PAMOGK: A Pathway Graph Kernel based Multi-Omics Clustering Approach for Discovering Cancer Patient Subgroups

Yasin Ilkagan Tepeli[1[0000−0002−3375−6678]] ⋆, Ali Burak Ünal[2,3[0000−0002−7279−620X]] ⋆,†, Furkan Mustafa Akdemir[3[0000−0003−0948−5756]], and Oznur Tastan[1[0000−0001−7058−5372]] ⋆⋆

[1] Faculty of Engineering and Natural Sciences, Sabanci University, 34956, Istanbul, Turkey
[2] Dept of Computer Science, University of Tübingen, 72076, Tübingen, Germany
[3] Dept of Computer Engineering, Bilkent University, 06800, Ankara, Turkey

Abstract. Accurate classification of patients into molecular subgroups is critical for the development of effective therapeutics and for deciphering what drives these subgroups to cancer. The availability of multi-omics data catalogs for large cohorts of cancer patients provides multiple views into the molecular biology of the tumors with unprecedented resolution. We develop PAMOGK (Pathway based Multi Omic Graph Kernel clustering) that not only integrates multi-omics patient data with existing biological knowledge on pathways. We develop a novel graph kernel that evaluates patient similarities based on a single molecular alteration type in the context of a pathway. To corroborate multiple views of patients evaluated by hundreds of pathways and molecular alteration combinations, we use multi-view kernel clustering. Applying PAMOGK to kidney renal clear cell carcinoma (KIRC) patients results in four clusters with significantly different survival times (p-value $= 1.24e$-11). When we compare PAMOGK to eight other state-of-the-art multi-omics clustering methods, PAMOGK consistently outperforms these in terms of its ability to partition KIRC patients into groups with different survival distributions. The discovered patient subgroups also differ with respect to other clinical parameters such as tumor stage and grade, and primary tumor and metastasis tumor spreads. The pathways identified as important are highly relevant to KIRC. PAMOGK is available at github.com/tastanlab/pamogk

Keywords: Patient Stratification · Graph Kernels · Multi-view Clustering · Pathways

1 Introduction

Cancer is a molecularly diverse disease; within the same cancer type, patients bear different molecular alterations, which manifest themselves as different clinical trajectories [6, 58, 28]. Discovering coherent subgroups of patients with similar molecular profiles is essential to developing better diagnostic tools and subtype specific treatment strategies. Knowledge of molecular subtypes is also key to finding the mechanisms that drive these different subtypes to cancer. The problem of stratifying patients based on their molecular profiles is also critical for other complex diseases.

Characterization of patients with omics technologies and the availability of multi-omics data on large cohorts of patients opens up opportunities for better stratification of patients [51, 48, 6]. Towards this goal, several multi-omics clustering methods have been proposed (reviewed in [33]) to integrate the multi-dimensional data collected on patients. The simple form of integration is early integration. In this case, patient features derived from single omic data are concatenated, and standard clustering is applied to this combined feature representation. However, this approach equally weighs each data type and suffers from a curse of dimensionality as the higher dimensional features dominate the clustering. There are more sophisticated early integration approaches that aim to overcome these problems. iClusterBayes and its earlier variants [41, 26, 25] and LRACluster assume a latent lower-dimensional distribution of data and uses regularization. A second strategy is to deploy late integration approaches [33]. In this case, the samples are clustered with each omic data type separately, and the ensemble's cluster assignments are combined into a single clustering solution. The consensus clustering by Monti et al. [27] is frequently used for cancer subtyping [13, 51]. PINS[31] and COCA[14] fall into this category as well. These approaches have the

⋆ These authors contributed equally to the work as first authors.
⋆⋆ Corresponding author: otastan@sabanciuniv.edu

drawback that they do not capture the correlations between the different data types. This strategy leads to poor clustering when each view individually contains a weak signal.

Alternatively, several intermediate integration algorithms have been developed [34]. For example, SNF [55] constructs a patient similarity network using each data type, and these similarities are then fused in a single similarity network through an algorithm based on message passing. Meng *et al.* [24] applies dimension reduction to the axes of maximal covariance between data types, JIVE [23] utilizes the variations in data. MCCA [59, 3] extends the canonical correlation analysis (CCA) [11] to a multi-view setting. There are also several algorithms, which are developed as generic multi-view algorithms ([62]). For example, [16] and [4] extend the spectral clustering algorithm [54], which relies on partitioning a similarity network of samples.

There are also kernel-based intermediate integration algorithms, which take a multi-view clustering approach. Kernel methods are powerful methods, where the samples' similarities are implicitly calculated in a higher-dimensional space [38]. Several generic multi-view kernel clustering methods (reviewed in [61]) have been developed where some have been applied to cancer subtyping [45, 10]. rMKL-LPP [45] extends the multi-view kernel framework [18] to the multi-omics clustering. A kernel matrix is computed from each omic data type, and a linear combination of kernels is sought for the clustering of the patients in kernel k-means. The localized multiple kernel k-means (LMKMM) [10] method also assumes a linear combination of the views but learns a sample specific kernel matrix weight in a k-means framework.

Although corroborating multi-omics data is important to construct a better view of patient similarities, it might not be sufficient to boost the signal as often only a small fraction of molecular alterations is common among the patients. Analyzing molecular data in the context of molecular networks is a widely used approach to overcome this heterogeneity and sparsity problem (reviewed in [5]). In this work, we present PAMOGK, a multi-view kernel clustering approach, which integrates multi-omics patient data with pathways using graphs [4]. PAMOGK represents each patient as a set of vertex labeled undirected graphs, where each graph represents the gene interactions in a biological pathway, and the vertex labels are assigned based on patient specific molecular alterations. To quantify patient similarity over a pathway and to attain an omic view, we introduce a novel graph kernel, the smoothed shortest path graph kernel (SmSPK), which extends the shortest path graph kernel [2]. While existing graph kernels are designed to capture the topological similarities of the graphs, SmSPK captures the similarities of the vertex label within the graph context. This allows us to capture patients' similarities that stem from the dysregulation of similar processes in the pathways. By utilizing multi-view kernel clustering approaches, PAMOGK stratifies patients into subgroups. PAMOGK also offers additional insights by showing how informative each pathway and the data type is to the clustering process based on the assigned kernel weights.

We apply our methodology to kidney renal cell carcinoma(KIRC) data made available through the Cancer Genome Atlas Project (TCGA) [29]. We integrate the patient somatic mutations, gene expression levels, and protein expression dataset. Compared to the state-of-the-art multi-omics clustering methods, PAMOGK consistently outperforms in terms of its ability to partition into groups with different prognosis. Extracting of the relative importance of pathways in the clustering process show that the discovered pathways that are relevant to KIRC. PAMOGK is available at https://github.com/tastanlab/pamogk.

