Supplementary Materials for

Sequential reinforcement active feature learning for gene signature identification in renal cell carcinoma

Meng Huang[1], Xiucai Ye[1,2,*], Akira Imakura[1,2] and Tetsuya Sakurai[1,2]

[1]Department of Computer Science, University of Tsukuba, Tsukuba, 3058577, Japan
[2]Center for Artificial Intelligence Research, University of Tsukuba, Tsukuba, 3058577, Japan
[*]Corresponding author: Xiucai Ye. Tel.: 029-853-5449; Fax: 029-853-5449; E-mail:yexiucai@cs.tsukuba.ac.jp

Supplementary Figures

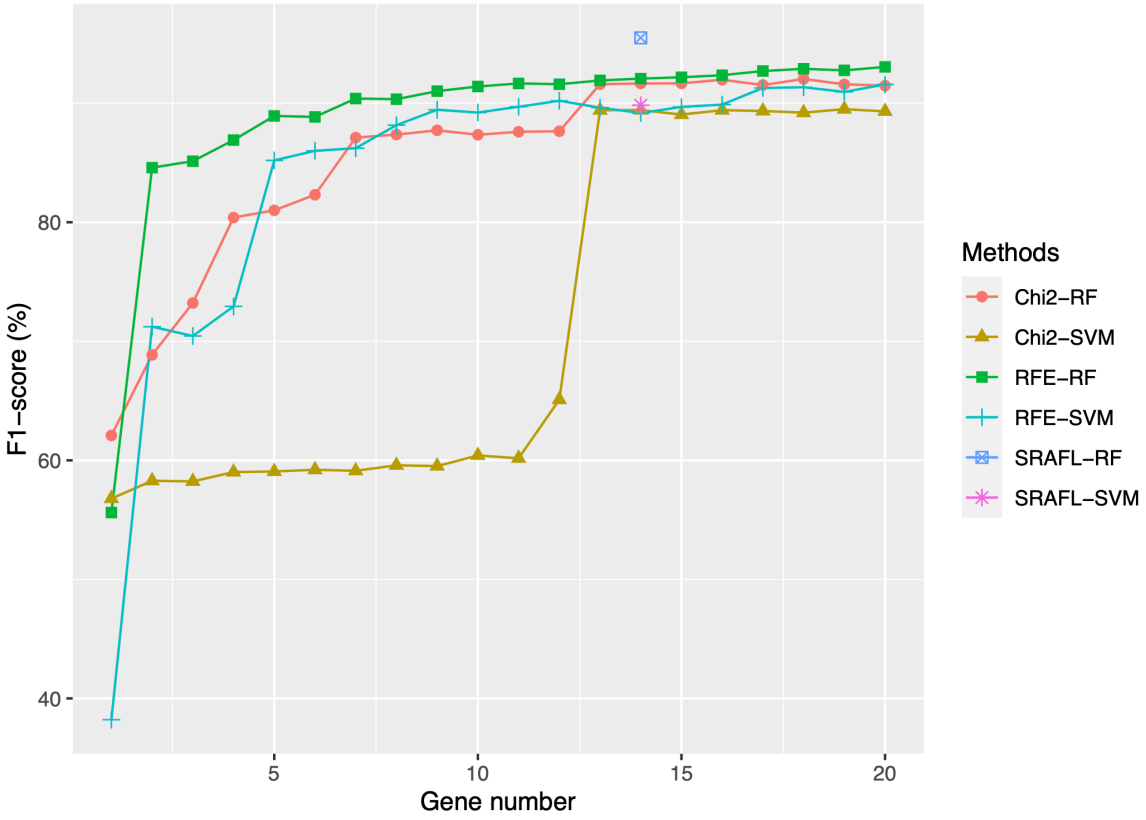

Figure S1. The F1-score comparison with other gene selection methods (Chi2 and RFE) on the RCC subtypes data. The SRAFL method uses the final selected gene signatures to train the classifiers. The number of genes used by the Chi2 and RFE methods varies from 1 to 20.

