IDENTIFYING STAGE 1 HEPATOCELLULAR CARCINOMA PATIENTS WITH POOR PROGNOSIS

J. B. NATION

ABSTRACT. **Aim:** The purpose of this note is to show that survival times for stage 1 HCC patients has a bimodal distribution, and to present a genetic signature for identifying the high-risk group.

Methods: Clinical data from TCGA gives the survival distribution. A predictive signature is extracted from TCGA gene expression data using the LUST algorithm.

Results: About 20% of stage 1 liver cancer patients die within the first two years, while 70% live at least four years, often much longer. A genetic signature of 2 genes related to immune response identifies the high-risk group.

Conclusion: Since the patients in the high-risk group can be recognized with fair accuracy, they should be potential candidates for alternate therapy.

Introduction. Early diagnosis of hepatocellular carcinoma (HCC) is perhaps the primary factor for patient outcome with liver cancer [1]. Patients treated for stage 1 HCC have a good chance of surviving five years or longer. However, the survival distribution for stage 1 HCC patients is bimodal: roughly 20% of these patients die within the first two years after diagnosis, another 10% during the next two years, while 70% live at least four years, often much longer. This is illustrated with a histogram in Figure 1. The survival histogram projected from the ecdf curve is given in Figure 2.

If the poor prognosis group could be recognized, then these patients would be candidates for alternate treatment. While the patients in the short-term survival group have other liver disease, such as hepatitis or cirrhosis, they are not different in that respect from the long-term survivors for whom current treatment protocols are effective. Moreover, our analysis shows that the gene expression profile for stage 1 HCC

Date: October 15, 2019.

Key words and phrases. hepatocellular carcinoma, stage 1, Kaplan-Meier survival analysis, genetic signature.

These results were presented at the 2018 Hepatobiliary and Pancreatic Cancer Conference at the University of Hawai'i Cancer Center, April 27–28, 2018, Honolulu.

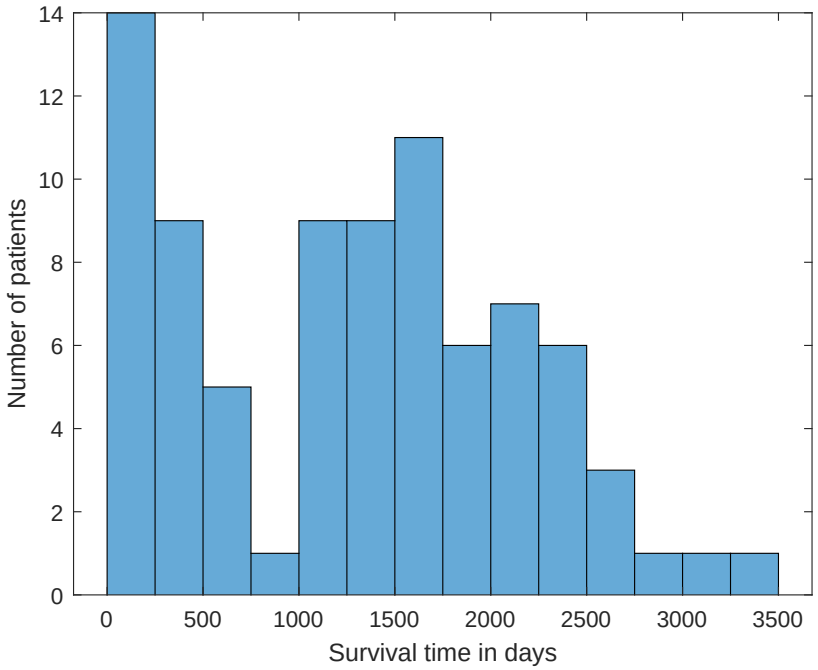

FIGURE 1. Survival histogram for stage 1 liver cancer patients, based on TCGA clinical records for 157 stage 1 HCC patients with tumors weighing under 500 grams. The mean survival time, censored and uncensored, for this group was 910 days. This histogram represents those 83 patients who either (1) died at any time, or (2) survived at least 910 days. For patients in the second group who are still alive, the number of *days to last follow-up* is used in place of *days to death*. Thus it greatly underestimates survival times for the low-risk group, as only 10 of the 54 patients indicated at times > 910 days were actually deceased.

patients is very different from that for later stages, which suggests that resection and observation may be appropriate for the low-risk cohort.