2 Methods

Given a set of cancer patients, $\mathcal{S}$, for which molecular profiles of the tumors are available, PAMOGK aims to stratify them into k subgroups through integrating pathways. Formally, we would like to find a partitioning $\mathcal{C}$ such that: $\mathcal{S}$ is grouped into k number of disjoint subsets C_i's where, $\mathcal{S} = \cup_{i=1}^{k} C_i$ and where $C_i \cap C_i = \emptyset$. In this section, we detail the steps of PAMOGK and data processing used in our experiment. Let M be the

[4] The early results of this work are presented as an extended abstract in [49].

number of pathways, D be the molecular alteration types (mutations, altered expression, etc.) available for the patients and N be the number of patients.

2.1 PAMOGK Overview

PAMOGK involves three main steps (Figure 1). In the first step, each pathway is represented with an undirected graph. Next, for a given molecular alteration type, i.e., somatic mutations, a patient's molecular alterations are mapped on the pathway. These alterations constitute the patient-specific node labels of the patient's graph. Thus, a "view" is constructed for each pathway-molecular alteration type pair. To assess a pair of patients' similarity under a view, in the second step, the novel graph kernel, SmSPK, is computed to quantify a patient pair's similarity over a pathway and a molecular alteration type. Each $N \times N$ kernel matrix constitute a *view* to the patient similarities. In the final step, to stratify cancer patients into meaningful subgroups, these multiple kernels are input to a multi-view kernel clustering algorithm. In the following sections, we elaborate on each step of PAMOGK with more technical details.

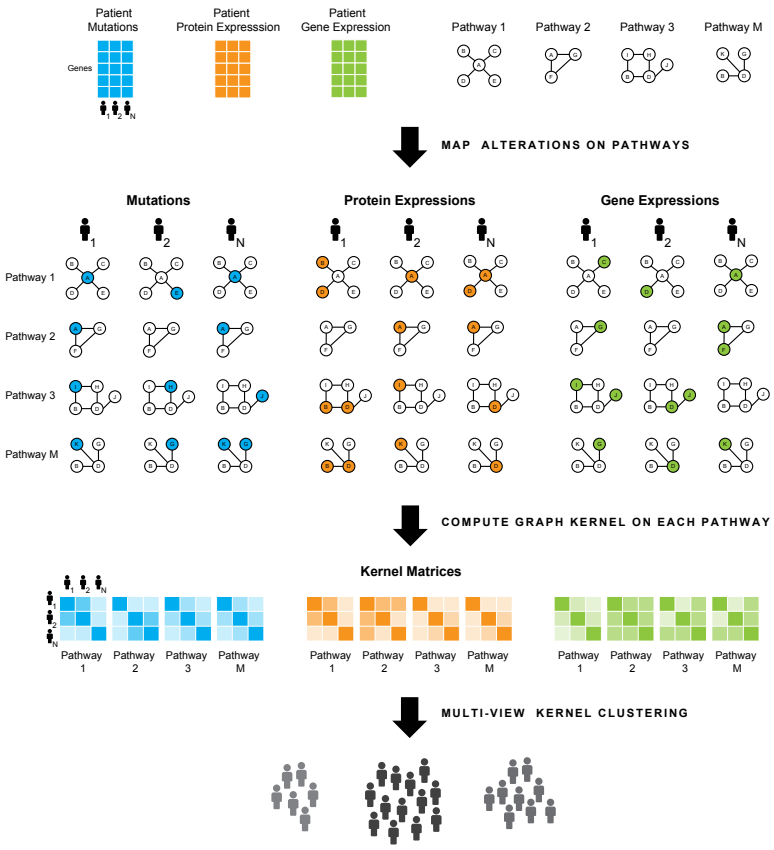


Fig. 1: The PAMOGK framework. PAMOGK takes different omic measurements (shown in different colors) and pathways as input. Note that pathway graphs are shown smaller than usual due to size constraints. Each pathway-omic pair constitute a view. In a view, each patient is represented with an undirected graph whose interactions are based on the pathway, and the node labels are molecular alterations of the genes for that patient. For each view, a patient-by-patient graph kernel matrix is computed to assess patient similarities under that pathway-alteration view. In the final step, these views are input to a multi-view kernel clustering method to obtain the patient clusters.

2.2 Step 1: Patient graph representation

We first convert each pathway to an undirected graph where nodes are genes, and an edge exists if there is an interaction between the two genes. For each pathway graph i and patient j, we define an undirected vertex

labeled graph $G_i^{(j)} = (V_i, E_i, \ell_i^{(j)})$. $V_i = \{v_1, v_2, \ldots, v_n\}$ is the ordered set of n genes in the pathway i and $E_i \subset V_i \times V_i$ is a set of undirected edges between the genes in this pathway. The label set $\ell_i^{(j)} = \{l_1, l_2, \ldots, l_n\}$ is in the same order of V_i and represents the corresponding vertex's label for patient j. For a specific pathway, the pathway graph structure is the same for all patients and is defined by the set of interactions in the pathway while the vertex labels are different and are based on each patient's molecular alterations. For a patient j, $\ell_i^{(j)}$ entries are assigned based on the patient's molecular alteration profile. For example, in the case of somatic mutations, if the corresponding gene k is mutated in patient j, label of value 1 is assigned to this gene (node), and 0 otherwise. At the end of this step, we have $N \times M \times D$ labeled pathway graphs.

2.3 Step 2: Computing Multi-View Kernels with Graph Kernels

In this step, we would like to assess the similarities of the patients on a given pathway for a given molecular data type. For this, we resort to graph kernel functions. While typical kernels take vectors as input, a graph kernel takes two graphs as input and returns a real-valued number that quantifies the similarity of two input graphs: $\mathcal{K} : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{R}$ [52]. Powerful graph kernels are presented in earlier work [42, 2, 30]. However, these graph kernels are designed to compare graphs with different graph structures and to identify similarities and differences that arise from these different structures. In our case, though, we would like to compare graphs with identical topology but different node label distribution. The graphs' structures are identical because they are from the same pathway, and the label distributions are different because of the patient specific alterations. To assess the similarity of topologically identical graphs with different node label distribution, we devise a new graph kernel for our purposes.

