[1] Moira Stillwell,[4] Zhen Xiao,[6] Thomas P. Conrads,[6] and Timothy Veenstra[6]* (Departments of [1] Pathology and Laboratory Medicine and Medical Biophysics, [2] Medical Oncology, and [5] Radiology, and [3] Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada; [4] Department of Radiology, University of British Columbia, Vancouver, BC, Canada; [6] Laboratory of Proteomics and Analytical Technologies, SAIC, Inc., Frederick, MD; * address correspondence to this author at: Department of Medical Biophysics, British Columbia Cancer Research Centre, 675 West 10th Ave., Vancouver, BC, Canada V5Z 1L3; fax 604-675-8049, e-mail akarsan@bccrc.ca)

Currently available serum tumor markers lack sufficient specificity and sensitivity as stand-alone diagnostic or screening tests *(1)*. Nevertheless, these assays are used extensively because of a lack of better alternatives. To accelerate the discovery of tumor markers for diagnosis and/or prognosis, there has been great enthusiasm in attempting to use mass spectrometry (MS)-based testing of serum to identify potential biomarkers or spectral patterns that can act as a fingerprints for specific diseases *(1)*. Analysis of serum by surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) MS was recently reported to be able to predict the existence of ovarian cancer without missing a single case *(2)*. In this method, capture of serum proteins on a biochip by use of surface chemistry is followed by MS analysis of all cap-

tured proteins, and data mining software is used to identify a pattern of spectral peaks that will predict the presence or absence of the particular disease state in question *(2)*. Although the seminal study using this technique has been criticized by statisticians, bioinformaticians, and clinical chemists because of the likelihood of systematic bias, there has been no empirical testing of the potential for systematic bias in SELDI-TOF serum analysis *(1, 3–6)*. This study was conducted to determine whether spectral patterns generated by SELDI-TOF MS could distinguish between patients with cancer and those with benign disease among women presenting with suspicious breast abnormalities on mammography or physical examination.

The study was approved by the Research Ethics Board at the University of British Columbia, and all participating patients provided informed consent. We prospectively recruited 136 consecutive consenting patients attending 3 different clinics from September 2002 to April 2004 for a core needle biopsy for histopathologic diagnosis of a suspicious breast lump. Of the 136 patients, 3 were lost to follow-up, and 1 sample was not frozen within 6 h. Of the remaining 132 patients, 63 came from clinic A, 64 from clinic B, and 5 from clinic C. All patients with a positive core biopsy for malignancy as well as a subset with a negative core biopsy had excisional biopsies. A total of 96 patients (72.7%) received a histopathologic diagnosis of breast cancer (ductal carcinoma in situ, n = 13; invasive ductal carcinoma, n = 78; lobular, tubulolobular, or mixed, n = 5).

Serum samples were collected before biopsy in 7-mL glass serum tubes with no additive (BD Vacutainer®), aliquoted, and frozen at −80 °C within 6 h of phlebotomy until used for SELDI-TOF analysis. A pilot study using chips with different surface chemistries led us to choose the immobilized metal affinity capture (IMAC3) chips with Cu(II) as the metal ion to use for this study because of good reproducibility of duplicate spectra and the generation of multiple features on the spectra for analysis (data not shown). IMAC3 chips from the same lot were used for all samples run to avoid lot-to-lot variability. Chips were charged with 100 mmol/L $CuSO_4$, fixed with 100 mmol/L sodium acetate (pH 4.0), and equilibrated with phosphate-buffered saline (PBS), pH 7.4. Serum samples (50 μL) were diluted 1:1 in 8 mol/L urea containing 10 mL/L CHAPS and after vortex-mixing were further diluted 1:5 in PBS. Diluted serum samples (100 μL) were then applied to IMAC3 chips in duplicate, on spots on different chips. After washes in PBS and water, chips were air-dried, and 1 μL of saturated sinnapinic acid in 500 mL/L acetonitrile–5 mL/L trifluoroacetic acid was applied to each spot. The 132 serum samples were prepared and spotted in duplicate on 3 consecutive days, and the spotted arrays were read on a PBS II ProteinChip reader (Ciphergen Systems) on 2 consecutive days (one-third on the first day and the remaining two-thirds on the following day).

