Genetic risk classifier to predict localised renal cell carcinoma recurrence

Authors' reply

We thank Shufang Liu and Jieping Lei for their interest in our Article.

First, Liu and colleagues suggested we could use the survival tree-based method to find a cutpoint for splitting the data into high-risk and low-risk groups. However, our goal for dividing data into these two groups was to evaluate the performance of the risk score classifier, and not to identify the optimal cutoff point that would provide the highest hazard ratio (HR) for the training set. The same number of experimental units in high-risk and low-risk groups is preferred if low variance and good reproducibility are desirable. In our study, we selected the median risk score as cutoff point; the HR of the classifier in the training set was 6·78 (95% CI 3·17–14·49) and the HRs for the three validation sets were 5·32 (95% CI 2·81–10·07; internal testing set), 5·39 (3·38–8·59; independent validation set), and 4·62 (2·48–8·61; The Cancer Genome Atlas set). If we select the cutoff by the survival tree-based method, the HRs in the training set and three validation sets are 7·22 (95% CI 3·38–15·43), 4·57 (2·50–8·36), 4·74 (3·06–7·36), and 4·26 (2·37–7·66), respectively. Compared with the survival tree-based method, the HR in the training set is lower with our approach, but higher in the three validation sets; therefore, our six-single-nucleotide polymorphism-based classifier has better reproducibility in the three validation sets when we use the cutoff of median genetic risk score. The use of median or tertiles as a cutoff point to divide data into two or three groups is very common for testing model performance in clinical studies.[1–3]

Second, with few training data, the parameter estimates will have greater variance, whereas with few testing data, our performance statistic will have greater variance. Therefore, there are no clear advantages of using the suggested 70:30 ratio over our approach (50:50). Considering the performance of our model, an even 50:50 ratio for training versus testing sets is preferred, and this ratio is also very common in clinical studies[1–4] to divide data in a way that neither variance is too high. Some previous research has shown that the optimal splitting proportion is dependent on model complexities, which are associated with the probability of error on the training and testing sets.[5]

We declare no competing interests.

*Jia-Xing Zhang, Yi-Hui Pan, Jun Lu, Guan-Nan Shu, Pei-Xing Li, Cai-Xia Li, *Jun-Hang Luo*

luojunh@mail.sysu.edu.cn

Department of Urology (J-XZ, Y-HP, JL, G-NS, J-HL), Department of Oncology (J-XZ), First Affiliated Hospital, and School of Mathematics and Computational Science (P-XL, C-XL), Sun Yat-sen University, Guangzhou 510080, China

1 Chen HY, Yu SL, Chen CH, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007; **356:** 11–20.

2 Liu N, Chen NY, Cui RX, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol* 2012; **13:** 633–41.

3 Qu L, Wang ZL, Chen Q, et al. Prognostic value of a long non-coding RNA signature in localized clear cell renal cell carcinoma. *Eur Urol* 2018; **74:** 756–63.

4 Qiu J, Peng B, Tang Y, et al. CpG methylation signature predicts recurrence in early-stage hepatocellular carcinoma: results from a multicenter study. *J Clin Oncol* 2017; **35:** 734–42.

5 Guyon I. A scaling law for the validation-set training-set size ratio. Murray Hill, NJ: AT&T Bell Laboratories, 1997.