



# Ordinal Multi-modal Feature Selection for Survival Analysis of Early-Stage Renal Cancer

Wei Shao<sup>1</sup>, Jun Cheng<sup>2</sup>, Liang Sun<sup>1</sup>, Zhi Han<sup>4</sup>, Qianjin Feng<sup>3</sup>,  
Daoqiang Zhang<sup>1(✉)</sup>, and Kun Huang<sup>4(✉)</sup>

<sup>1</sup> College of Computer Science and Technology, Nanjing University of Aeronautics  
and Astronautics, Nanjing 211106, China

dqzhang@nuaa.edu.cn

<sup>2</sup> School of Biomedical Engineering, Shenzhen University, Shenzhen 518073, China

<sup>3</sup> School of Biomedical Engineering, Southern Medical University, Guangzhou  
510515, China

<sup>4</sup> School of Medicine, Indiana University, Indianapolis, IN 46202, USA  
kunhuang@iu.edu

**Abstract.** Existing studies have demonstrated that combining genomic data and histopathological images can better stratify cancer patients with distinct prognosis than using single biomarker, for different biomarkers may provide complementary information. However, these multi-modal data, most high-dimensional, may contain redundant features that will deteriorate the performance of the prognosis model, and therefore it has become a challenging problem to select the informative features for survival analysis from the redundant and heterogeneous feature groups. Existing feature selection methods assume that the survival information of one patient is independent to another, and thus miss the ordinal relationship among the survival time of different patients. To solve this issue, we make use of the important ordinal survival information among different patients and propose an ordinal sparse canonical correlation analysis (i.e., OSCCA) framework to simultaneously identify important image features and eigengenes for survival analysis. Specifically, we formulate our framework basing on sparse canonical correlation analysis model, which aims at finding the best linear projections so that the highest correlation between the selected image features and eigengenes can be achieved. In addition, we also add constrains to ensure that the ordinal survival information of different patients is preserved after projection. We evaluate the effectiveness of our method on an early-stage renal cell carcinoma dataset. Experimental results demonstrate that the selected features correlated strongly with survival, by which we can achieve better patient stratification than the comparing methods.

---

This work was supported in part by the National Natural Science Foundation of China (61473149,61732006), the NCI ITCR (U01CA188547) and Shenzhen Peacock Plan (KQTD2016053112051497).

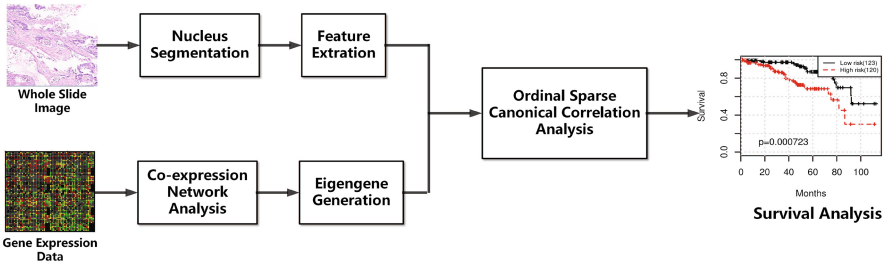
© Springer Nature Switzerland AG 2018

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11071, pp. 648–656, 2018.

[https://doi.org/10.1007/978-3-030-00934-2\\_72](https://doi.org/10.1007/978-3-030-00934-2_72)

# 1 Introduction

Cancer is the most common form of diseases worldwide. It is reported that the number of affected people in developing country will reach 20 million annually as early as 2025 [1]. Effective and accurate prognosis prediction of human cancer, especially at its early stage has attracted much attention recently. So far, many biomarkers have been shown to be associated with the prognosis of cancers, including the histopathological images, genomic signatures, proteomics makers and demographical information.



**Fig. 1.** The flowchart of the proposed method.

Early studies on the prognosis of cancer often focus on using single-modality biomarker (*e.g.*, imaging or genomic data). However, in these studies, some useful complementary information across different modalities of data is ignored. Recently, some studies explored to combine both imaging and genomic biomarkers for survival analysis [2–4]. For instance, Cheng *et al.* [2] constructed a novel framework that can predict the survival outcomes of patients with renal cell carcinoma by using a combination of quantitative image features and gene expression feature. Yuan [3] *et al.* integrated both image data and genomic data to improve the survival prognosis for breast cancer patients. These existing studies have suggested that different modalities of data complement with each other and provide better patient stratification when used together.