Spectra were calibrated externally and analyzed by mapping the raw (nonfiltered, non–baseline-subtracted) TOF spectra to mass spectra consisting of 16 384 channels, with mass calibration given by $m/z = aC^2$, where C is the channel number and $a = 0.0001\ m/z$, and normalized to the same total area. An automated procedure to find and fit the peaks in the mass spectra has been developed by one of the authors and is freely available (sflibotte@bcgsc.ca). This procedure generates an average spectrum from all samples, which is divided into several sections by a heuristic approach to obtain the best possible fit. Each section is then fitted iteratively with the appropriate number of gaussian peaks superimposed on a locally quadratic background. The duplicate pairs of spectra from each specimen in the dataset were averaged and fitted, with each section from the average spectrum used as a template. Duplicate spectra from individual samples showed a high degree of reproducibility as demonstrated by a median Pearson correlation coefficient of 0.9704 for all pairs of spectra evaluated. The Pearson correlation coefficients were calculated in the mass region used in the fitting procedure (m/z 533 to 26 840). Examples of duplicate raw spectra are shown in Fig. 1 of the Data Supplement that accompanies the online version of this Technical Brief at http://www.clinchem.org/content/vol51/issue8/. The position, width, and height of the peaks, and the local background were all fitted at the same time; this procedure corrects for local gain, matching variations between spectra because the absolute position of each peak is free to vary, although the relative position of each peak is fixed. A total of 445 peaks were fitted for each spectrum.

Two machine-learning algorithms, a support vector machine (SVM) and C4.5, were used in various analyses using all 445 peaks as described below. SVMs perform well in situations in which the number of samples in the dataset is not large compared with the number of attributes, i.e., peak areas in this case, and have been used successfully in microarray and SELDI-TOF experiments *(7, 8)*. The C4.5 algorithm, a decision tree algorithm, is also a widely used machine learning algorithm applied in many settings *(8)*. A 10-fold cross-validation was performed 10 times to assess each classification scheme. In other words, the dataset was divided into 10 equal groups, 9 of which were used to build a classifier and predict the classes of the samples in the remaining group. All 10 groups were assessed in this way. This 10-fold cross-validation procedure was repeated 9 more times with a breakdown of the dataset into 10 different but random groups. A majority predictor, which simply predicts the majority class in the dataset, was used as a comparator. Classification accuracy with means, SDs, and probability values to assess significant differences from the majority predictor (using a Student t-test) were generated by the Weka machine learning software *(8)*.

Our findings demonstrate that specimen collection and processing introduce significant biases in the spectral pattern, such that machine learning algorithms can differentiate between sample source, day that the chips were set up, and days that they were read. In contrast, accuracy of predicting cancer was much poorer.

Table 1. Accuracy of 2 machine learning algorithms in diagnosis and prediction of variables associated with the proteomic analysis of serum from patients with breast cancer.

		SVM		C4.5	
Variables	**Majority predictor,[a] %**	**Accuracy,[a] %**	***P***	**Accuracy,[a] %**	***P***
Cancer (A)[b]	72.8 (3.5)	70.9 (12.0)	0.67	66.7 (11.1)	0.12
Cancer (A, P)	72.8 (3.5)	72.8 (3.5)	1.00	70.8 (7.7)	0.46
Cancer (S)	70.0 (4.3)	78.4 (13.4)	0.08	78.1 (15.3)	0.13
Cancer (S, B)	56.7 (8.2)	67.0 (21.8)	0.19	62.9 (21.6)	0.42
Day chips read	67.2 (2.3)	97.0 (5.0)	10^{-28}	88.3 (9.7)	10^{-8}
Day chips set up	35.9 (3.1)	92.5 (7.1)	10^{-40}	80.6 (12.2)	10^{-17}
Source of serum	47.5 (1.8)	84.1 (9.2)	10^{-18}	85.3 (9.2)	10^{-20}

[a] Mean (SD).

[b] A, all patients in the study (n = 132); P, peaks from literature used for classification (see text); S, subset of patients with reproducible spectra (n = 70); B, patients from clinic B only (n = 36).

As demonstrated in Table 1, neither machine learning algorithm was able to classify patients with breast cancer any better than the majority predictor. Two previous studies using IMAC3 chips identified 2 different sets of peaks that were able to classify patients with breast cancer *(9, 10)*. Attempts to classify the spectra by use of these published peaks were also unsuccessful. We then eliminated every sample in which the duplicate spectra could not be overlaid by visual inspection. Both algorithms performed slightly better than the majority predictor in classifying cancer in this reduced subset of 70 patients, but the results were not statistically significant. To reduce possible source-related biases, we next analyzed specimens that showed reproducible spectra but came from only one clinic. However, the use of samples from only one clinic did not improve classification accuracy by either machine learning algorithm (Table 1).

