

Building Classification Models with Combined Biomarker Tests: Application to Early Detection of Liver Cancer

Dion Chen

Biostatistics, Janssen R&D, LLC,
Spring House, PA 19477, USA

Surbhi Jain

Biomarkers, JBS Science, Inc.,
Doylestown, PA 18902, USA

Ying-Hsu Su

Biomarkers, The Baruch S. Blumberg Institute,
Doylestown, PA 18902, USA

Wei Song

Biomarkers, JBS Science, Inc.,
Doylestown, PA 18902, USA

Early detection of hepatocellular carcinoma (HCC) is critical for the effective treatment. Alpha fetoprotein (AFP) serum level is currently used for HCC screening, but the cutoff of the AFP test has limited sensitivity (~50%), indicating a high false negative rate. We have successfully demonstrated that cancer derived DNA biomarkers can be detected in urine of patients with cancer and can be used for the early detection of cancer (Jain et al., 2015; Lin et al., 2011; Song et al., 2012; Su, Lin, Song, & Jain, 2014; Su, Wang, Norton, Brenner, & Block, 2008). By combining urine biomarkers (uBMK) values and serum AFP (sAFP) level, a new classification model has been proposed for more efficient HCC screening. Several criteria have been discussed to optimal the cutoff for uBMK score and sAFP score. A joint distribution of sAFP and uBMK with point mass has been fitted using maximum likelihood method. Numerical results show that the sAFP data and uBMK data are very well described by proposed model. A tree-structured sequential test can be optimized by selecting the cutoffs. Bootstrap simulations also show the robust classification results with the optimal cutoff.

Keywords: Classification Model, Biomarker Data Analysis, Joint Distribution, Sensitivity, Specificity, Liver Cancer.

Introduction

Hepatocellular carcinoma (HCC) is the world's 2nd leading cause of cancer death, 83% of patients die within 5 years and is now one of the fastest growing cancers in the United States (Block, Mehta, Fimmel, & Jordan, 2003; Davis et al., 2008; Di Bisceglie et al., 2003; El Serag, 2001; Ferlay J, 2013; Howlader N, 2016; Marrero, 2006). Although the 5-year relative survival rate for patients with liver cancer is low, among patients in whom cancer is found at an early stage, the 5-year survival rate is 31% compared with only 3% in patients in whom the cancer is found after it has spread to distant organs (American Cancer Society, 2017). Early detection is critical for the effective treatment of HCC (A. Lok & McMahon, 2001; Marrero & Pelletier, 2006). The current standard marker, serum alpha fetoprotein (sAFP) is of limited value with only 40-60% sensitivity for all stages of HCC (Bruix & Sherman, 2011; Wright, Kreikemeier, & Fimmel, 2007). Unfortunately, sAFP consistently has lower sensitivity (25-50%) and specificity (75-90%) in detecting early HCC (stage I/II) as

compared to advanced HCC (A. S. Lok et al., 2010; Marrero et al., 2009; Marrero et al., 2005; Singal et al., 2012). It is nearly impossible to detect HCC early using current methods, even when a well-defined, high-risk population infected with the hepatitis C virus (HCV) or hepatitis B virus (HBV), which includes 90% of all HCC cases, can be identified (Bruix & Sherman, 2011; Smith, Cokkinides, Brooks, Saslow, & Brawley, 2010). The need for a more sensitive screening test to detect HCC early to improve the prognosis is urgent.

We have successfully demonstrated that cancer derived DNA biomarkers can be detected in urine of patients with cancer and can be used for the early detection of cancer (Hann et al., 2017; Jain et al., 2015; Lin et al., 2011; Song et al., 2012; Su et al., 2014; Su et al., 2008). The purpose of this article is to propose a process to improve early detection of HCC by combining sAFP and urine biomarkers (uBMK). The method proposed in this article can easily be extended to using multiple test processes.

Data and Method

The sAFP and uBMK data used in our study were collected from 203 non-HCC patients either cirrhosis (106) or hepatitis (97) and 84 HCC patients. The serum AFP values and urine samples were obtained from three different sources under institutional review board approvals and with written informed consent of the patients from the National Cheng-Kung University Medical Center, Taiwan; Buddhist Tzu Chi Medical Center in Hualien, Taiwan; and non-identifiable archived urine samples collected from patients with hepatitis, provided from Trovagene, Inc (San Diego, CA, USA).

Urine DNA Isolation and Bisulfite Treatment

Freshly collected urine was immediately mixed with 0.5 mol/L EDTA, pH 8.0, to a final concentration of 10 mmol/L EDTA and stored at -70°C . Total urine DNA was isolated by adding an equal volume of 6 mol/L guanidine thiocyanate (Sigma, St. Louis, MO) to thawed urine as described previously. The LMW urine DNA, DNA less than 1 kb, was obtained from total urine DNA using carboxylated magnetic beads (Agentcourt Bioscience Corporation, Beverly, MA), as previously developed by us. Bisulfite (BS) treatment was performed using EZ DNA Methylation-Lightning™ Kit (Zymo Research, Irvine, CA) following manufacturer's guidelines.

HCC urine Biomarker Quantification

TP53249T mutation detection assay was performed in duplicate using the low molecular weight urine DNA from 200-500 μl of urine per reaction as per the manufacturer's specifications (JBS Science, Inc). Both methylation (*mRASSF1A* and *mGSTP1*) detection assays were performed in duplicate using the bisulfite treated low molecular weight urine DNA from 200-500 μl of urine per reaction as per the manufacturer's specifications (JBS Science, Inc). All biomarker values below the lower limit of detection (LLOD) have been replaced by half of LLOD value for analysis purpose.

Logistic model with sAFP has been established for HCC screening by using cutoff 0.224 ng/mL to achieve 63.1% sensitivity and keep 90.1% specificity, which is comparable to the results reported in literature (Singal et al. 2012; Lok et al. 2012; Marrero et al, 2009; Marrero et al. 2005).

Three urine biomarkers (*TP53249T*, *mRASSF1A*, and *mGSTP1*) have been shown in previous studies to be associated with HCC in high incidence (Csepregi et al., 2008; Feng et al., 2010; Harder et al., 2008; Jain et al., 2012; Jain et al., 2015; Katoh et al., 2006; Lee et al., 2003; Lin et al., 2011; Moribe et al., 2009; Su et al., 2014; Yang, Guo, Herman, & Clark, 2003). Urine biomarker HCC screening model has been built using logistic

model with the linear combination of 3 urine biomarkers. The cutoff based on the uBMK linear score is 0.2 to achieve 84.5% sensitivity and 90.1% specificity, where uBMK linear score is calculated as $0.1118 \times TP53249T + 0.0193 \times mRASSF1A + 0.0015 \times mGSTP1$.