Methods. Clinical and mRNA expression data for HCC patients was obtained from The Cancer Genome Atlas (TCGA). The data for stage 1 patients with tumor weight less than 500 grams was extracted, giving 157 subjects. (Later analysis showed that excluding the larger tumors had little effect on the results.) All calculations were done in MATLAB (MathWorks, Natick, MA, USA).

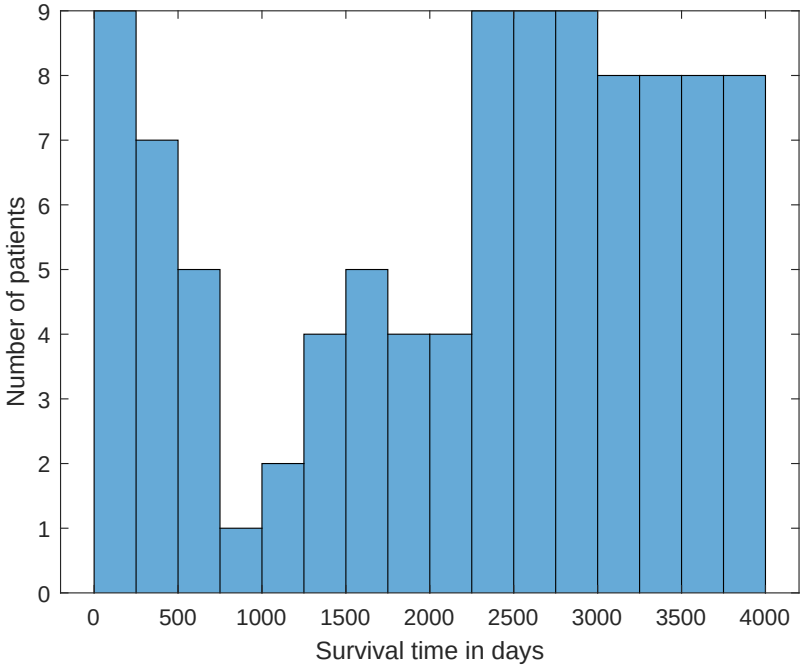

FIGURE 2. Survival histogram for 100 patients predicted by the empirical cumulative distribution curve based on Figure 1. The ecdf makes no prediction past 2500 days, so that part of the histogram has been flattened to represent the number of cases (50/100).

The mRNA expression data was analyzed using the LUST algorithm to find sets of genes with coordinated expression patterns [2]. This is a two-step algorithm. The first step looks for sets of genes that maximize a graph-theoretic objective function, unsupervised by clinical information. These *metagenes* are then refined in the second step, using clinical data to find subsets that separate the Kaplan-Meier survival curves. The algorithm produces a number of signatures predicting survival, and ranks them.

For the stage 1 liver patients, a signature consisting of 7 genes relating to immune response was chosen by the algorithm: BIN2, C1QB, CD53, DOCK2, EVI2A, ITGB2, NCKAP1L. Later refinements reduced the signature to 2 genes: BIN2 and C1QB. Each patient is assigned a score that is a linear combination of his/her expression of those genes, by projecting the patient's expression onto the singular vectors of the expression matrix for those genes only, as in [2, 3, 4].

Patients with a score lower than a chosen threshold, representing reduced expression of these immune system genes, were assigned to the high-risk group. The accuracy of the predictor was then analyzed in terms of the standard confusion table; see [5] for a thorough summary. The results were fairly constant over a range of thresholds.

Results. Let us designate patients who die in the first 750 days after diagnosis as *short-term survivors*. To analyze the performance of the signature for predicting short-term survivors, subjects who are censored at less than 750 days are removed from the data set. If we use a test score threshold of $s = -.32$, we obtain the results shown in Figure 3 and summarized in Table 1. Patients in the high-risk group have just under a 50% chance of surviving for 750 days, while those in the low-risk group have about an 85% chance. The Kaplan-Meier survival curves for the low-risk and high-risk groups are shown in Figure 4.

In terms of accuracy, the test correctly places about 70% (19/27) of the short-term survivors in the high-risk group (true positive rate). The true negative rate with this threshold is 74% (50/68), for an overall accuracy of 73% (69/95).

	low score	high score
long survival	18	50
short survival	19	8

TABLE 1. Scores vs. survival. A low score is $s < -0.32$ based on the expression of the genes BIN2 and C1QB; a high score is $s \geq -0.32$. Long survival is ≥ 750 days, while short survival is < 750 days.

The survival statistics are based on deaths from all causes. Of the 27 deaths during the first two years, 11 occurred during the first four months (≤ 120 days). These could perhaps be attributed to general liver disease or complications of surgery; the clinical record does not specify. The test places 7 of these 11 in the high-risk group.