Although integrating imaging and genomic features can better predict the clinical outcome for cancer patients, simply combing these features may bring redundant features that will deteriorate the prediction performance, and thus feature selection is a key step for multi-modal feature fusion. In the existing studies [2, 3], the authors usually simply concatenate multi-modal data together at first, and then apply traditional feature selection methods (*e.g.*, LASSO) to select components that are related to cancer prognosis. However, these feature selection methods assume that the survival time of one patient is independent to another, and thus missing the strong ordinal relationship among the survival time of different patients, *e.g.*, the survival time of patient A is longer than that of patient B. In addition, most of the studies [2, 3] directly combine morphological and genomic data together for survival analysis, which neglects correlation

among the multi-modal data. As a matter of fact, the exploitation of multi-modal association has been widely accepted as a key component of the state-of-the-art multi-modality based machine learning approaches [5].

Based on the above consideration, in this paper, we take advantage of the ordinal survival information among different cancer patients, and propose an ordinal sparse canonical correlation analysis (OSCCA) framework that can select features from multi-modal data for survival analysis. Specifically, we formulate our framework based on sparse canonical correlation analysis (*i.e.*, SCCA), which is a powerful association method that can identify linear projections to achieve the highest correlation between the selected imaging and genomic components. In addition, we add constraints to ensure that the ordinal information of different groups of patients is preserved, *i.e.*, the average projection of the patients from the long-term survival groups should be larger than that of short-term survival groups. The experimental results on a public available early-stage clear cell renal cell carcinoma (ccRCC) dataset demonstrate that the proposed method outperforms comparing methods in terms of patient stratification.

**Table 1.** Demographics and clinical characteristics

Characteristics	Summary	Characteristics	Summary
Patients		Stage	
Censored	188	Stage I	201
Non-censored	55	Stage II	42
Age (Y):	$59.1 \pm 12.2$	Follow-up (M):	$43.2 \pm 25.2$

## 2 Method

Figure 1 shows the flowchart of our framework, which includes three major steps, *i.e.*, feature extraction, ordinal sparse canonical correlation analysis based feature selection (OSCCA), and prognostic prediction. Before giving the detailed descriptions of these steps, we will firstly introduce the dataset used in this study.

**Dataset:** The Cancer Genome Atlas (TCGA) project has generated multimodal genomic and imaging data for different types of cancer. Renal cell carcinoma is the most common type of cancer arising from kidney. In this study, we test our method on an early-stage (*i.e.*, stage I and stage II) ccRCC dataset [2] derived from TCGA. Specifically, this dataset contains pathological imaging, genomic, and clinical data for 243 early-stage renal cell carcinoma patients. Of the 243 samples, 188 patients are censored, which means that the death events of them were not observed during the follow-up period, and their exact survival times are longer than the recorded data. The remaining 55 samples are non-censored patients, and their recorded survival times are the exact time from initial diagnosis to death. Table 1 summarizes the demographics of all the samples.

**Feature Extraction:** For each image, we firstly apply the method in [6] to segment the nucleus in the whole-slide image, and then for each segmented nucleus, we extract ten different features [2], *i.e.*, nuclear area (denoted as area), lengths of the major and minor axes of cell nucleus, and the ratio of major axis length to minor axis length (major, minor, and ratio), mean pixel values of nucleus in RGB channels respectively (rMean, gMean, and bMean), and mean, maximum, and minimum distances (distMean, distMax, and distMin) to its neighboring nuclei. Next, for each type of feature, a 10-bin histogram and five statistic measurements (*i.e.*, mean, SD, skewness, kurtosis, and entropy) are used to aggregate the cell-level features into patient-level features, and thus a 150-dimensional imaging feature for each patient can be obtained. Here, we use area\_bin1 to represent the percentage of very small nuclei while area\_bin10 indicates the percentage of very large nuclei in the patient sample. As to gene expression data, we firstly use co-expression network analysis algorithms to cluster genes into co-expressed modules, and then summarize each module as an eigengene (gene modules are shown in Supplementary Materials). This algorithm yields 15 coexpressed gene modules. More details about the genomic feature extraction can be found in [2].