A double-log-transformation is applied to sAFP and uBMK linear score as below for normalizations:

- sAFP score = $\log[\log(\text{sAFP level}) + 0.5]$;
- uBMK score = $\log[\log(\text{uBMK liner score}) + 0.5]$,

where constant 0.5 is added to ensure positive values for log-transformation.

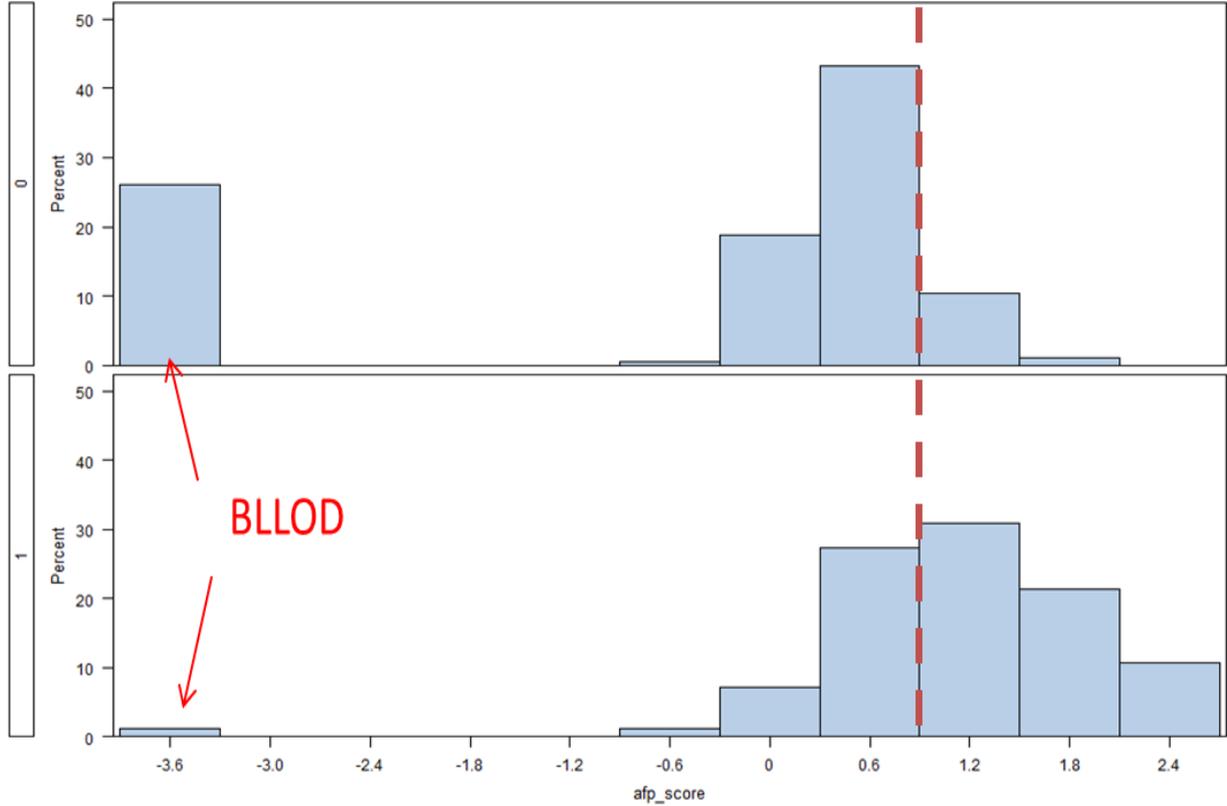


Figure 1. Distribution of sAFP score.

The upper panels of Figures 1 and 2 are non-HCC population, and lower panels are HCC population. A scatter plot of joint distribution of sAFP and uBMK is displayed in Figure 3.

Note there is a non-trivial point mass density at LLOD (-3.81 for sAFP score and -2.41 for uBMK score). Let $S = \text{sAFP} + 3.81$ and $U = \text{uBMK} + 2.41$. The joint distribution of sAFP score S and uBMK score U is formulated as below:

$$f_D(s, u) = \begin{cases} p_{D12}, & \text{if } s = 0 \text{ and } u = 0; \\ p_{D1}g_D(u), & \text{if } s = 0 \text{ and } u > 0; \\ p_{D2}h_D(s), & \text{if } s > 0 \text{ and } u = 0; \\ (1 - p_{D12} - p_{D1} - p_{D2})k_D(s, u), & \text{if } s > 0 \text{ and } u > 0. \end{cases} \quad (1)$$

Where D denote the disease population as $D=1$ and 0 for HCC and non-HCC, respectively. The marginal distributions $h_D(s)$ and $g_D(u)$ for S and U , respectively, are log-normal distributions, and the joint distribution $k_D(s, u)$ of S and U is bivariate log-normal.

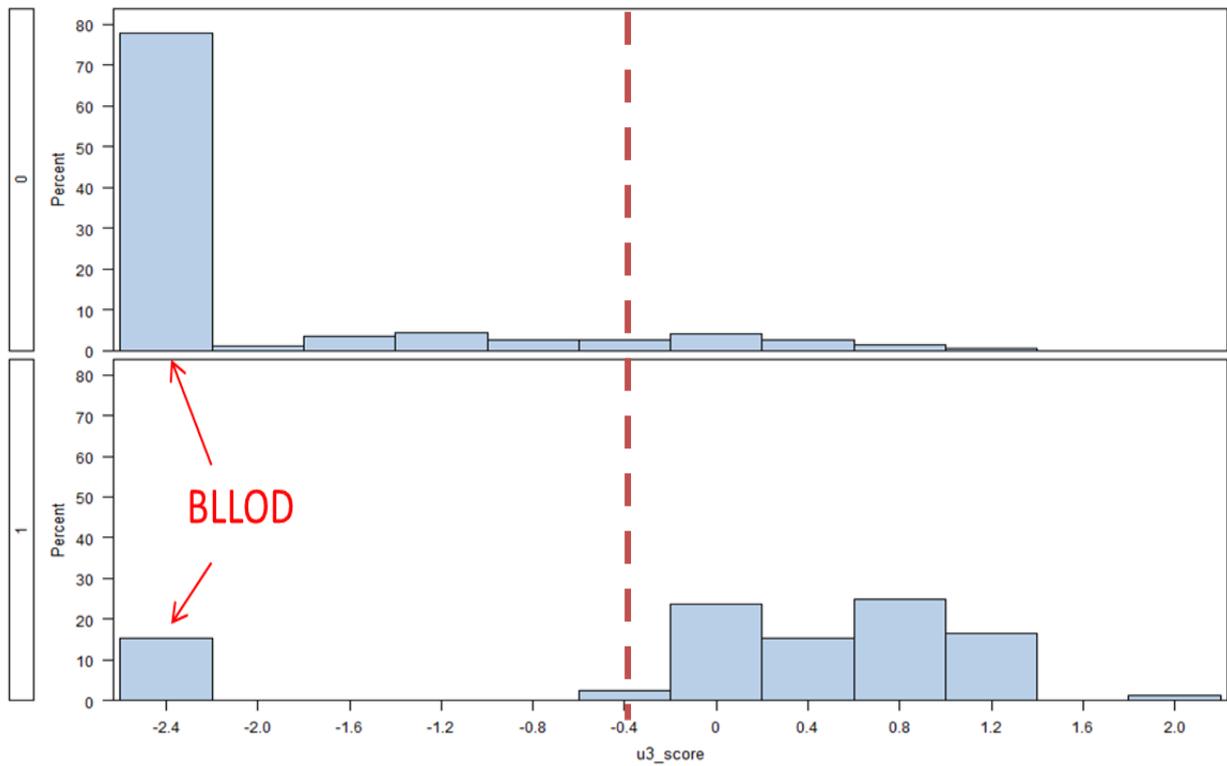


Figure 2. Distribution of uBMK score.