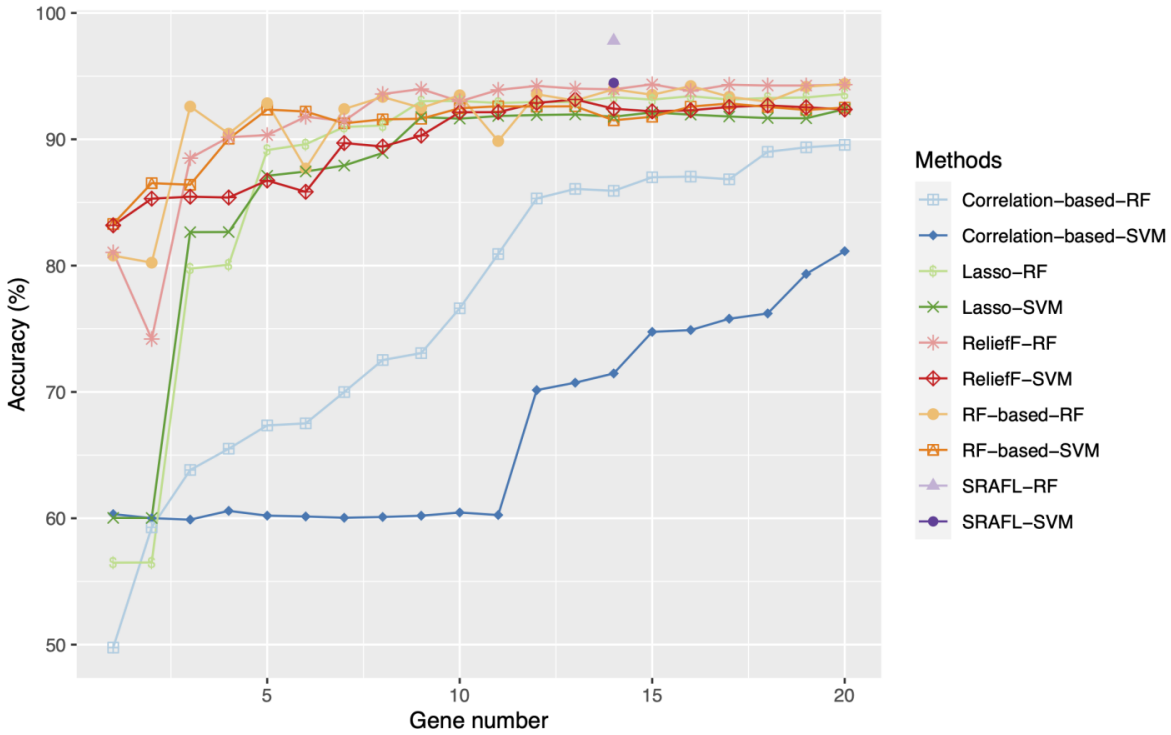

Figure S2. The accuracy comparison with traditional feature selection methods (RF-based, Correlation-based, Lasso, and ReliefF) on the RCC subtypes data. The SRAFL method uses the final selected gene signatures to train the classifiers. The number of genes used by the RF-based, Correlation-based, Lasso, and ReliefF methods varies from 1 to 20.