**Sparse Canonical Correlation Analysis:** For the derived imaging and eigengene features, we implement our feature selection model under SCCA framework. Specifically, let  $\mathbf{X}_H \in R^{N \times p}$  be the histopathological imaging data, and  $\mathbf{X}_G \in R^{N \times q}$  be the extracted eigengenes data, where  $N$  is the number of the patients, and  $p$  and  $q$  are the feature number of imaging data and eigengene data, respectively. The objective function of SCCA is:

$$\begin{aligned} \min_{\boldsymbol{\omega}_H, \boldsymbol{\omega}_G} & -(\boldsymbol{\omega}_H)^T (\mathbf{X}_H)^T \mathbf{X}_G \boldsymbol{\omega}_G + r_1 \|\boldsymbol{\omega}_H\|_1 + r_2 \|\boldsymbol{\omega}_G\|_1 \\ \text{s.t.} & \|\mathbf{X}_G \boldsymbol{\omega}_G\|_2^2 \leq 1; \|\mathbf{X}_H \boldsymbol{\omega}_H\|_2^2 \leq 1 \end{aligned} \quad (1)$$

where the first term in Eq. (1) seeks linear transformations (*i.e.*,  $\boldsymbol{\omega}_H, \boldsymbol{\omega}_G$ ) to achieve the maximal correlation between imaging and eigengene data, the second and third L1-norm regularized terms are used to select a small number of feature that can maximize the association between the multi-modal data.

**Ordinal Sparse Canonical Correlation Analysis:** In the SCCA model, we only consider the mutual dependency between imaging and genomic data, and thus ignore the survival information of patients. Although the study in [2] used the survival information for feature selection, they assume that the survival information of one patient is independent to another, and thus miss the strong ordinal relationship among the survival time of different patients. To address this problem, we propose an ordinal sparse canonical correlation analysis (OSCCA) method to simultaneously identify important features from the multi-modal data. Specifically, we divide  $\mathbf{X} = [\mathbf{X}_H, \mathbf{X}_G] \in R^{N \times (p+q)}$  into  $\mathbf{X}^C$  and  $\mathbf{X}^{NC}$ , where  $\mathbf{X}^C \in R^{k \times (p+q)}$  and  $\mathbf{X}^{NC} \in R^{(N-k) \times (p+q)}$  correspond to the multi-modal features for censored and non-censored patients, respectively, and  $k$  denotes the number of censored patients. We also define  $\mathbf{Y} = [\mathbf{Y}^C, \mathbf{Y}^{NC}]$ ,

where  $\mathbf{Y}^C \in R^k$  and  $\mathbf{Y}^{NC} \in R^{(N-k)}$  indicate the recorded survival time for censored and non-censored patients, respectively. In order to reduce the chance that all patients in one group are censored, we divide all the patients (include both censored and non-censored patients) into four groups with equal size based on the quartiles of their recorded survival time, where each patient in group  $i$  ( $i = 1, 2, 3, 4$ ) has longer survival time than that in group  $j$  if  $i > j$ . We denote the mean imaging and eigengene feature for censored patients in group  $i$  as  $\mathbf{u}_H^i$  and  $\mathbf{u}_G^i$ , and those for non-censored patients in group  $i$  as  $\mathbf{v}_H^i$  and  $\mathbf{v}_G^i$ , respectively. We show the objective function of the OSCCA model as:

$$\min_{\omega_H, \omega_G} -(\omega_H)^T (\mathbf{X}_H)^T \mathbf{X}_G \omega_G + r_1 \|\omega_H\|_1 + r_2 \|\omega_G\|_1 + r_3 \left\| \mathbf{X}^{NC} \omega - \mathbf{Y}^{NC} \right\|_2^2 \tag{2}$$

$$s.t. \quad (\mathbf{v}_G^{i+1} - \mathbf{v}_G^i) \omega_G > 0, \quad (\mathbf{v}_H^{i+1} - \mathbf{v}_H^i) \omega_H > 0 \tag{3}$$

$$(\mathbf{u}_G^{i+1} - \mathbf{u}_G^i) \omega_G > 0, \quad (\mathbf{u}_H^{i+1} - \mathbf{u}_H^i) \omega_H > 0 \tag{4}$$

where the first three terms in Eq. (2) are as same as they are stated in the SCCA model, the fourth part is a regression term, where  $\omega = [\omega_H, \omega_G] \in R^{p+q}$ . We use this term to estimate the relationship between the multi-modal data and the survival time for non-censored patients, since their survival information are accurate. We add two linear inequalities in (3) to ensure that the ordinal survival information of different groups of non-censored patients is preserved after the projections are adopted on both imaging and eigengenes data. In addition, since the genuine survival time for censored patients are longer than the recorded data, it is easy to infer that the average projection for the censored patients in groups  $i + 1$  should be larger than that for non-censored patients in group  $i$ , and we also add this ordinal relationship for both eigengene and imaging data by adding two inequality constraints shown in (4).