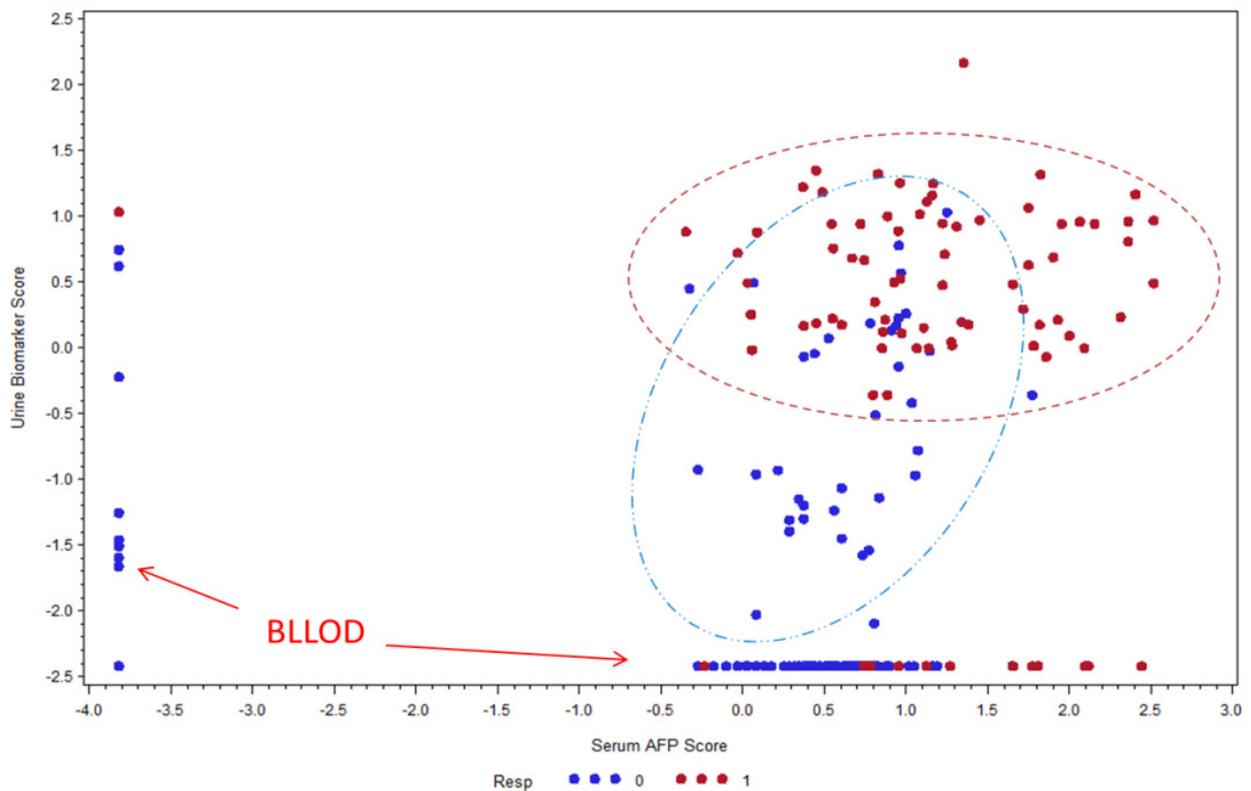


Figure 3. Scatter plot of joint distribution of sAFP score and uBMK score.

The maximum likelihood estimates of the unknown parameters have been obtained with the sample data as below.

$$f_0(s,u) = \begin{cases} 0.22, & \text{if } s = 0 \text{ and } u = 0; \\ 0.04\varphi_1([\ln(u) + 0.791]/1.02), & \text{if } s = 0 \text{ and } u > 0; \\ 0.56\varphi_1([\ln(s) - 0.48]/0.337), & \text{if } s > 0 \text{ and } u = 0; \\ 0.18\varphi_2\left[\begin{pmatrix} \ln(s) - 0.66 \\ \ln(u) + 0.55 \end{pmatrix}, \begin{pmatrix} 0.19 & 0.1 \\ 0.1 & 0.66 \end{pmatrix}\right], & \text{if } s > 0 \text{ and } u > 0, \end{cases} \quad (2)$$

$$f_1(s,u) = \begin{cases} 0.01, & \text{if } s = 0 \text{ and } u > 0; \\ 0.15\varphi_1([\ln(s) - 1.33]/0.74), & \text{if } s > 0 \text{ and } u = 0; \\ 0.84\varphi_2\left[\begin{pmatrix} \ln(s) - 1.18 \\ \ln(u) + 0.59 \end{pmatrix}, \begin{pmatrix} 0.47 & 0.01 \\ 0.01 & 0.25 \end{pmatrix}\right], & \text{if } s > 0 \text{ and } u > 0, \end{cases} \quad (3)$$

where $\varphi_1(x)$ and $\varphi_2(x,y)$ are the density functions of the standard univariate and bivariate normal distributions, respectively.

Evaluations of HCC Screening Tests

The HCC screening tests can be formulated as classification models. Let S and U denote the test results and c_s and c_u be the cutoffs for sAFP and uBMK scores, respectively. The positive result will classify as HCC by sAFP or uBMK, if $S \geq c_s$ or $U \geq c_u$, respectively, and non-disease otherwise.

The joint distribution of (S, U) is denoted as

$$F_D(s, u) = \int_{x < s} \int_{y < u} f_D(x, y) dx dy$$

where $f_D(s, u)$ is the joint density function given by Equation (2) and (3).

The precision of tests can be described as sensitivity (the probability of true positive classification) and specificity (the probability of true negative). For given cutoff, the area under curve of receiver operating characteristic (AUROC, Hanley & McNeil, 1982; Mason & Graham, 2002) of the test is simplified as (sensitivity + specificity)/2, the half of the sum of specificity and sensitivity, which is an overall assessment of the test accuracy.