Inspired from the shortest path graph kernel [2], SmSPK makes use of all shortest paths of the graphs to characterize them. We also smooth the node labels of a patient in the pathway so that if two patients have alterations in genes in close proximity, they contribute to the similarity even though the set of altered genes are not identical. To propagate node labels along the pathway, we use the random walk with restart, which is a common strategy used in various tasks (reviewed in [5]). For a single graph indexed by g, the label propagation is performed by employing the following formula for all patients:

$$\mathbf{S}_g^{(t+1)} = \alpha \mathbf{S}_g^{(\mathrm{t})} \mathbf{A}_g + (1 - \alpha)\, \mathbf{S}_g^{(0)}, \tag{1}$$

where $\mathbf{S_g}^{(0)}$ is a patient-by-gene matrix which represents the labels of the vertices in the graph g at time $t = 0$ and each row (patient) is determined by $\ell_g^{(j)}$. In this case, $\mathbf{S}_{g,ji}^{(0)} = 1$ where j is index of the patient and i is index of the vertex. $\mathbf{S_g}^{(t)}$ is the node label matrix at time t. $\mathbf{A}_g$ is the degree normalized adjacency matrix of the pathway graph g. $\alpha \in [0, 1]$ is the parameter that defines the degree of smoothing. We iterate over propagation until convergence is attained. We assign node attributes of the graph for each patient based on the final $\mathbf{S}$. Once we attain the label smoothed graphs of $G_g^{(i)}$ and $G_g^{(j)}$, we compute the similarities of these two graphs to each other as follows:

$$\mathcal{K}(G_g^{(i)}, G_g^{(j)}) = \sum_{p=1}^{P} \mathbf{s}_p^{(i)} . \mathbf{s}_p^{(j)} \tag{2}$$

Here, $\mathbf{s}_p^{(i)}$ is the vector that represents the labels of the vertices of the graph G_g on the shortest path p for patient i after smoothing, P P is the number of all pairs of shortest paths on the graph. The above function is a valid kernel function, as the dot product is the linear kernel, and the kernel property is preserved under summation.

For a given molecular alteration type and a pathway, we compute the kernel over all pairs of patients. $\mathbf{K}$ matrix is a symmetric $N \times N$ matrix, for which the i, j-th entry is the kernel function evaluated for patient i and patient j pair. By computing kernel matrices for each pathway and for each molecular alteration type,

we obtain $M \times D$ kernel matrices. We normalize the kernel matrices such that all kernel entries are in $[0-1]$ by dividing the kernel matrix entry $\mathbf{K}(i,j)$ entry by $\sqrt{(\mathbf{K}(i,i) * \mathbf{K}(j,j)}$

2.4 Step 3: Multi-View Kernel Clustering

Each of the kernel matrices computed in the previous section represents a view of the patients' similarities. To integrate these views, we resort to existing multi-view kernel clustering approaches. We experiment with different approaches (see Section 3.3); multiple kernel k-means with matrix-induced regularization (MKKM-MR) [21] performs the best. Thus the final model of PAMOGK uses MKKM-MR; yet, this step can be replaced by any multi-view clustering approach as long as the method accepts kernel matrices as input. In this section, for completeness, we provide a brief overview of the selected multi-view kernel clustering methods with which we experimented.

Multiple Kernel K-Means with Matrix-Induced Regularization: MKKM-MR algorithm objective is to minimize sum-of-squared loss over the cluster assignments. To reduce redundancy among kernel matrices and enhance the diversity of the selected kernel matrices, MKKM-MR [21] uses the matrix-induced regularization. The algorithm solves the following optimization problem:

$$\begin{aligned} \min_{\mathbf{H}\in\mathbb{R}^{nxk},\gamma\in\mathbb{R}^m_+} \quad & \mathrm{Tr}(\mathbf{K}_\gamma(\mathbf{I}_n - \mathbf{H}\mathbf{H}^\mathrm{T})) + \frac{\lambda}{2}\gamma^\mathrm{T}\mathbf{M}\gamma \\ \text{s.t.} \quad & \mathbf{H}^\mathrm{T}\mathbf{H} = \mathbf{I}_k \\ & \gamma^\mathrm{T}1_m = 1 \end{aligned} \tag{3}$$

Here, k is the number of clusters, n denotes the number of samples, m is the number of kernel matrices. $\mathbf{H}$ is the relaxed clustering assignment matrix, $\gamma = [\gamma_1, \gamma_2, \cdots, \gamma_m]$ are the weights of input kernel matrices. $\mathbf{K}_\gamma$ is the best kernel matrix, $\mathbf{M}$ is the matrix that measures the relation between kernel matrices. $\mathbf{I}_x$ is the x-by-x dimensional identity matrix, 1_m is m dimensional vector of ones. λ is the parameter that adjusts the trade-off between clustering cost and the regularization term.

Average Kernel K-Means (AKKM): Kernel k-means (KKM) [37] is simple but a strong baseline. Since it accepts a single kernel matrix, we input the average of the kernel matrices. We will refer to this method as average kernel k-means (AKKM).

Localized multiple kernel k-means (LMKMM): LMKMM [10] is another powerful method that optimizes not only the weight of the kernel matrices but also the weight of the samples. We reimplemented LMKKM in Python, which is originally provided in Matlab and R.

Similarity Network Fusion (SNF): SNF[56] calculates a similarity matrix of samples using an exponential kernel based on the view created by each data types separately, and construct a similarity network for each view. In these networks, samples are nodes and edge weights are the similarities. Through an iterative procedure based on message passing, the networks are fused into a single network. We use this fusion step of SNF as a multi-view clustering method, where we compute the similarities based on smSPK and cluster the samples with kernel k-Means or spectral clustering.

2.5 Dataset and Data Preprocessing

Pathway data: As the pathway source, we use National Cancer Institute - Pathway Interaction Database (NCI-PID) at NDEXBio [36][5]. NCI-PID is a curated database with focus on processes that are relevant to cancer research (download date: Apr 24, 2019). We filter out a pathway if it does not contain any overlapping gene with the omic data genes, which leaves out 165 pathways. The statistics of pathways sizes are provided in Supplementary Table 2.

[5] https://ndexbio.org/#/networkset/8a2d7ee9-1513-11e9-bb6a-0ac135e8bacf

Patient molecular and clinical data: The molecular and clinical data for KIRC is obtained from the TCGA PanCancer project [57]. We retrieve the data directly from Synapse[6]. We only consider the primary solid tumor samples and make use of three different molecular data types that can directly be mapped to pathways: somatic mutations, transcriptomics, and proteomics data. The transcriptomic data include the RNAseq gene expression levels, while protein expression is quantified through Reverse Phase Protein Array (RPPA). The exact data files are listed in Supplementary Table 1.