The remaining 16 deaths under two years occurred between 170 and 700 days. Of these, 7 were due to recurrence of liver cancer, 5 were due to extrahepatic recurrence, while the cause for 4 deaths was not specified in the record. The test puts 12 of these 16 patients in the high-risk group.

If we remove the patients who died in the first 120 days, a clearer picture emerges. In doing so, we make the implicit assumption that

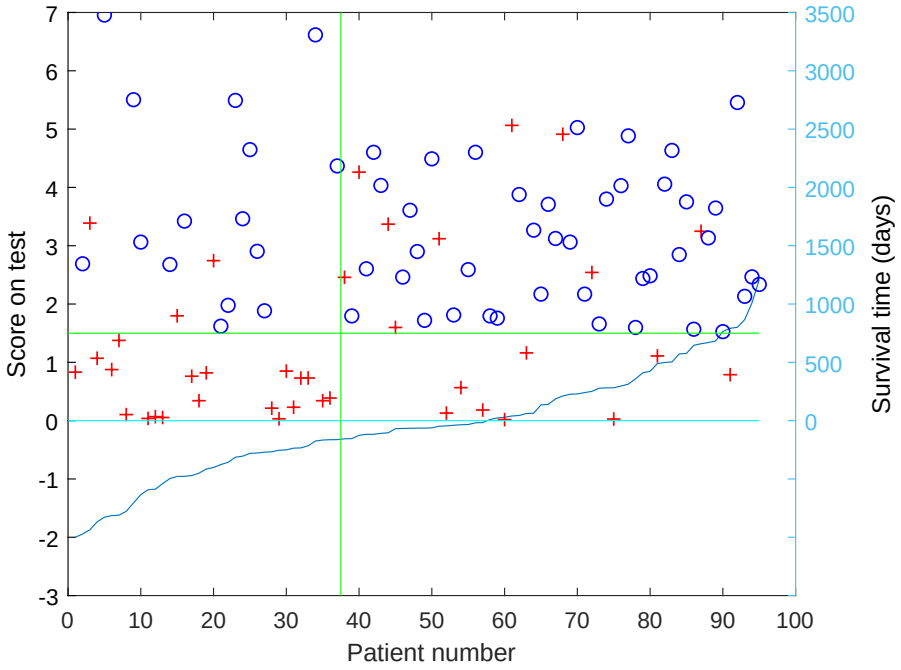

FIGURE 3. This graph plots each patient's score on the 2-gene test *vs.* survival. The score s on the test, arranged in increasing order, is indicated by the blue curve, in units shown on the left y-axis. The vertical green line is at the threshold score -0.32, so that patients to the left of the green line are in the high-risk group, and patients to the right represent the low-risk group. Survival is measured on the right y-axis in days The horizontal green line is at 750 days (just over 2 years). A red + indicates death, while a blue $\circ$ is the survival time at last follow-up. Censored patients (those alive at last follow-up $<$ 750 days) are not shown.

the cause of death for these patients is more than stage 1 liver cancer. The results shown in Figure 5 and summarized in Table 2. Patients in the high-risk group have a 60% chance of surviving for 750 days, while those in the low-risk group have over a 90% chance.

Some other risk factors for liver cancer failed to predict the short-term survivors. Fibrosis and hepatitis B even appeared to convey a slight patient benefit, though this is surely a statistical anomaly. However, 8 of the 28 short-term survivors (29%) had hepatitis C, as compared with 24 of the remaining 127 stage 1 HCC patients (19%). Most