There are several commonly used methods to combine two test results S and U together to improve the accuracy of screening result. The most often used method is to have a linear combination of two tests $Z = a_0 + a_1S + a_2U$, or additive model $Z = b_s(S, \beta_1) + b_u(U, \beta_2)$ with some function $b_s(\cdot)$ and $b_u(\cdot)$ and parameters β_1 and β_2 . Also, one can use “both positive” rule $Z = (S > c_s \text{ and } U > c_u)$ for higher specificity and “either positive” rule $Z = (S > c_s \text{ or } U > c_u)$ for higher sensitivity. However, all of methods mentioned above need to have both tests done, although the order of the tests is not the matter.

From clinical practice point of view, sequential tests are usually preformed. For example, a “standard” screening test is followed by a newly developed test; an easy test is followed by a complicate test; non-invasive test is followed by an invasive test, or preferred test is followed by a non-preferred test. Using sequential test process, both patients and clinicians can save time and money by avoiding non-necessary test. Also, two cutoff values could be adjusted for each test to increase the accuracy of screening tests for difference purposes. For example, higher sensitivity will be requested to detect disease indicators, and the overall accuracy is needed to establish disease absence or presence. Certainly, the order of tests is the matter in sequential test process. Based on a classification tree structure, two-step sequential tests will be discussed in the following sections

Two-Step Sequential Tests

A classification tree based two-step sequential test can be performed as either positive or negative results of Test 1 followed by Test 2. It is followed from some simple algebraic operations that, if positive (negative) results of Test 1 are followed by Test 2, respectively sensitivity (specificity) will never be increased and specificity (sensitivity) will never be decreased.

Assume that there are two HCC screening tests available, and consider the sequential test that Test 1 negative followed by Test 2, which is equivalent “either positive” rule, i.e., a patient classified as HCC only if either sAFP or uBMK positive to increase the sensitivity of the HCC screening test.

For given cutoff values c_s and c_u for sAFP and uBMK, respectively, it follows from Eq. 3 that, for the sequential test, there are

$$\text{Sensitivity} = P(\text{diagnosis as HCC} \mid D = 1)$$

$$= P(S > c_s \text{ or } U > c_u \mid D = 1) = 1 - F_I(c_s, c_u); \quad (4)$$

$$\text{Specificity} = P(S \leq c_s \text{ and } U \leq c_u \mid D = 0) = F_0(c_s, c_u); \quad (5)$$

$$\text{AUROC} = (\text{Sensitivity} + \text{Specificity})/2 = [F_0(c_s, c_u) + 1 - F_I(c_s, c_u)]/2; \quad (6)$$

The sensitivity, specificity and AUROC can be calculated with the density functions estimated from equations (2) and (3).

Optimal Criteria for HCC Screening Tests

There optimal criteria can be applied to achieve the optimization by selecting the cutoff values for sAFP and uBMK for HCC screening tests.

Global Optima

Maximum AUROC = (Specificity + Sensitivity)/2 to have the best overall accuracy. The optimization can be obtained by selecting cutoff c_s and c_u to maximum the objective function given by Equation (6).

$$L_g(c_s, c_u) = F_0(c_s, c_u) - F_I(c_s, c_u).$$

However, the global optimal may not achieve a sufficient level on both sensitivity and specificity.

Conditional Optima

When sAFP is used for HCC screening test alone, the cutoff 0.2 ng/mL achieve about 90% specificity and about 50% sensitivity. In sequential test, it is reasonable to maximize the sensitivity on the condition of specificity $s_p = 90\%$, which can be obtained with Lagrange-multiplier method. The objective function is

$$L_c(c_s, c_u, \lambda) = -F_I(c_s, c_u) + \lambda[F_0(c_s, c_u) - s_p].$$

Local Optima

Local optima can be obtained by fixing one cutoff of either c_s or c_u and searching for the other to maximize sensitivity, specificity or AUROC. Since the process of two-dimension optimal is reduced to one-dimension search, it is easily to be implemented in computation.

Optimal Results

With the sAFP and uBMK data collected, the optimal cutoffs of sAFP and uBMK are calculated with various criterions (Table 1).

Table 1
Optimal Cutoffs with Various Criteria

Criterion	sAFP Cutoff	uBMK Cutoff	Sensitivity	Specificity	AUROC
Max AUROC	1.25	-0.25	0.903	0.905	0.904
90% Specificity	1.20	-0.27	0.907	0.900	0.903
Marginal ^a	0.90	-0.40	0.847	0.867	0.857
uBMK+sAFP ^b	1.65	-0.40	0.875	0.908	0.891
sAFP+uBMK ^c	0.90	0.75	0.769	0.877	0.823

^a The cutoffs were based on the marginal distributions of sAFP and uBMK.

^b Local optimal searching cutoff for sAFP conditionally on the fixed uBMK cutoff ($c_u = -0.4$).

^c Local optimal searching cutoff for uBMK conditionally on the fixed sAFP cutoff ($c_s = 0.9$).

The global optimal criterion to maximize AUROC leads the similar result as condition optimal criterion that maximizes sensitivity on the condition of keeping at least 90% of specificity. See the first two rows of Table 1 for detail.

The result of local optima on AUROC by searching for uBMK cutoff conditional on fixed sAFP cutoff (0.9) is better than the result of searching for sAFP cutoff conditional on fixed uBMK cutoff (-0.4), and the latter doesn't achieve at least 90% specificity.

If the marginal cutoffs, which were obtained with sAFP and uBMK models separately without optimal process, $c_s = 0.9$ and $c_u = -0.4$ are used, the results is better than the HCC screening test with sAFP alone by the means of higher AUROC. However, it doesn't achieve at least 90% specificity.

Comparing the logistic models with sAFP, uBMK and both of sAFP and uBMK, the sequential model with maximum AUROC is the best not only in overall accuracy but also achieving at least 90% specificity (Table 2). Table 2 shows that the sequential HCC screening test can be improved by searching the cutoff values of sAFP and uBMK.

Table 2
Comparison of HCC Screening Tests

Model	Cutoff	Sensitivity	Specificity	AUROC
sAFP ^a	$c_s = 0.9$	0.631	0.901	0.766
uBMK ^b	$c_u = -0.4$	0.845	0.901	0.873
sAFP + uBMK ^c	$c = 0.16$	0.893	0.901	0.897
uBMK/sAFP ^d	$c_s = 1.25, c_u = -0.25$	0.903	0.905	0.904

^a Logistic model with sAFP, the cutoff was selected to achieve 90% specificity

^b Logistic model with uBMK, the cutoff was selected to achieve 90% specificity

^c Logistic model with both sAFP and uBMK, the cutoff was selected to achieve 90% specificity

^d Sequential model that negative results of uBMK is followed by sAFP, and the cutoffs were selected to achieve maximum AUROC.