Assigning node labels based on molecular alterations: In the case of mutations, the patient node label is assigned as a binary label based on the presence or absence of the mutation. In the expression datasets, the gene and protein expression values are normalized and converted to z-scores relative to other patients. For each data type, if a gene(a protein)'s z-score is greater than 1.96(which stands for 95% confidence), the gene(the protein) is considered overexpressed while the genes(the proteins) with z-score lower than -1.96 is considered underexpressed. For the underexpressed and the overexpressed values, we use the z value as the node attribute. If the z-values are between -$1.96 < z < 1.96$, the node label is assigned to zero. In assigning the expression values, we considered different alternative methods; as discussed in the results section, yields slightly better results. Finally, from the three different omic data sources, five different types of alterations are defined: absence and presence of a somatic mutation in a gene, over and underexpression based on gene expression, over and underexpression based on protein expression.

3 Results and Discussion

3.1 Experimental Set up

We apply PAMOGK to discover different subgroups of KIRC patients. The dataset contains 361 patients whose molecular profiles come from three separate data types: somatic mutation, gene expression, and protein expression. We define five different molecular alteration types based on these three types of omics data (see Section 2.5). We compute one kernel matrix for each pathway-molecular alteration type; this results in 825 kernels (165 pathways $\times$ 5 molecular alteration types), each one of which constitutes a distinct view.

Throughout all experiments, we evaluate four different cluster numbers, $k = 2, 3, 4, 5$. When computing SmSPK, we try 12 different alpha α values (Supplementary Table 3). We conduct experiments by using different multi-view clustering methods. These include average kernel k-means(AKKM), LMKKM, MKKM-MR, SNF-kernel k-means and SNF-spectral clustering. If a pathway kernel includes a few or no altered genes, we eliminate it before inputting it into multi-view kernel clustering methods to increase time efficiency. The criteria for this is to eliminate those whose nonzero entries constitute at most 1% of all entries. The parameter λ in MKKM-MR is chosen using grid-search. (Supplementary Table 3).

We evaluate the clustering solutions through survival analysis in accordance with previous work [20, 17, 35, 9]. We compare the survival distributions of the clusters using Kaplan-Meier (KM) survival curves [15] and log-rank test's p-value [12]. In the log-rank test, we test whether there is a statistical difference between the survival times of the clusters. In comparing alternative methods, we use the p-value of this log-rank test as the performance criteria.

3.2 Assessing the Need of a new Graph Kernel

Constructing kernels, which reflect the similarity of patients, is a crucial step of PAMOGK. First, we would like to understand whether there is any merit in using SmSPK as opposed to deploying an already existing and powerful graph kernel. The motivation behind proposing a new kernel is that the existing graph kernels are designed to capture topological similarities. Since we compare the two patients on the same pathway, the structure of graphs shall always be the same. On the other hand, the node label distribution is different as it is patient specific. Thus, the existing graph kernels computed over the same pathway will consider patients as

[6] https://www.synapse.org/#!Synapse:syn300013

overly similar and would not serve our purpose. To check if this intuition holds, we analyze the distribution of the kernel values computed over all the pathways and the overexpressed molecular alteration type. Since the overexpressed genes are the densest kernels, we choose this data type. We compare SmSPK with kernels that accept continuous node attributes. These kernels include the propagation kernel [30], Graph Hopper kernel [8] and Wasserstein Weisfeiller Lehman graph kernel [47]. We use the implementation provided by the Grakel library [43] for propagation and graph hopper kernel. We use WWL Python library[47] provided by authors for the Wasserstein Weisfeiler Lehman graph kernel. We also tried to compare our method with node attributed shortest path graph kernel [2], but due to slow run times, we abandoned this comparison. Also, to asses, whether we need a graph kernel at all, we also compare to the case where RBF kernel is used. When RBF kernels are computed, they are directly evaluated on the omic data. Thus, they are computed over all the genes regardless of their participation in a pathway. The gamma values of RBF is determined by the median heuristic [39] (Supplementary Table 3). In order to make the comparison fair, we also apply smoothing and choose the results with the best smoothing parameter assignment for each method (Supplementary Table 3).

First we analyze the kernel values computed by each kernel. Figure 2a display the heatmaps computed by each kernel method on an example pathway (more examples are provided in Supplementary Figure Figure 1). The rows and the columns are patients the cell entries colors are proportional to the kernel value computed for the two patients over a single pathway and data type. Figure 1 clearly shows how kernels assign patient similarities of 1 very frequently. To better analyze this over all pathways, we analyze the distribution of kernel values assigned to patients by each of the different kernels. We bin the kernel matrix entries into groups for each kernel and calculate the frequency of each bin. Next, we calculate the average frequency for each bin across all kernel matrices computed (Figure 2b). Figure 2b shows, for each graph kernel, how the kernel values are distributed on average. All the kernels other than SmSPK, assign patient similarities of 1 very frequently (the darkest bin). These results confirm our intuition that due to the identical graph structures, the existing graph kernels are unable to distinguish patients with different molecular alterations on the same pathway graph.

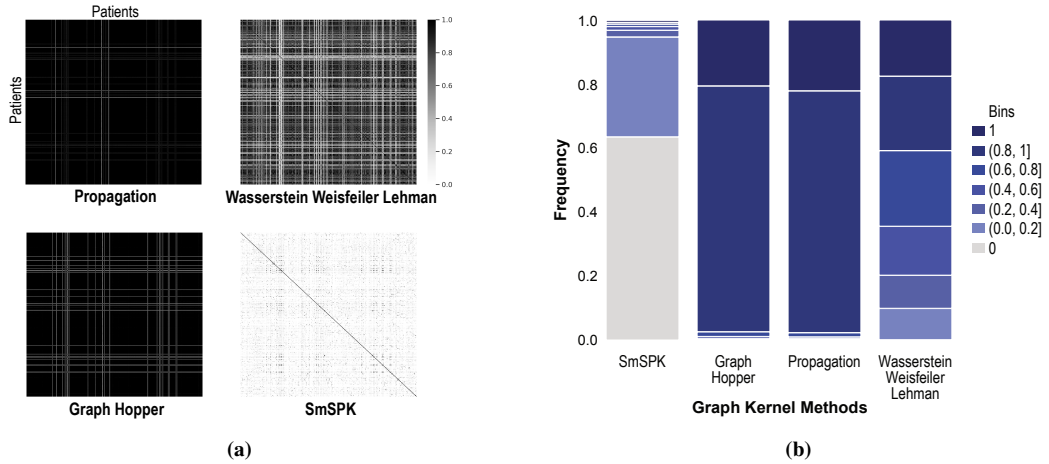
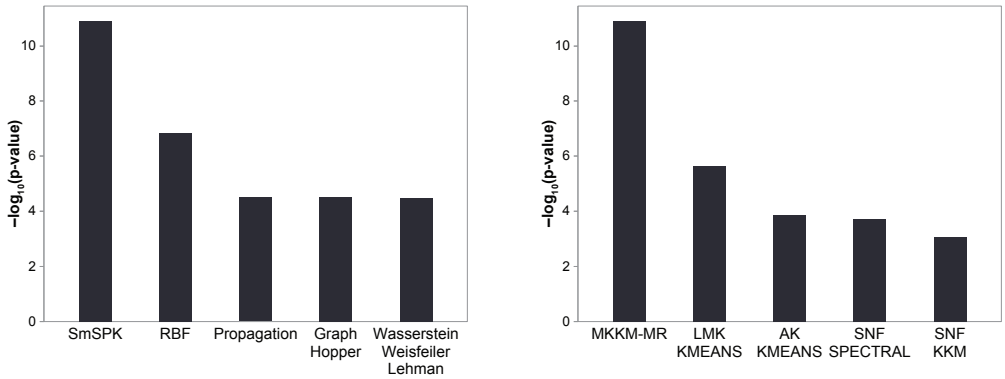