**Optimization:** We adopt an alternating strategy to optimize  $\omega_H$  and  $\omega_G$  in the proposed OSCCA model. Specifically, given the fixed  $\omega_H$ , the optimization problem for  $\omega_G$  can be reformulated as:

$$\min_{\omega_G} \frac{1}{2} (\omega_G)^T \mathbf{A} \omega_G + (\omega_G)^T \mathbf{B} + r_2 \|\omega_G\|_1 \tag{5}$$

$$s.t. \quad \mathbf{C} \omega_G > 0$$

where  $\mathbf{B} = -(\mathbf{X}_G)^T \mathbf{X}_H \omega_H + 2r_3 (\mathbf{X}_G^{NC})^T (\mathbf{X}_H^{NC} \omega_H - \mathbf{Y}^{NC})$ , in which  $\mathbf{X}_G^{NC} \in R^{(N-k) \times p}$  and  $\mathbf{X}_H^{NC} \in R^{(N-k) \times q}$  correspond to the imaging and eigengene data for non-censored patients, respectively. Also,  $\mathbf{A} = 2r_3 (\mathbf{X}_G^{NC})^T \mathbf{X}_G^{NC}$ , and  $\mathbf{C} = [\mathbf{v}_G^4 - \mathbf{v}_G^3; \mathbf{v}_G^3 - \mathbf{v}_G^2; \mathbf{v}_G^2 - \mathbf{v}_G^1; \mathbf{u}_G^4 - \mathbf{u}_G^3; \mathbf{u}_G^3 - \mathbf{u}_G^2; \mathbf{u}_G^2 - \mathbf{u}_G^1] \in R^{6 \times q}$ . For the optimization problem in (5), we adopt the alternating direction method of multipliers (*i.e.*, ADMM) algorithm to solve it. To change the problem in (5) into ADMM form, we introduce variables  $\mathbf{J} \in R^q$  and non-negative vector  $\theta \in R^6$ , which is used to transform the inequality constraints  $\mathbf{C} \omega_G > 0$  into equality

constrains  $C\omega_G - \theta = 0$ , Eq. (5) can be reformulated as:

$$\begin{aligned} \min_{\omega_G} & \frac{1}{2}(\omega_G)^T A\omega_G + (\omega_G)^T B + r_2\|J\|_1 \\ \text{s.t.} & \quad J = \omega_G, \quad C\omega_G - \theta = 0 \end{aligned} \tag{6}$$

Then, the augmented Lagrangian form of Eq. (6) can be written as:

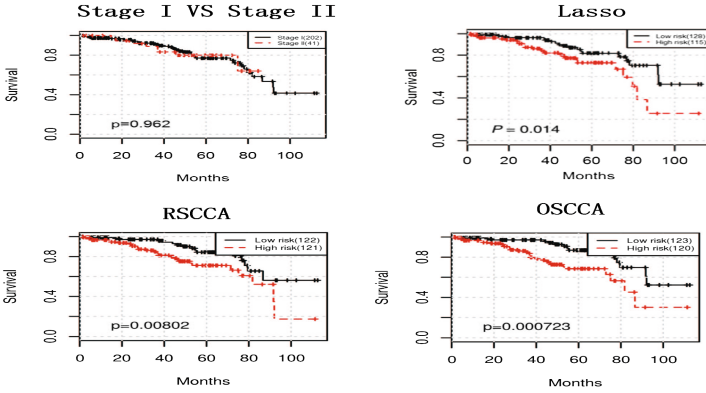
$$\begin{aligned} L(\omega_G, J, \theta, Q, R) &= \frac{1}{2}(\omega_G)^T A\omega_G + (\omega_G)^T B + r_2\|J\|_1 + \langle Q, \omega_G - J \rangle \\ &+ \frac{\rho_1}{2} \|\omega_G - J\|_2^2 + \langle R, C\omega_G - \theta \rangle + \frac{\rho_2}{2} \|C\omega_G - \theta\|_2^2 \end{aligned} \tag{7}$$