Numerical Results of Bootstrap Simulations

To evaluate the robustness of the sequential model proposed, bootstrap simulations have been conducted. With N=1000 bootstrap samples, the mean, standard deviation (SD) and 90% confident interval (CI) of AUROC, specificity and sensitivity are presented in Tables 3-5, respectively.

The numerical results of the bootstrap simulation show that, for the sequential model propose, the coefficient of variation (CV%) of AUROC is about 15%, CV% of sensitivity is about 27%, and about CV% of

specificity is about 20%. Note that the lower bound of 90% CI is 0.89 for AUROC, 0.88 for both sensitivity and specificity. The numerical results of the bootstrap simulation indicate that the sequential model is relative robust.

Table 3

Bootstrap simulation results for AUROC (N=1000)

Criterion	sAFP cutoff	uBMK cutoff	Estimate	Mean (SD)	90% CI
Max AUROC	1.25	-0.25	0.904	0.909 (0.012)	(0.889, 0.930)
90% Specificity	1.20	-0.27	0.903	0.908 (0.012)	(0.888, 0.929)
Marginal*	0.90	-0.40	0.857	0.882 (0.012)	(0.861, 0.902)
uBMK+sAFP**	1.65	-0.40	0.891	0.895 (0.014)	(0.872, 0.918)
sAFP+uBMK***	0.90	0.75	0.823	0.828 (0.015)	(0.804, 0.852)

Table 4

Bootstrap simulation results for sensitivity (N=1000)

Criterion	sAFP cutoff	uBMK cutoff	Estimate	Mean (SD)	90% CI
Max AUROC	1.25	-0.25	0.903	0.913 (0.021)	(0.882, 0.949)
90% Specificity	1.20	-0.27	0.907	0.916 (0.020)	(0.886, 0.952)
Marginal *	0.90	-0.40	0.847	0.851 (0.017)	(0.828, 0.879)
uBMK+sAFP**	1.65	-0.40	0.875	0.882 (0.024)	(0.844, 0.925)
sAFP+uBMK***	0.90	0.75	0.769	0.780 (0.025)	(0.739, 0.822)

Table 5

Bootstrap simulation results for specificity (N=1000)

Criterion	sAFP cutoff	uBMK cutoff	Estimate	Mean (SD)	90% CI
Max AUROC	1.25	-0.25	0.905	0.905 (0.014)	(0.882, 0.927)
90% Specificity	1.20	-0.27	0.900	0.901 (0.014)	(0.877, 0.923)
Marginal *	0.90	-0.40	0.867	0.813 (0.019)	(0.782, 0.844)
uBMK+aAFP**	1.65	-0.40	0.908	0.908 (0.014)	(0.886, 0.930)
aAFP+uBMK***	0.90	0.75	0.877	0.877 (0.015)	(0.853, 0.903)

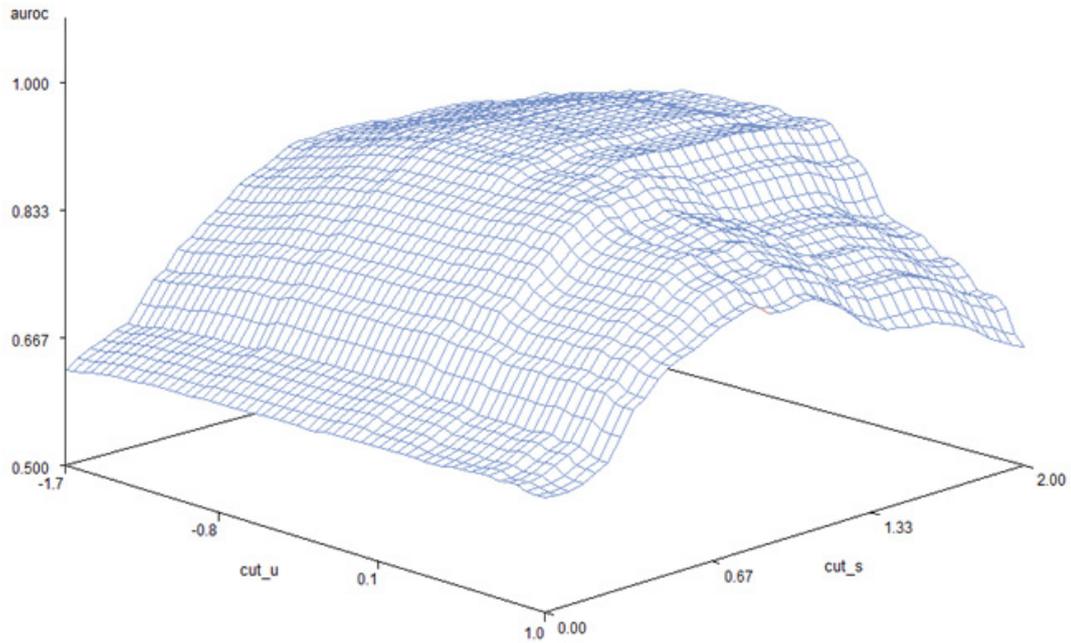
Conclusions and Discussions

Early detection of HCC is critical for the effective treatment. A classification tree-structured two-step sequential model that combines uBMK score and sAFP level has been proposed for more efficient HCC screening. Comparing with the logistic model using linear combination of sAFP and uBMK, the overall accuracy of the HCC screening has been improved in both sensitivity and specificity.

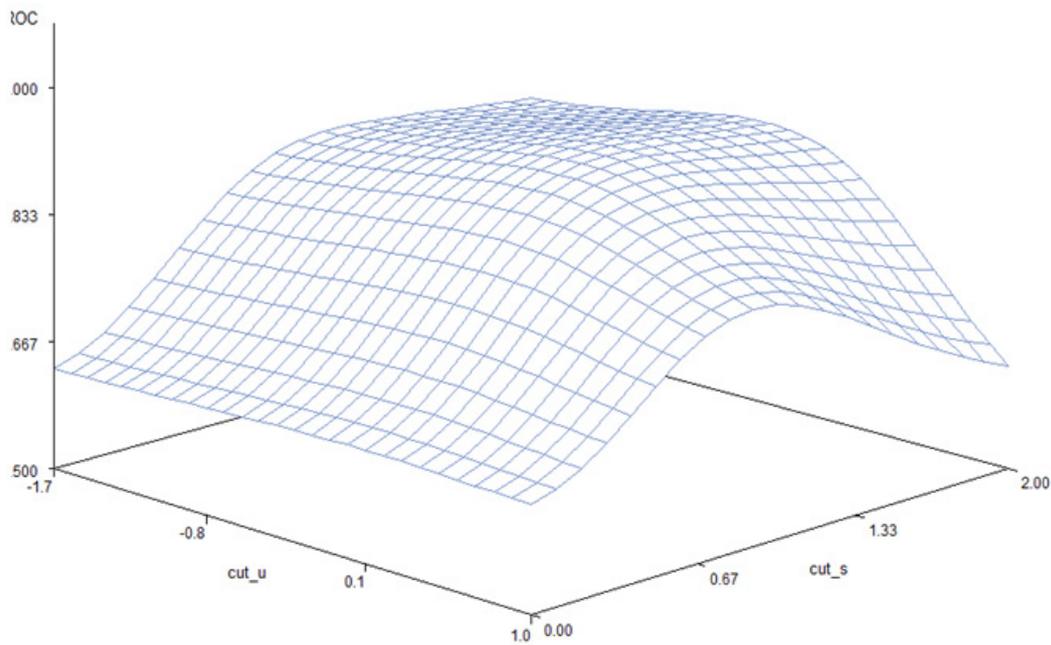
The sequential test can reduce the cost and time by avoiding non-necessary tests. The sequential test proposed is to conduct the second test only if the negative result was obtained from the first test, which is equivalent to the rule of “either positive”. Therefore, the order of the two tests will not impact on the screening results. Based on the consideration of “non-invasive” test followed by invasive test, it can be recommended to clinical practice to conduct urine test first, and then do blood test only if the urine test result is