Fig. 2: a) Example heatmaps of patient-by-patient kernel matrices calculated by different kernel choices. The kernel functions include the propagation kernel, graph hopper kernel, Wasserstein Weisfeiler Lehman, and SmSPK graph kernel methods. Each kernel belongs to *the direct p53 effectors pathway* and overexpressed gene data type. The color black indicates that the similarity of the two patients is evaluated as 1. **b)** The frequency of patient similarities for different kernels over all pathways with the overexpression molecular data. For example, the darkest navy indicates the kernel value of 1, and the height of this bar is the proportion of patient-pairs for which the kernel value is 1. All the kernels other than SmSPK assign patient similarities of 1 very frequently.



(a) **(b)**

Fig. 3: a) The log-rank test p-values obtained with different choices of kernels employed with MKKM-MR multi-view kernel clustering algorithm. Kernel construction methods consist of SmSPK(our method), propagation graph kernel [30], graph hopper kernel [8], Wasserstein Weisfeiler Lehman graph kernel [46] and radial basis function (RBF) kernel. **b)** The log-rank test p-values obtained with different choices of multi-view kernel clustering methods withSmSPK as the kernel construction method. The clustering methods include average kernel k-means (AKKM), localized multiple kernel k-means (LMKKM) [10], multiple kernel k-means with matrix-induced regularization (MKKM-MR) [21], SNF[55] with spectral clustering and kernel k-means(KKM).

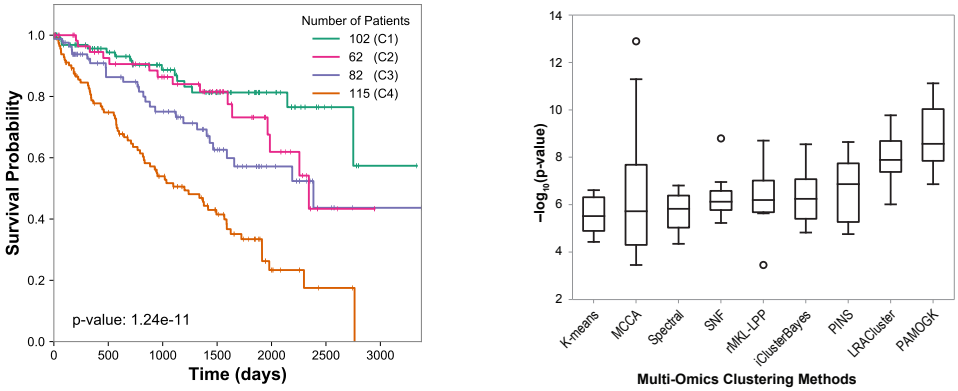
Additionally, we compare performances of the kernels based on the lowest p-value attained in the log-rank test on the survival distributions of clusters. In each experiment, each kernel is used with MKKM-MR method, and they are allowed to choose from a set of predetermined values for each of the hyperparameters. These include k for clustering, the smoothing parameter α for SmSPK, λ for MKKM-MR. The best clustering solution obtained for each method is compared in Figure 3a. We observe that SmSPK outperforms other graph kernels compared. This can be explained based on the previous remark that this graph kernel is formulated to distinguish graphs with different topologies. Additionally, although the use of RBF kernel generally yields good results, the integration of pathway information through SmSPK brings an improvement to the cluster separations in terms of survival.

3.3 Deciding on the Multi-view Kernel Clustering Algorithm to Use in PAMOGK

To determine the multi-view kernel clustering algorithm to be used in PAMOGK, we experiment with different alternatives. The multi-view kernel clustering methods that we analyze include the MKKM-MR [21], AKKM, LMKKM [10] and SNF [55] with KKM(Kernel K-Means) and spectral clustering (see Section 2.4. For each method, we report the best clustering solution, which is determined based on the lowest p-value attained in the log-rank test on the survival distributions of clusters. In each experiment, we allow the methods to choose from a set of predetermined values for each of the hyperparameters. These include k for clustering, the smoothing parameter α for SmSPK, λ for MKKM-MR.

Figure 3b summarizes the results in these experiments for the best clustering solution, where $k = 4$. When comparing the three multi-view kernel clustering methods, we observe that MKKM-MR produces the best results. LMKKM, AKKM and SNF based methods yield similar results with the difference that LMKKM performs slightly better than others. Overall, PAMOGK that uses the MKKM-MR multi-view clustering outperforms all the other clustering alternatives. Thus, we employ MKKM-MR in PAMOGK.

The best clustering solution by PAMOGK is obtained when $k = 4$, smoothing parameter, α is set to 0.3, and λ for MKKM-MR is set to 8. The KM plot of the resulting clustering is provided in Figure 4a. The survival distributions significantly differ (log-rank test, p-value $= 1.24e$-11). We should note that the solution with $k = 3$ is also quite good, p-value $= 8.13e$-11 (Supplementary Figure 4b).



(a) **(b)**

Fig. 4: a) Kaplan-Meier survival curves of the best clustering solution for KIRC. The p-value was obtained from a log-rank test between the groups **b)** Comparison of PAMOGK with the multi-omics clustering methods over 10 different trials. Each trial contains a random subsample of KIRC patients. The boxplot shows the -log (p-values) of the log-rank tests conducted on the survival distributions of these clusters. The higher the values, the better the clusters are separated in terms of survival distributions. (Note that PINS method results are over 9 experiments since in one of trial, it did not return a result.)