where  $Q$  and  $R$  are Lagrange multipliers. A general ADMM scheme for Eq. (7) repeats the following 5 steps until convergence: (1)  $\omega_G \leftarrow \arg \min_{\omega_G} L(\omega_G, J, \theta, Q, R)$ : It is a convex problem with respect to  $\omega_G$  and we can solve it via gradient descent method. (2)  $J \leftarrow \arg \min_J L(\omega_G, J, \theta, Q, R)$ : This optimization problem can be formulated as:  $\min_J \frac{\rho_1}{2} \|\omega_G - J\|_2^2 + r_2\|J\|_1 - (Q)^T J$ . Since the L1-norm is non-differentiable at zero, a smooth approximation has been estimated for L1 term by including an extremely small value. Then, by taking the derivative regarding to  $J$  and let it to be zero, we can obtain  $J = (r_2 D + \rho_1 I)^{-1}(Q + \rho_1 \omega_G)$ , where  $D$  is a diagonal matrix with the  $k$ -th element as  $1/\|J_k\|_1$ . Here,  $J_k$  denotes the  $k$ -th element in  $J$ . (3)  $\theta \leftarrow \arg \min_{\theta} L(\omega_G, J, \theta, Q, R)$ : It has a close form solution with the  $k$ -th element  $\theta_k = \max(0, T_k)$ , where  $T_k$  corresponds to the  $k$ -th element in  $T = C\omega_G + \frac{1}{\rho_2} R$ . (4)  $Q = Q + \rho_1(\omega_G - J)$ . (5)  $R = R + \rho_2(C\omega_G - \theta)$ . After  $\omega_G$  is determined, we use similar method to optimize  $\omega_H$ .

**Prognostic Prediction:** We build Cox proportional hazards model [2] for survival analysis. Specifically, we firstly divide all patients into 10 folds, with 9 folds used for training the proposed OSCCA model and the remaining for testing, then the Cox proportional hazards model is built on the selected features in the training set. After that, the median risk score predicted by the cox proportional hazards model is used as a threshold to split patients into low-risk and high-risk groups. Finally, we test if these two groups has distinct survival outcome using Kaplan-Meier estimator and log-rank test [2].

### 3 Experimental Results

**Experimental Settings:** The parameters  $r_1, r_2, r_3$  in the OSCCA model are tuned from  $\{2^{-4}, 2^{-5}\}$ ,  $\{2^{-5}, 2^{-6}\}$  and  $\{2^{-5}, 2^{-6}\}$ , respectively,  $\rho_1$  and  $\rho_2$  in Eq. (7) are fixed as  $2^{-3}$ . All the algorithms are implemented using MATLAB 2017.



**Fig. 2.** Comparisons of the survival curves by applying different feature methods.

**Results and Discussion:** We compare the prognostic power of our proposed OSCCA method with several other methods, including LASSO [2] and RSCCA, as well as tumor staging. Compared to the proposed OSCCA model, RSCCA method has the same objective function (shown in Eq. (2)), but neglect to take the ordinal survival information (shown in the inequalities in (3) and (4)) into consideration. We show the survival curves of these four methods in Fig. 2. It is observed that on one hand, the Kaplan-Meier curves for tumor Stage I and Stage II are intertwined (log-rank test  $P = 0.962$ ), which demonstrates that the stratification of the early-stage renal cell carcinoma patients is a challenging task, on the other hand, OSCCA could achieve significantly better patient stratification (log-rank test  $P = 7.2e - 3$ ) than the comparing methods, which shows the advantage of using ordinal survival information for feature selection. In addition, it is worth noting that the proposed RSCCA could provide better prognostic prediction than the LASSO method, this is because RSCCA considers the correlation among different modalities for feature selection, which is better than the direct combination strategy.

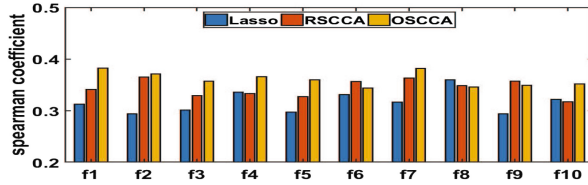
Next, in order to investigate the association between the selected imaging feature and eigengenes, the spearman coefficients between  $X_H\omega_H$  and  $X_G\omega_G$  on 10-fold testing data are shown in Fig. 3. Obviously, OSCCA generally outperforms the comparing methods in term of identifying high correlation between imaging data and genomic data, and the better exploration of the inherent correlation within multi-modal data may be the reason for the better patient stratification performance of the proposed OSCCA method.