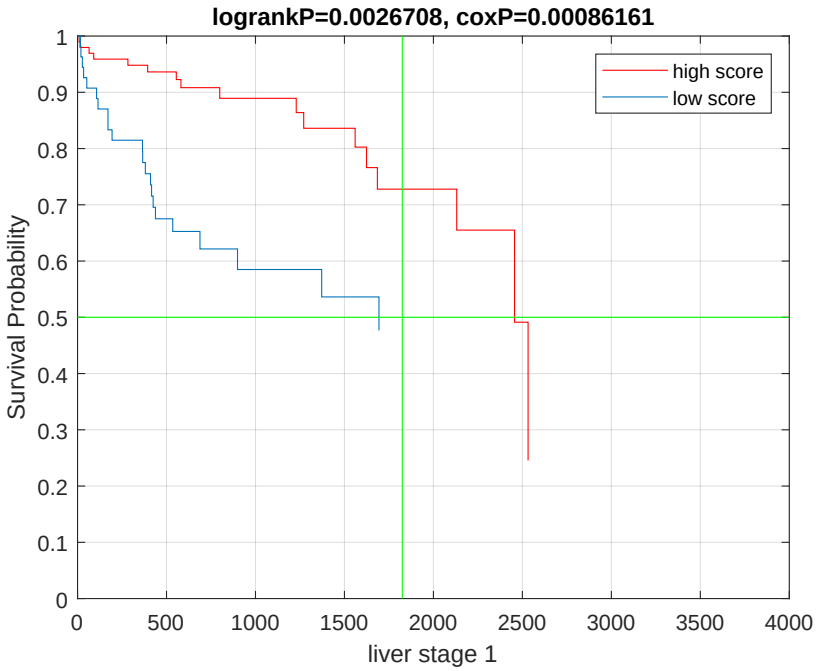

FIGURE 4. Kaplan-Meier curves for the high-risk and low-risk groups with respect to survival

	low score	high score
long survival	18	50
short survival	12	4

TABLE 2. Scores vs. survival with patients who died in the first 120 days removed.

of the patients with larger tumors were censored, so no conclusion could be drawn there. An analysis of mutation data from TCGA likewise yielded nothing useful for predicting survival or recurrence.

Discussion. For the majority of patients diagnosed with early stage HCC and small tumors, treatment involving primarily resection is effective [6]. However, there is a distinct cohort, consisting of about 20% of these patients, that is at great risk during the first two years after diagnosis. The main objective of this note is to distinguish this cohort as a group that needs to be considered separately as potential candidates for alternate therapy. Perhaps equally important is to identify those for whom resection is likely to be curative.

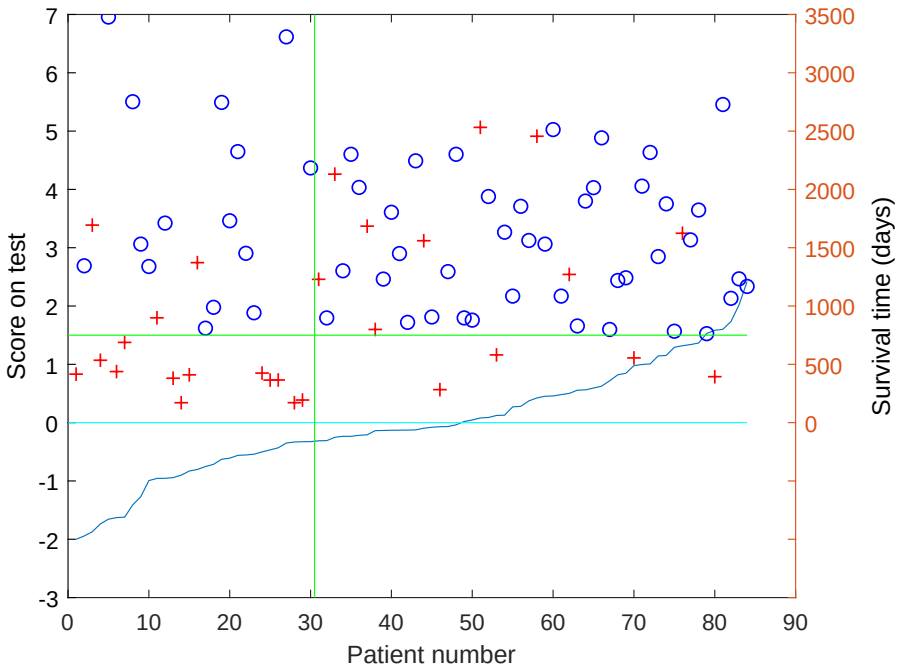

FIGURE 5. Patient scores vs. survival with patients who died in the first 120 days removed.

The second point is that high-risk stage 1 HCC patients can be identified using a genetic signature. This suggests genetic testing after resection or transplantation as part of a rubric, along with clinical considerations, to identify high-risk patients.

All this suggests that the high-risk group could benefit from additional treatment, but necessarily begs the question of what form that alternate treatment might take. That is a more difficult clinical question, not to be answered from a broad data analysis. The options for adjuvant treatment of HCC are very limited, and many patients would not be candidates for further treatments because of poor liver function. The fact that the signature is comprised of immune regulatory genes suggests that perhaps some sort of immunotherapy could be appropriate; see e.g. [7, 8, 9, 10]. The data could also be interpreted as supporting increased use of neoadjuvant therapy for early stage liver cancer, which has shown promising results [11, 12].