negative.

Based on the data collected from uBMK and sAFP, a joint distribution with point mass probability has been derived to model the non-trivial portion of below LLOD values. Figures 4, 5 and 6 are the visualization of the objective functions by cutoff values for AUROC, sensitivity and specificity. The numerical results show that the statistical model very well describes the sAFP and uBMK data.

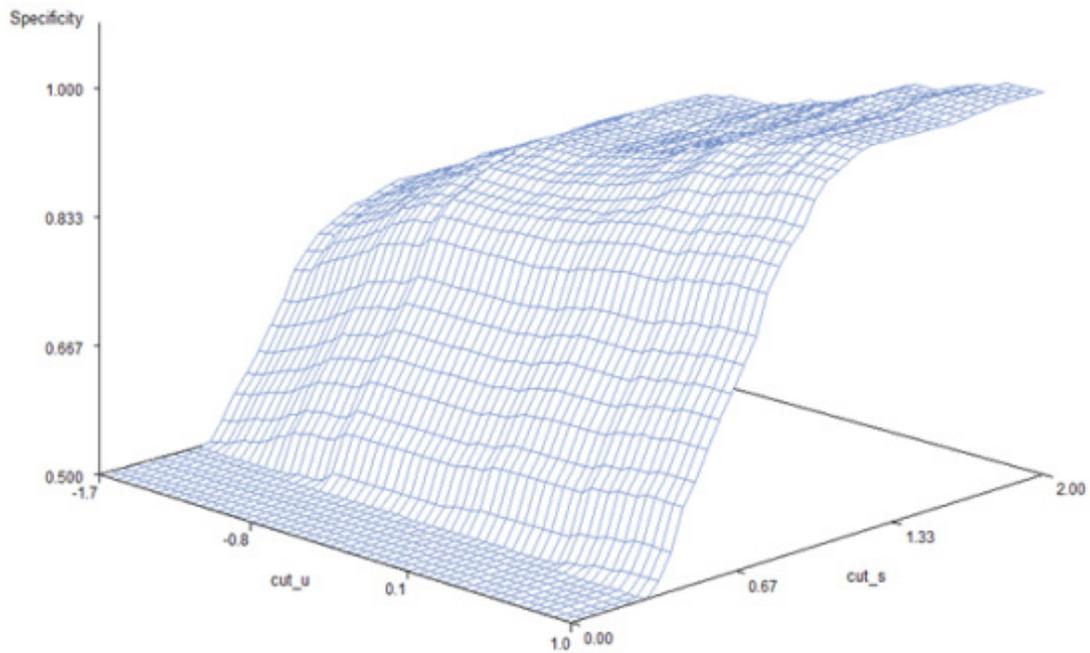


Raw Data

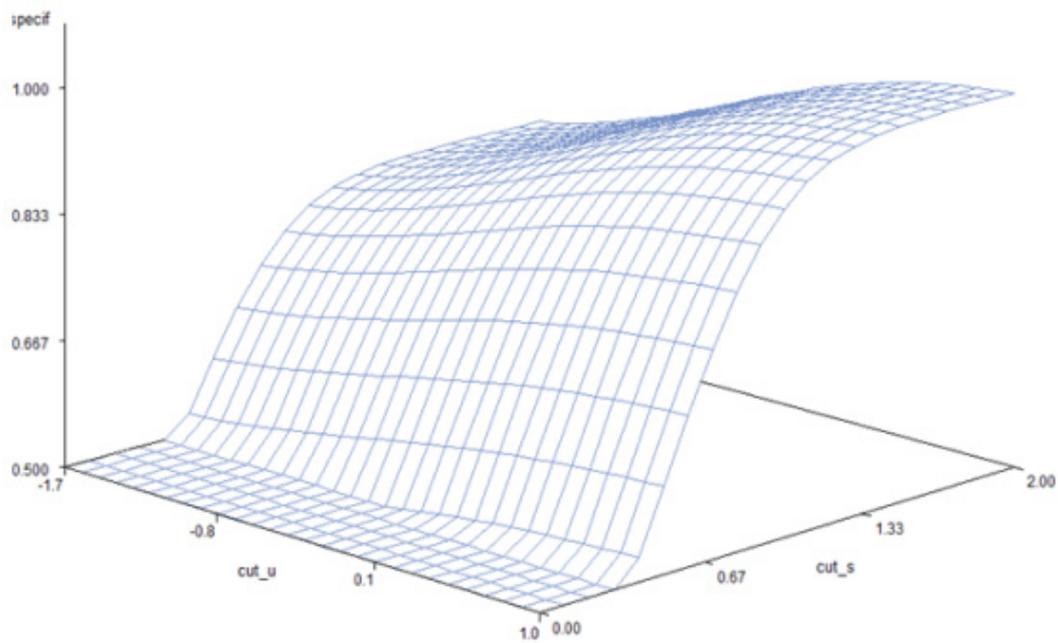


Model Data

Figure 4. AUROC by cutoff.