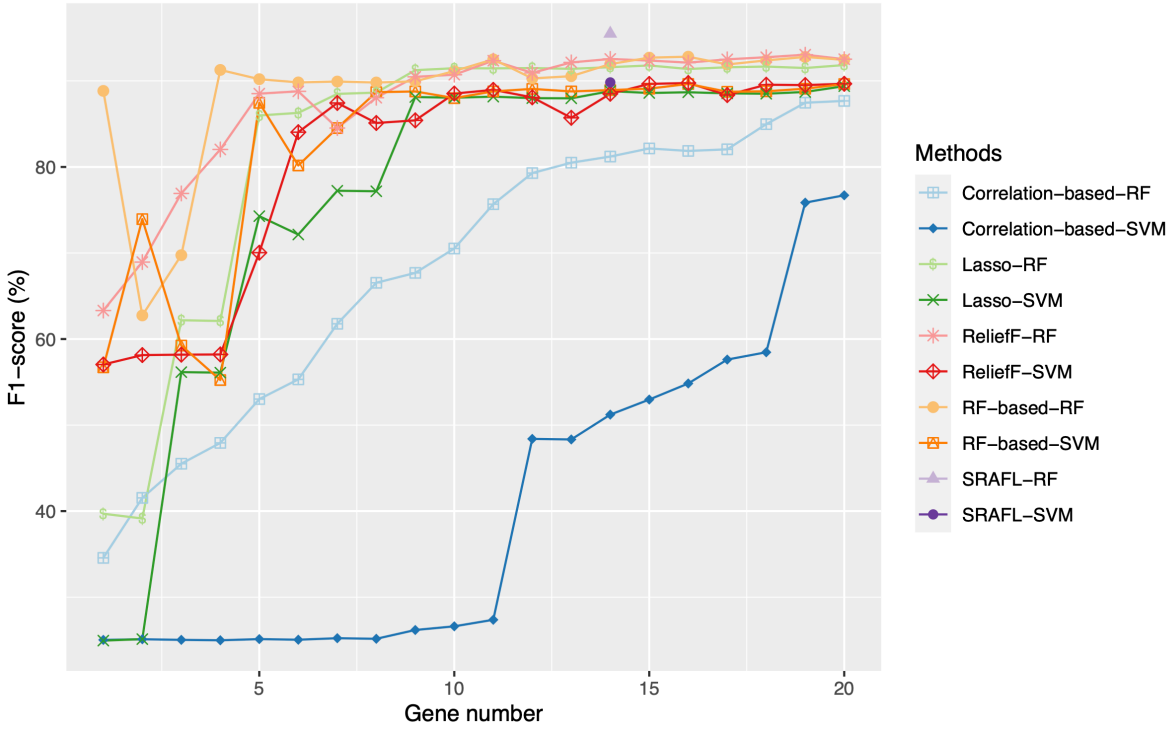

Figure S3. The F1-score comparison with traditional feature selection methods (RF-based, Correlation-based, Lasso, and ReliefF) on the RCC subtypes data. The SRAFL method uses the final selected gene signatures to train the classifiers. The number of genes used by the RF-based, Correlation-based, Lasso, and ReliefF methods varies from 1 to 20.

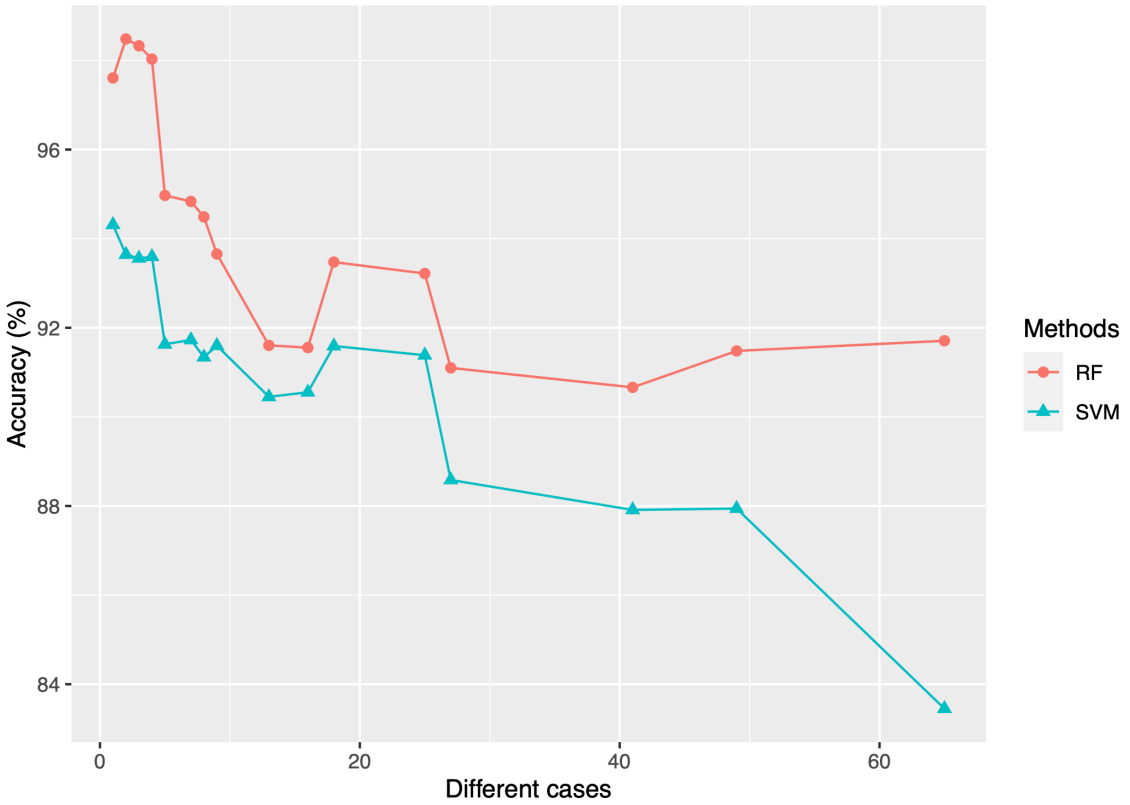

Figure S4. Accuracy by varying the number of selection genes in different cases. Each case consists of different gene signatures in different RCC subtypes. All selected gene signatures are divided to 16 cases including case ⩾ 1, case ⩾ 2, case ⩾ 3, case ⩾ 4, case ⩾ 5, case ⩾ 7, case ⩾ 8, case ⩾ 9, case ⩾ 13, case ⩾ 16, case ⩾ 18, case ⩾ 25, case ⩾ 27, case ⩾ 41, case ⩾ 49, and case ⩾ 65.