3.4 Effect of Other Design Choices

We checked some of the design choices. The first one is that the alternative strategies of how to map the expression changes on the network. We compare the following three alternatives. In the first one, we use the z-score as the node attribute. In the second one, we assigned the node attribute a binary value (0 or 1) depending on it is overexpressed or not, and similarly for the underexpressed case, underexpressed or not. The third one is the case; we assigned the z-scores of expression if the gene (or the protein) is over or underexpresssed to the degree of the change, and is 0 if it is not over or underexpressed. When we compare these three alternatives, we observe that the directly mapping z-scores do not yield good results, probably because when network smoothing is applied, everything accumulates change. The binary scoring yields also good results, but the third alternative yields better results. This comparison is shown in Supplementary Figure 2. We also assessed whether smoothing effects or not. Supplementary Figure 3 shows that smoothing helps, especially with mutation data, as these alterations are sparse compared to expression data.

3.5 Comparison with the State-of-the Art Multi-Omics Methods

Performance comparison: We compare PAMOGK with eight other multi-omics methods. These include k-means [22], MCCA [59], LRACluster [60], rMKL-LPP [45], iClusterBayes [25], PINS [31], SNF [55], and finally Spectral Clustering [63]. These methods cover all methods that are included in a recent comparative benchmark study by Rappoport et al. [34] with the exception of multiNMF [19], which we are not able to run properly. In running these algorithms, we set the maximum number of clusters to five and choose the other parameter configurations for each algorithm exactly as in the benchmark study [34].

To assess the performance of different methods, we repeatedly subsample the original patient set, and for each subsample, run the algorithms to find the patient clusters. Each subsample contained 300 patients. Due to prohibiting runtime of iClusterBayes, we were able to conduct this experiment 10 times. The distribution of log-rank test p-values attained by each method is displayed in Figure 4b. The comparison over ten runs shows that PAMOGK is the best performer among the nine methods. Not only the median performance is high, but even the 90-th percentile of the trials is superior to almost all methods. It also displays low variance across different runs. For all methods, for all trials, the resulting clusters are balanced in terms of the number of patients participating in the clusters except two trials of MCCA. The log-rank test is known to result in

unrealistically low p-values when one of cluster size is small [50]. In those two trials, MCCA's extremely low p-values are due to cluster sizes of 9 and 14.

Table 1: The runtimes in seconds for clustering 361 KIRC patients with the three types of omic data for different methods and PAMOGK.

Method	PAMOGK	LRACluster	PINS	SNF	rMKL-LPP	iClusterBayes	Spectral	K-means	MCCA
Time	352	289	56	7	109	10,898	3	47	6

Runtime comparisons: We conduct a runtime comparison of the algorithms for clustering all the KIRC patients using the three different data types. PAMOGK demands more time to run in comparison to the other methods, with the exception of iClusterBayes. This is because it calculates many more views of the data based on pathways. A second time limiting step is the weight optimization of the kernels in the MKKM-MR algorithm. Despite these additional requirements, the runtime is within reasonable limits, and a typical run takes less than 30 minutes without any parallelization. Replacing the multi-view clustering step with a less demanding algorithm and parallelization could reduce the runtime. Experiments are conducted on the following system configuration: CPU: Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40GHz CPU. Memory: 256Gb. Operating system: Ubuntu 16.04.4 LTS.

3.6 Detailed Analysis of KIRC Subgroups Discovered by PAMOGK

Table 2: Summary of statistical analyses of clinical variables.

Clinical Parameter	Test	p-value
Age	One-way ANOVA	$2.200e$-01
Gender	χ^2	$4.080e$-01
Stage	χ^2	$3.476e$-08
Primary Tumor Pathologic Spread	χ^2	$1.349e$-07
Distant Metastasis Pathologic Spread	χ^2	$3.766e$-04
Neoplasm Histologic Grade	χ^2	$2.104e$-09

KIRC Subgroups' Associations with Other Clinical Parameters: We analyze the association of clinical parameters of the discovered subgroups other than survival. The parameters include age, gender, tumor stage, primary tumor pathological spread, distant metastasis pathological spread and neoplasm histological grade. The associations of categorical variables are determined using χ^2 test while the continuous variables are tested with one-way ANOVA. We find no statistically significant difference in terms of age (p-value $= 0.220$) and gender (p-value $= 0.408$). All the other clinical parameters differ across groups at a statistically significant level (see Table 2). The distributions of these variables across groups are provided in Supplementary Section 3. The best prognosis group is cluster 1, and the worst prognosis group is cluster 4.

For all prognostic tumor-related features, cluster 1 always has more patients with a lower degree stage and grade, whereas cluster 4 always has more patients with a higher degree stage and grade. Overall, this analysis provides additional evidence that PAMOGK partitions KIRC patients into clinically meaningful subgroups The best prognosis group is cluster 1, and the worst prognosis group is cluster 4 (Figure 4a). There are clear differences between these two groups in terms of these additional clinical parameters. More specifically, 53.9% of the patients in cluster 1 are in stage I, whereas 67.8% of the patients in Cluster 4 are either in stage III or Stage IV. (Supplementary Table 6). Also, nearly half of the patients in cluster 1 have primary tumor T1, whereas 60% of the patients in cluster 4 have primary tumor T3 (see Supplementary Table 7). While only 8.82% of the patients of cluster 1 have distant metastasis, this ratio is 29.6% for cluster 4 patients. (Supplementary Table 8). Finally, the fraction of cluster 1 patients with histologic grade G1 and G2 is 61.7%, and those with G4 is 5.9%. For cluster 4, the percentage for G1 and G2 drops to 20.5% and G4 increases to 35.7%. (Supplementary Table 9). For all prognostic tumor-related features, cluster 1 always has more patients with a lower degree stage and grade, whereas cluster 4 always has more patients with a

higher degree stage and grade. Overall, this analysis provides additional evidence that PAMOGK partitions KIRC patients into clinically meaningful subgroups.