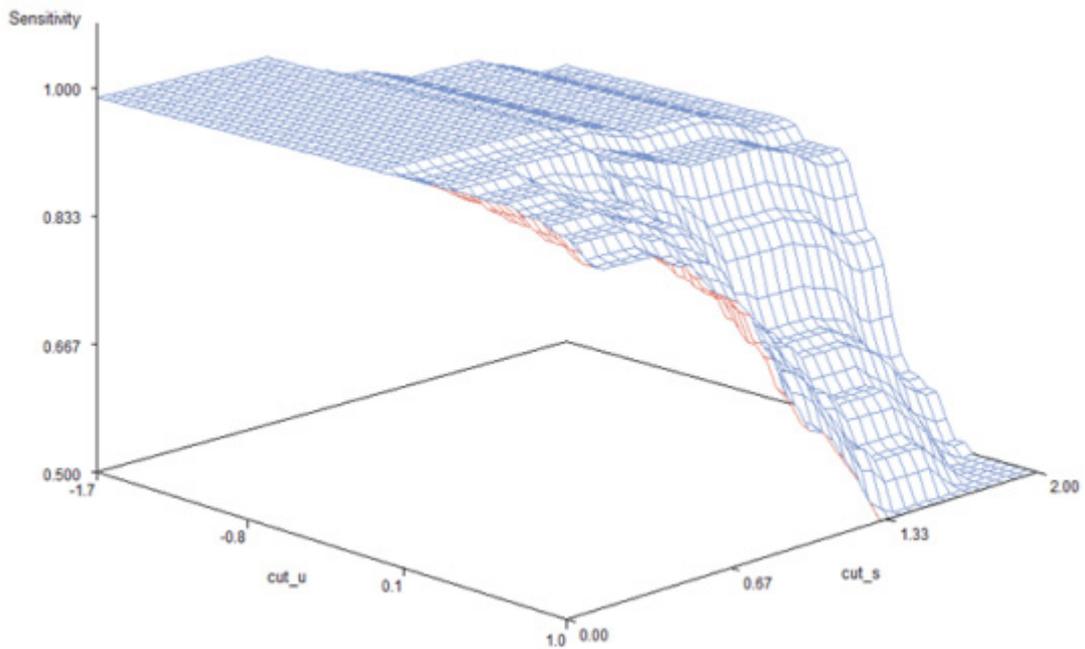


Raw Data

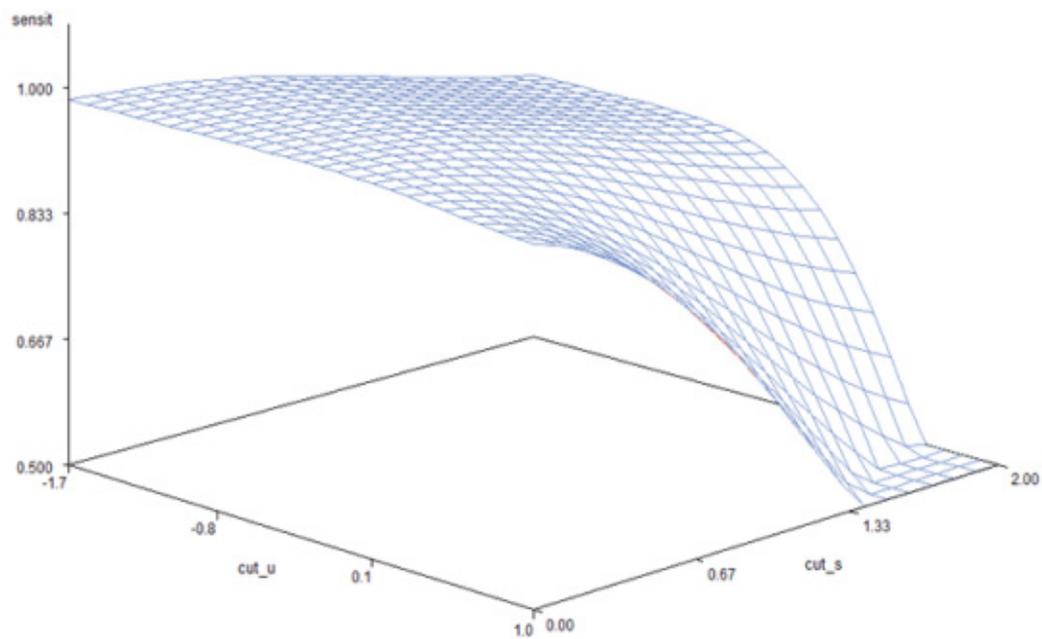


Model Data

Figure 5. Sensitivity by cutoff.



Raw Data



Model Data

Figure 6. Specificity by cutoff.

Several optimal criteria have been discussed, and the sequential test can be optimized by selecting the cutoffs. Bootstrap sample simulations also show the robust classification results with the optimal cutoff.

The data presented in this article should be considered as a “Training Set”. To apply the sequential test for HCC screening in clinical practice, the model needs to be validated by more double-blinded test data. Also, to

establish a standard HCC screening tool, demographics factor may be taken into the model, such as race, gender, and age.

From statistics point of view, the joint distribution with point mass probability can be used for other biomarker data analysis with non-trivial below LLOD values. The bivariate distribution can also be extended to multivariate distribution. However, statistical properties of the model parameter estimate need to be further discussed in general. For example, the asymptotical properties MLE may need to be investigated by checking the regularity conditions with the point mass probability.

Acknowledgement

The authors thank Trovagene Inc. (San Diego, CA) for the generosity of providing the archived, non-identifiable urine samples previously collected from patients with hepatitis for other studies.

Funding Information

This work is supported by R44 CA165312.

References

- American Cancer Society. (2017). Cancer Facts and Figures
- Block, T. M., Mehta, A. S., Fimmel, C. J., & Jordan, R. (2003). Molecular viral oncology of hepatocellular carcinoma. *Oncogene*, 22(33), 5093-5107.
- Bruix, Jordi, & Sherman, Morris. (2011). Management of hepatocellular carcinoma: An update. *Hepatology*, 53(3), 1020-1022. doi: 10.1002/hep.24199
- Csepregi, Antal, Rocken, C., Hoffmann, J., Gu, P., Saliger, S., Muller, O., Ebert, M. P. A. (2008). APC promoter methylation and protein expression in hepatocellular carcinoma. *J Cancer Res Clin Oncol*, 134(5), 579-589.
- Davis, G. L., Dempster, J., Meler, J. D., Orr, D. W., Walberg, M. W., Brown, B., Goldstein, R. M. (2008). Hepatocellular carcinoma: management of an increasingly common problem. *Proc (Bayl Univ Med Cent)*, 21(3), 266-280.
- Di Bisceglie, A. M., Lyra, A. C., Schwartz, M., Reddy, R. K., Martin, P., Gores, G., Befeler, A. S. (2003). Hepatitis C-related hepatocellular carcinoma in the United States: influence of ethnic status. *Am J Gastroenterol*, 98(9), 2060-2063.
- El Serag, H. B., Mason, A. C., and Key, C. (2001). Trends in survival of patients with hepatocellular carcinoma between 1977 and 1996 in the United States. *Hepatology*, 33, 62-65.
- Feng, Qinghua, Stern, Joshua E., Hawes, Stephen E., Lu, Hiep, Jiang, Mingjun, & Kiviat, Nancy B. (2010). DNA methylation changes in normal liver tissues and hepatocellular carcinoma with different viral infection. *Experimental and Molecular Pathology*, 88(2), 287-292.
- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: <http://globocan.iarc.fr>, accessed on 14/03/2017.
- Hanley, J A, & McNeil, B J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36. doi: 10.1148/radiology.143.1.7063747
- Hann H.W., Jain S., Park G., Steffen J.D., Song W., Su Y-H. (2017). Detection of urine DNA markers for monitoring recurrent hepatocellular carcinoma. *Hepatoma Res*, 3:105-11.
- Harder, Jan, Opitz, Oliver G., Brabender, Jan, Olschewski, Manfred, Blum, Hubert E., Nomoto, Shuji, & Usadel, Henning. (2008). Quantitative promoter methylation analysis of hepatocellular carcinoma, cirrhotic and normal liver. *International Journal of Cancer*, 122(12), 2800-2804.
- Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). (2016). SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2013/, based on November 2015 SEER data submission, posted to the SEER web site, April 2016.
- Jain, Surbhi, Chen, Sitong, Chang, Kung-Chao, Lin, Yih-Jyh, Hu, Chi-Tan, Boldbaatar, Batbold, Su, Ying-Hsiu. (2012). Impact