In contrast to the lack of predictive ability of the spectral patterns for the diagnosis of breast cancer, both machine learning algorithms demonstrated an excellent ability to predict on which day the chips were read and on which day they were prepared, albeit the second variable may be a function of the first (Table 1). Even more surprisingly, there were distinct spectral features that the algorithms successfully applied to classifying the clinics from which the samples were acquired (Table 1). Fig. 1 shows a dot plot demonstrating that, using only 2 peaks at *m/z* 2992 and 5643, the C4.5 algorithm was able to distinguish between samples obtained in clinic A or B.

The very high probability values assigned to the classifications of distinct analytical and preanalytical variables suggest that in previous reports there may have been inadvertent biases in sample collection, storage, or processing between patients from different groups being tested. There are several potential reasons for the analytical and preanalytical biases seen, but the specific mechanisms remain to be elucidated. The findings presented here empirically validate the concerns of clinical chemists and bioinformaticians that serum profiling of unfractionated serum may be detecting preanalytical and analytical variables that are not reflective of the disease state *(3, 4, 11)*. The inability to use previously published peaks to classify breast cancer patients in this study also suggests that there are likely site-specific findings that reflect analytical and preanalytical biases for SELDI-TOF MS. As has been pointed out, it can be difficult to obtain stable, reproducible SELDI-TOF MS results over time and across laboratories *(12)*. Although a recent study has shown some reproducibility across laboratories, only 28 optimal spectra out of a cohort of more than 1000 were used for the validation, and whether the reproducible peaks actually represent cancer biomarkers or artifacts was not addressed *(13, 14)*. Recent critiques have argued that the two main potential problems with observational studies arise from chance and bias *(15–17)*. The current study highlights the effects of bias; thus, future studies attempting to profile serum by proteomic approaches will have to

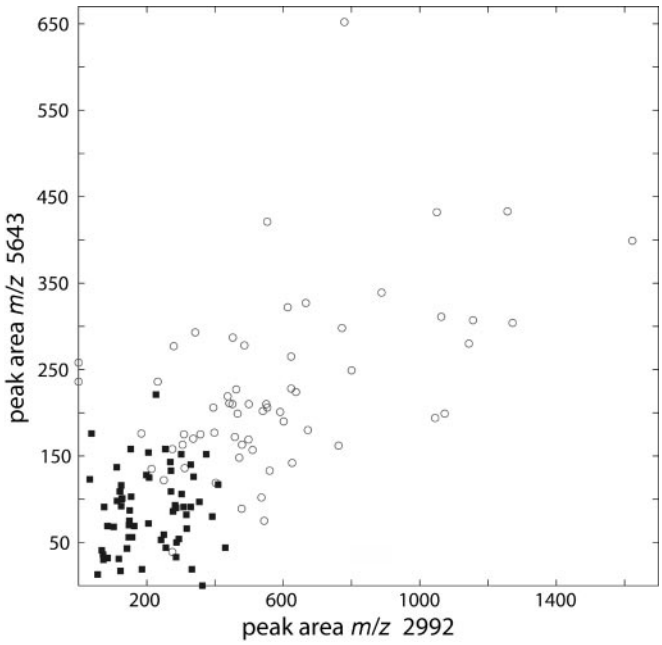

Fig. 1. Ability of the C4.5 algorithm to identify the source of serum samples based on the area under the peaks at *m/z* 5643 and 2992. ○, clinic A; ■, clinic B.

take extreme care in specimen handling and storage, as well as in randomization of specimen preparation and spectrum collection times, to discover true disease-related spectral profiles.

We thank Ingrid Pollet and Fred Wong for assistance with serum sample freezing. This study was funded by grants to A.K. from the National Cancer Institute of Canada with funds from the Canadian Cancer Society and to A.K. and K.G. from the Canadian Breast Cancer Foundation (BC Chapter), and in part with US funds from the National Cancer Institute, National Institutes of Health, under Contract NO1-CO-12400 to T.V., T.P.C., and Z.X. A.K. is supported by a personnel award from the Heart and Stroke Foundation of Canada and a Scholarship from the Michael Smith Foundation for Health Research.