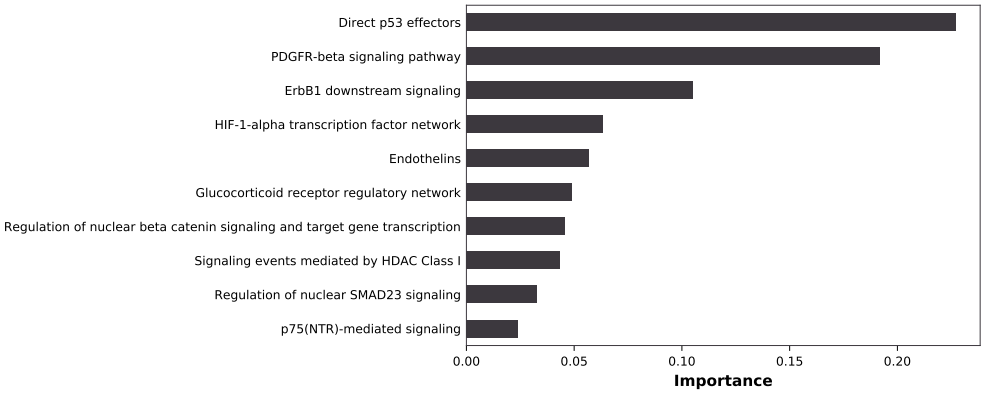


Fig. 5: The top 10 pathways, which have the highest relative importance in clustering.

Influential pathways and data types: By inspecting the assigned kernel weights, we can quantify the relative importance of pathways and molecular data types. For KIRC $(k = 4)$, the *direct p53 effectors* pathway and gene overexpression kernel emerge as the most important pathway-molecular alteration pair (see Supplementary Figure 5 for the top 10 pairs). By averaging the weights associated with each omic data type, we find that the gene expression is the top important data type, while protein expression data have almost no effect on the clustering (Supplementary Figure 6). This could be arising from the fact that the protein expression data covers only a small number of proteins.

The top relevant pathway in clustering the patients into subgroups emerges as the *the direct p53 effectors pathway* (Figure 5). p53 is a tumor suppressor transcription factor that regulates the cell division to prevent uncontrolled growth of cells [53]. Similarly, the second and the third pathways, *PDGFR-beta signaling* and *ErbB1 downstream signaling pathway* (better known as EGFR) are critical signaling pathways [44]. Additionally, Hypoxia-inducible factors (HIFs) regulates the expressions of many genes that are related to tumorigenesis [1] and [40] shows that HIF1α is a target of 14q loss, which is commonly associated with poor prognosis in kidney cancer. The firth pathway is Endothelin pathway and earlier results report that Endotelin-1 promotes cell survival in renal cell carcinoma [32].

4 Conclusion and Future Work

We present PAMOGK for discovering subgroups of patients, which not only operates by integrating different omics data sets derived from patients but also incorporates existing knowledge on biological pathways. To corroborate these data sources, we develop a novel graph kernel that evaluates patient similarities based on their molecular alterations in the context of known pathways. We employ a multi-view kernel clustering technique to leverage views constructed by different molecular alteration types and pathways. Our results indicate that the suggested methodology, when applied to KIRC, results in patient clusters that differ significantly in their survival distributions and other clinical parameters. The proposed methodology also provides quantitative evidence for the decisive role of known driver pathways on the clustering process. We further show that PAMOGK performs better compared to the state-of-the-art multi-omics approaches.

One limitation of the current work is that we used the bulk expression results provided by the TCGA project. However, it is known that there could be a high level of intra-tumor heterogeneity [7], and the bulk tumor might include a diverse collection of cells harbouring distinct molecular signatures. Future work would be to adapt PAMOGK framework to single-cell measurements as they become available for large cohorts of patients.

The work can be extended in several directions. In this current work, we use omic datasets containing somatic mutations, gene and protein expression as they give more direct information on the alterations in the pathways. In the future work, one can map the copy number variations and methylation levels as well to the genes and also use them as additional views. Also, we use binary node labels to indicate if the gene for that patient is perturbed or not. PAMOGK can be easily extended to accept continuous node labels to incorporate the extent of the molecular alterations. Furthermore, in the present study, we ignore the direction information of the edges in the pathways. A kernel that explicitly accounts for edge directions can be more devised. In place or addition to the pathways, protein-protein interaction networks can be used to generate views, which will be explored in future work.

5 Acknowledgements

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant #117E140. Oznur Tastan thanks to the Science Academy of Turkey under The Young Scientist Award Program (BAGEP). Yasin Tepeli and Ali Burak Unal acknowledge TUBITAK-BIDEB for the 2210-A scholarship program.

References

1. Banumathy, G., Cairns, P.: Signaling pathways in renal cell carcinoma. Cancer Biology & Therapy **10**(7), 658–664 (Oct 2010). https://doi.org/10.4161/cbt.10.7.13247, https://doi.org/10.4161/cbt.10.7.13247
2. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: Data Mining, Fifth IEEE International Conference on. pp. 8–pp. IEEE (2005)
3. Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K.: Multi-view clustering via canonical correlation analysis. In: Proceedings of the 26th annual International Conference on Machine Learning. pp. 129–136. ACM (2009)
4. Chikhi, N.F.: Multi-view clustering via spectral partitioning and local refinement. Information Processing & Management **52**(4), 618–627 (Jul 2016). https://doi.org/10.1016/j.ipm.2015.12.007, https://doi.org/10.1016/j.ipm.2015.12.007
5. Cowen, L., Ideker, T., Raphael, B.J., Sharan, R.: Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics **18**(9), 551 (2017)
6. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature **486**(7403), 346 (2012)
7. Dagogo-Jack, I., Shaw, A.T.: Tumour heterogeneity and resistance to cancer therapies. Nature reviews Clinical oncology **15**(2), 81 (2018)
8. Feragen, A., Kasenburg, N., Petersen, J., de Bruijne, M., Borgwardt, K.: Scalable kernels for graphs with continuous attributes. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 26, pp. 216–224. Curran Associates, Inc. (2013), http://papers.nips.cc/paper/5155-scalable-kernels-for-graphs-with-continuous-attributes.pdf
9. Gabasova, E., Reid, J., Wernisch, L.: Clusternomics: Integrative context-dependent clustering for heterogeneous datasets. PLoS computational biology **13**(10), e1005781 (2017)
10. Gönen, M., Margolin, A.A.: Localized data fusion for kernel k-means clustering with application to cancer biology. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1. pp. 1305–1313. NIPS'14, MIT Press, Cambridge, MA, USA (2014), http://dl.acm.org/citation.cfm?id=2968826.2968972
11. H., H.: Relations between two sets of variables. Biometrika pp. 321–327 (1936)
12. Harrington, D.P., Fleming, T.R.: A class of rank test procedures for censored survival data. Biometrika **69**(3), 553–566 (1982). https://doi.org/10.1093/biomet/69.3.553, https://doi.org/10.1093/biomet/69.3.553
13. Hayes, D.N., Monti, S., Parmigiani, G., Gilks, C.B., Naoki, K., Bhattacharjee, A., Socinski, M.A., Perou, C., Meyerson, M.: Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. Journal of Clinical Oncology **24**(31), 5079–5090 (2006)
14. Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V., et al.: Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell **158**(4), 929–944 (2014)
15. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American Statistical Association **53**(282), 457–481 (Jun 1958). https://doi.org/10.1080/01621459.1958.10501452, https://doi.org/10.1080/01621459.1958.10501452