- of the Location of CpG Methylation within the GSTP1 Gene on Its Specificity as a DNA Marker for Hepatocellular Carcinoma. *PLoS ONE*, 7(4), e35789. doi: 10.1371/journal.pone.0035789
- Jain, Surbhi, Xie, Lijia, Boldbaatar, Batbold, Lin, Selena Y., Hamilton, James P., Meltzer, Stephen J., Su, Ying-Hsiu. (2015). Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis. *Hepatology Research*. doi: 10.1111/hepr.12449
- Katoh, Hiroto, Shibata, Tatsuhiro, Kokubu, Akiko, Ojima, Hidenori, Fukayama, Masashi, Kanai, Yae, & Hirohashi, Setsuo. (2006). Epigenetic Instability and Chromosomal Instability in Hepatocellular Carcinoma. *American Journal of Pathology*, 168(4), 1375-1384.
- Lee, Sun, Lee, Hyeon Joo, Kim, Jae Hoon, Lee, Hyo Suk, Jang, Ja June, & Kang, Gyeong Hoon. (2003). Aberrant CpG Island Hypermethylation Along Multistep Hepatocarcinogenesis. *American Journal of Pathology*, 163(4), 1371-1378.
- Lin, Selena Y, Dhillon, Veerpal, Jain, Surbhi, Chang, T-T, Hu, C-T, Lin, Y-J, Su, Y-H. (2011). A locked nucleic acid clamp-mediated PCR assay for detection of a p53 codon 249 hotspot mutation in urine. *Journal of Molecular Diagnostics*, 13(5), 474-484.
- Lok, Anna S., Sterling, Richard K., Everhart, James E., Wright, Elizabeth C., Hoefs, John C., Di Bisceglie, Adrian M., . . . Dienstag, Jules L. (2010). Des- γ -Carboxy Prothrombin and α -Fetoprotein as Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Gastroenterology*, 138(2), 493-502. doi: 10.1053/j.gastro.2009.10.031
- Lok, AS, & McMahon, BJ. (2001). Chronic hepatitis B. *Hepatology*, 34(6), 1225-1241.
- Marrero, J. A. (2006). Hepatocellular carcinoma. *Curr Opin Gastroenterol*, 22(3), 248-253.
- Marrero, J. A., & Pelletier, S. (2006). Hepatocellular carcinoma. *Clinics in Liver Disease*, 10(2), 339-351.
- Marrero, Jorge A., Feng, Ziding, Wang, Yinghui, Nguyen, Mindie H., Befeler, Alex S., Roberts, Lewis R., . . . Schwartz, Myron. (2009). α -Fetoprotein, Des- γ Carboxyprothrombin, and Lectin-Bound α -Fetoprotein in Early Hepatocellular Carcinoma. *Gastroenterology*, 137(1), 110-118. doi: 10.1053/j.gastro.2009.04.005
- Marrero, Jorge A., Romano, Patrick R., Nikolaeva, Olga, Steel, Laura, Mehta, Anand, Fimmel, Claus J., . . . Block, Timothy M. (2005). GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. *Journal of Hepatology*, 43(6), 1007-1012. doi: 10.1016/j.jhep.2005.05.028
- Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, 128(584), 2145-2166. doi: 10.1256/003590002320603584
- Moribe, T., Iizuka, N., Miura, T., Kimura, Naoki, Tamatsukuri, S., Ishitsuka, H., . . . Oka, M. (2009). Methylation of multiple genes as molecular markers for diagnosis of a small, well-differentiated hepatocellular carcinoma. *Int.J.Cancer*, 125, 388-397.
- Singal, Amit G., Conjeevaram, Hari S., Volk, Michael L., Fu, Sherry, Fontana, Robert J., Askari, Fred, . . . Marrero, Jorge A. (2012). Effectiveness of Hepatocellular Carcinoma Surveillance in Patients with Cirrhosis. *Cancer Epidemiology Biomarkers & Prevention*, 21(5), 793-799. doi: 10.1158/1055-9965.epi-11-1005
- Smith, Robert A., Cokkinides, Vilma, Brooks, Durado, Saslow, Debbie, & Brawley, Otis W. (2010). Cancer Screening in the United States, 2010: A Review of Current American Cancer Society Guidelines and Issues in Cancer Screening. *CA: A Cancer Journal for Clinicians*, 60(2), 99-119. doi: 10.3322/caac.20063
- Song, Benjamin P, Jain, Surbhi, Lin, Selena Y, Chen, Quan, Block, T M, Song, Wei, . . . Su, Y-H. (2012). Detection of Hypermethylated Vimentin in Urine of Patients with Colorectal Cancer. *Journal of Molecular Diagnostics*, 14(2).
- Su, Ying-Hsiu, Lin, Selena Y, Song, Wei, & Jain, Surbhi. (2014). DNA markers in molecular diagnostics for hepatocellular carcinoma. *Expert Review of Molecular Diagnostics*, 14(7), 803-817. doi: doi:10.1586/14737159.2014.946908
- Su, Ying-Hsiu, Wang, Mengjun, Norton, P. A., Brenner, Dean E., & Block, T. M. (2008). Detection of mutated K-ras DNA in urine, plasma and serum from patients with colorectal carcinoma or adenomatous polyps. *Annals of the New York Academy of Sciences*, 1137, 197-201.
- Wright, Lorinda M., Kreikemeier, Jeff T., & Fimmel, Claus J. (2007). A concise review of serum markers for hepatocellular cancer. *Cancer Detection and Prevention*, 31(1), 35-44.
- Yang, Bin, Guo, Mingzhou, Herman, James G., & Clark, Douglas P. (2003). Aberrant Promoter Methylation Profiles of Tumor Suppressor Genes in Hepatocellular Carcinoma. *American Journal of Pathology*, 163(3), 1101-1107.