Lastly, we compare the features selected by OSCCA with those selected by [2] in Table 2. We find that our method can identify new types of image features (*i.e.*, area\_bin10) that are related to large nuclei. It has been demonstrated that the ccRCC patients with large values of nuclei size have worse prognosis [7] than other patients. As to genomic features, two novel eigengenes (*i.e.*, eigengene 9, eigengene 14) are identified. The enrichment analysis on eigengene 9 shows that it is related to mitotic cell cycle process and genome stability, and genes in this

**Table 2.** Comparisons of the selected features by the method in [2] and OSCCA.

	Image feature	Eigengenes
LASSO [2]	rMean_bin8, rMean_mean, distMean_bin10, rMean_mean	Eigengene3, Eigengene13
OSCCA	rMean_bin9, major_kurtosis, area_bin10, distMin_bin1, bMean_bin8	Eigengene3, <i>Eigengene9</i> , <i>Eigengene14</i>

module are frequently observed to co-express in multiple types of cancers [8]. In addition, eigengene 14 is enriched with genes that are associated with immune response, and it is reported that the deregulation of the immune response genes are associated with the initiation and progression of cancers [9], and our discovery can potentially shed light on the emerging immunotherapies. These results further shows the promise of OSCCA to identify biologically meaningful biomarkers for the prognosis of early-stage renal cancer patients.



**Fig. 3.** The spearman coefficients between  $X_H\omega_H$  and  $X_G\omega_G$  on 10-fold testing data.

## 4 Conclusion

In this paper, we develop OSCCA which is an effective multimodal feature selection method for patient stratification aiming at identifying subgroups of cancer patients with distinct prognosis. The strength of our approach is its capability of utilizing the ordinal survival information among different patients to identify features that are associated with patient survival time. Experimental results on an early-stage multi-modal renal cell carcinoma dataset have demonstrated that the proposed OSCCA can identify new types of image features and gene modules that are associated with patient survival, by which we can achieve significantly better patients stratification than the comparing methods. Such prediction is particularly important for early stage patients when the prediction is important yet staging information from pathologists is not sufficient to meet the needs. OSCCA is a general framework and can be used to find multi-modal biomarkers for other cancers or predict response of specific treatment, which allow for better patients management and cancer care in precision medicine.



## References

1. Torre, L., Bray, F., Siegel, R.: Global cancer statistics. *Cancer J. Clin.* **65**(2), 87–108 (2012)
2. Cheng, J., Zhang, J., Han, Y., Huang, K.: Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. *Cancer Res.* **77**(21), 91–100 (2017)
3. Yuan, Y., Failmezger, H., Rueda, O.: Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Transl. Med.* **4**(157), 143–157 (2012)
4. Yao, J., Zhu, X., Zhu, F., Huang, J.: Deep correlational learning for survival prediction from multi-modality data. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 406–414. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-66185-8\\_46](https://doi.org/10.1007/978-3-319-66185-8_46)
5. Liu, F., Chen, H., Shen, D.: Inter-modality relationship constrained multi-modality multi-task feature selection for alzheimer’s disease and mild cognitive impairment identification. *Neuroimage* **84**(1), 466–475 (2014)
6. Phoulady, H., Goldgof, D., Mouton, P.: Nucleus segmentation in histology images with hierarchical multilevel thresholding. In: Proceedings of SPIE 2016, pp. 1–8 (2016)
7. Kanamaru, H., Akino, H., Suzuki, Y.: Prognostic value of nuclear area index in combination with the world health organization grading system for patients with renal cell carcinoma. *Urology* **57**(2), 257–261 (2001)
8. Zhang, J., Lu, K., Xiang, Y., Huang, K.: Weighted frequent gene co-expression network mining to identify genes involved in genome stability. *Plos Comput. Biol.* **8**(8), 1–14 (2012)
9. Kim, S., Park, C., Kim, H., Kang, M.: Deregulation of immune response genes in patients with epstein-barr virus-associated gastric cancer and outcomes. *Gastroenterology* **148**(1), 137–147 (2015)