The increasing incidence of HCC makes finding an effective treatment for the high-risk group a problem worthy of attention.

1. Declarations

Acknowledgments. These results were presented at the 2018 Hepatobiliary and Pancreatic Cancer Conference at the University of Hawai'i Cancer Center, April 27–28, 2018, Honolulu. The author wishes to thank Drs. Linda Wong and Robert Lewandowski for helpful discussions.

Data source and availability. This research is based clinical and mRNA expression data for hepatocarcinoma, publicly available from The Cancer Genome Atlas (TCGA).

Financial support and sponsorship. None.

Conflicts of interest. There are no conflicts of interest.

Patient consent. Not applicable.

Ethics approval. Not applicable.

References

[1] Llovet, J., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwarz, M., Sherman, M., Gores, G.: Hepatocellular carcinoma. Nat. Rev. Dis. Primers (2016 Apr 14), doi: 10.1038/nrdp.2016.18.

[2] Nation, J., Okimoto, G., Wenska, T., Achari, A., Holmes, T., Maligro, J., Yoshioka, T., Zitello, E.: A comparative analysis of mRNA expression for 33 different cancers. Draft available at https://github.com/tristanh314/lust-cancer-2019.

[3] Okimoto, G., Zeinalzadeh, A., Wenska, T., Loomis, M., Nation, J., Fabre, T., Tiirikainen, M., Hernandez, B., Wong, L., Kwee, S.: The joint analysis of multiple, high-dimensional data types using sparse matrix factorizations of rank-1 with applications to ovarian and liver cancer. BioData Mining, 2016, 9:24, doi: 10.1186/s13040-016-0103-7.

[4] Nation, J.: A genetic signature predicting survival and metastasis for melanoma patients. Preprint 2019. Available at https://github.com/tristanh314/lust-cancer-2019.

[5] J. M. Hughes-Oliver, Assessment of prediction algorithms for ranking objects. Notices Amer. Math. Soc., Feb. 2019, 182–190.

[6] Bruix, J., Reig, M., Sherman, M.: Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. Gastroenterology 150 (2016 Apr), 835–853, doi: 10.1053/j.gastro.2015.12.041.

[7] Büttner, N., Schmidt, N., Thimme, R.: Perspectives of immunotherapy for hepatocellular carcinoma (HCC). Z. Gastroenterol. 54 (2016 Dec), 1334–1342, doi: 10.1055/s-0042-120417.

[8] Korangy, F., Höchst, B., Manns, M.P., Greten, T.F.: Immune response in hepatocellular carcinoma. Dig. Dis. 28 (2010 May 7), 150–154, doi: 10.1159/00028279.

[9] Sia, D., Jiao, Y., Martinez-Quetglas, I., Kuchuk, O., Villacorta-Martin, C., Castro de Moura, M., Putra, J., Camprecios, G., Bassaganyas, L., Akers, N., Losic, B., Waxman, S., Thung, S.N., Mazzaferro, V., Esteller, M., Friedman, S.L., Schwartz, M., Villanueva, A., Llovet, J.M.: Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features, Gastroenterology 153 (2017 Sep), 812–826, doi: 10.1053/j.gastro.2017.06.007.

[10] Sprinzl, M.F., Galle, P.R.: Immune control in hepatocellular carcinoma development and progression: role of stromal cells. Semin. Liver Dis. 34 (2014 Nov), 376–388, doi: 10.1055s-0034-1394138.

[11] Lewandowski, R.J., Gabr, A., Abouchaleh, N., Ali, R., Al Asadi, A., Mora, R.A., Kulik, L., Ganger, D., Desai, K., Thornburg, B., Mouli, S., Hickey, R., Caicedo, J.C., Abecassis, M., Riaz, A., Salem, R.: Radiation Segmentectomy: Potential Curative Therapy for Early Hepatocellular Carcinoma. Radiology (2018 Apr 24) 171768. doi: 10.1148/radiol.2018171768.

[12] Tsutsui, R., Nagamatsu, H., Itano, O., Deguchi, A., Tsutsumi, T., Hiraki, M., Mizukami, N., Akiba, J.: Neoadjuvant hepatic arterial infusion chemotherapy for resectable hepatocellular carcinomas. Hepatoma Res. (2018) 4:13, doi: 10.20517/2394-5079.2018.20.

Department of Mathematics, University of Hawai'i, Honolulu, HI 96822, USA

Email address: jb@math.hawaii.edu