References

1. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics 2004;3:367–78.
2. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.
3. Sorace JM, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. BMC Bioinformatics 2003;4:24.
4. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004;20:777–85.
5. Diamandis EP. Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? Clin Chem 2003;49:1272–5.
6. Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst 2004;96:353–6.
7. Zhang Z, Bast RC Jr, Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64:5882–90.
8. Witten IH, Frank E. Data mining: practical machine learning tools and techniques with JAVA implementations. San Francisco: Morgan Kaufman, 2000:371pp.
9. Pusztai L, Gregory BW, Baggerly KA, Peng B, Koomen J, Kuerer HM, et al. Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. Cancer 2004;100:1814–22.
10. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem 2002;48:1296–304.
11. Diamandis EP, van der Merwe DE. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. Clin Cancer Res 2005;11:963–5.
12. Coombes KR, Morris JS, Hu J, Edmonson SR, Baggerly KA. Serum proteomics profiling-a young technology begins to mature. Nat Biotechnol 2005;23:291–2.
13. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin Chem 2005;51:102–12.
14. Hortin GL. Can mass spectrometric protein profiling meet desired standards of clinical laboratory practice? Clin Chem 2005;51:3–5.
15. Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. J Natl Cancer Inst 2005;97:315–9.
16. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst 2005;97:307–9.
17. Liotta LA, Lowenthal M, Mehta A, Conrads TP, Veenstra TD, Fishman DA, et al. Importance of communication between producers and consumers of publicly available experimental data. J Natl Cancer Inst 2005;97:310–4.

Previously published online at DOI: 10.1373/clinchem.2005.050708

Comparison of HPLC and Capillary Electrophoresis for Confirmatory Testing of the Alcohol Misuse Marker Carbohydrate-Deficient Transferrin, *Anders Helander,[1*] Jos P.M. Wielders,[2] Riekie te Stroet,[2] and Jonas P. Bergström[1]* ([1] Department of Clinical Neuroscience, Karolinska Institutet & University Hospital, Stockholm, Sweden; [2] Department of Clinical Chemistry, Meander Medical Center, Amersfoort, The Netherlands; * address correspondence to this author at: Alcohol Laboratory, L7:03, Karolinska University Hospital Solna, SE-171 76 Stockholm, Sweden; fax 46-8-51771532, e-mail Anders.Helander@cns.ki.se)

The major form of the iron-transport glycoprotein transferrin in blood contains 2 N-linked disialylated biantennary oligosaccharide chains (glycans) and is named tetrasialotransferrin. Regular high alcohol consumption (mean of at least 50–80 g/day) generally alters the glycosylation profile of transferrin *(1)*, increasing the relative amounts of glycoforms lacking one (disialotransferrin) or both (asialotransferrin) N-glycans *(2, 3)*. The alcohol-related glycoforms are collectively referred to as carbohydrate-deficient transferrin (CDT). CDT measurements are widely used for identifying individuals with alcohol problems in various medical settings (e.g., addiction treatment) and for monitoring abstinence from alcohol in outpatient treatment programs (e.g., when drunk-driving offenders reapply for a driver's license) *(4)*. When drinking is discontinued, the CDT concentration normalizes with a half-life of 1.5–2 weeks *(5, 6)*. The main advantage of CDT over the conventional alcohol biomarkers, such as the liver function test γ-glutamyltransferase, is the higher specificity for alcohol misuse with resulting lower risk for false-positive identifications *(7, 8)*.

Since the discovery of CDT as an alcohol marker *(1)*, a multitude of analytical techniques and methods have been applied for its measurement *(1, 9)*. The most widely used assays worldwide today are the Axis-Shield %CDT immunoassay and various automated applications thereof, such as %CDT TIA from Bio-Rad and Tina-quant® %CDT from Roche *(10, 11)*. These assays are based on ion-exchange minicolumn chromatographic isolation of the CDT fraction, separate measurement of CDT and total transferrin using the same transferrin antibody, and calculating CDT as a percentage of total transferrin (%CDT). Immunologic methods are convenient and time-efficient for routine use in central laboratories with high specimen throughput, but because these tests separate CDT from non-CDT moieties on the basis of differences in isoelectric point (pI), they will be disturbed by genetic transferrin polymorphisms *(12)* and by congenital disorders of glycosylation *(13)*, which can cause falsely high or low results that may lead to false-positive or -negative identification of patients for alcohol misuse. Accordingly, when the %CDT ion-exchange immunoassay combination is used in medico-legal cases, such as traffic medicine *(14, 15)*, there is a need to confirm the test result by an independent separation method to rule out analytical interferences as the cause of a high value *(12, 16, 17)*.

This study compared the performances of 2 laboratory