16. Kumar, A., Rai, P., Daumé, III, H.: Co-regularized multi-view spectral clustering. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. pp. 1413–1421. NIPS'11, Curran Associates Inc., USA (2011), http://dl.acm.org/citation.cfm?id=2986459.2986617
17. Liang, R., Wang, M., Zheng, G., Zhu, H., Zhi, Y., Sun, Z.: A comprehensive analysis of prognosis prediction models based on pathway-level, gene-level and clinical information for glioblastoma. International Journal of Molecular Medicine (Jul 2018). https://doi.org/10.3892/ijmm.2018.3765, https://doi.org/10.3892/ijmm.2018.3765
18. Lin, Y.Y., Liu, T.L., Fuh, C.S.: Multiple kernel learning for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(6), 1147–1160 (2010)
19. Liu, J., Wang, C., Gao, J., Han, J.: Multi-view clustering via joint nonnegative matrix factorization. In: Proceedings of the 2013 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics (May 2013). https://doi.org/10.1137/1.9781611972832.28, https://doi.org/10.1137/1.9781611972832.28
20. Liu, J., et al.: An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell **173**(2), 400–416.e11 (Apr 2018). https://doi.org/10.1016/j.cell.2018.02.052, https://doi.org/10.1016/j.cell.2018.02.052
21. Liu, X., Dou, Y., Yin, J., Wang, L., Zhu, E.: Multiple kernel k-means clustering with matrix-induced regularization. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. pp. 1888–1894. AAAI'16, AAAI Press (2016), http://dl.acm.org/citation.cfm?id=3016100.3016163
22. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2), 129–137 (Mar 1982). https://doi.org/10.1109/tit.1982.1056489, https://doi.org/10.1109/tit.1982.1056489
23. Lock, E.F., Hoadley, K.A., Marron, J.S., Nobel, A.B.: Joint and individual variation explained (jive) for integrated analysis of multiple data types. The annals of applied statistics **7**(1), 523 (2013)
24. Meng, C., Kuster, B., Culhane, A.C., Gholami, A.M.: A multivariate approach to the integration of multi-omics datasets. BMC bioinformatics **15**(1), 162 (2014)
25. Mo, Q., Shen, R., Guo, C., Vannucci, M., Chan, K.S., Hilsenbeck, S.G.: A fully bayesian latent variable model for integrative clustering analysis of multi-type omics data. Biostatistics **19**(1), 71–86 (May 2017). https://doi.org/10.1093/biostatistics/kxx017, https://doi.org/10.1093/biostatistics/kxx017
26. Mo, Q., Wang, S., Seshan, V.E., Olshen, A.B., Schultz, N., Sander, C., Powers, R.S., Ladanyi, M., Shen, R.: Pattern discovery and cancer gene identification in integrated cancer genomic data. Proceedings of the National Academy of Sciences **110**(11), 4245–4250 (2013)
27. Monti, S., Tamayo, P., Mesirov, J., Golub, T.: Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine learning **52**(1-2), 91–118 (2003)
28. Müller, M.F., Ibrahim, A.E., Arends, M.J.: Molecular pathological classification of colorectal cancer. Virchows Archiv **469**(2), 125–134 (2016)
29. Network, T.C.G.A.R., et al.: Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature **499**(7456), 43–49 (Jun 2013). https://doi.org/10.1038/nature12222, https://doi.org/10.1038/nature12222
30. Neumann, M., Garnett, R., Bauckhage, C., Kersting, K.: Propagation kernels: Efficient graph kernels from propagated information. Mach. Learn. **102**(2), 209–245 (Feb 2016). https://doi.org/10.1007/s10994-015-5517-9, http://dx.doi.org/10.1007/s10994-015-5517-9
31. Nguyen, T., Tagett, R., Diaz, D., Draghici, S.: A novel approach for data integration and disease subtyping. Genome research **27**(12), 2025–2039 (2017)
32. Pflug, B.R., Zheng, H., Udan, M.S., D'Antonio, J.M., Marshall, F.F., Brooks, J.D., Nelson, J.B.: Endothelin-1 promotes cell survival in renal cell carcinoma through the eta receptor. Cancer letters **246**(1-2), 139–148 (2007)
33. Rappoport, N., Shamir, R.: Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic acids research **46**(20), 10546–10562 (2018)
34. Rappoport, N., Shamir, R.: Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research **46**(20), 10546–10562 (10 2018). https://doi.org/10.1093/nar/gky889, https://doi.org/10.1093/nar/gky889
35. Ricketts, C.J., et al.: The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. Cell Reports **23**(1), 313–326.e5 (Apr 2018). https://doi.org/10.1016/j.celrep.2018.03.075, https://doi.org/10.1016/j.celrep.2018.03.075
36. Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., Buetow, K.H.: PID: the pathway interaction database. Nucleic Acids Research **37**(suppl_1), D674–D679 (Oct 2008). https://doi.org/10.1093/nar/gkn653, https://doi.org/10.1093/nar/gkn653
37. Schölkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation **10**(5), 1299–1319 (Jul 1998). https://doi.org/10.1162/089976698300017467, https://doi.org/10.1162/089976698300017467
38. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA (2001)
39. Sejdinovic, D., Gretton, A., Bergsma, W.: A kernel test for three-variable interactions (2013)

40. Shen, C., Beroukhim, R., Schumacher, S.E., Zhou, J., Chang, M., Signoretti, S., Kaelin, W.G.: Genetic and functional studies implicate HIF1 as a 14q kidney cancer suppressor gene. Cancer Discovery **1**(3), 222–235 (Jun 2011). https://doi.org/10.1158/2159-8290.cd-11-0098, https://doi.org/10.1158/2159-8290.cd-11-0098
41. Shen, R., Olshen, A.B., Ladanyi, M.: Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics **25**(22), 2906–2912 (09 2009). https://doi.org/10.1093/bioinformatics/btp543, https://doi.org/10.1093/bioinformatics/btp543
42. Shervashidze, N., Schweitzer, P., Leeuwen, E.J.v., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-lehman graph kernels. Journal of Machine Learning Research **12**(Sep), 2539–2561 (2011)
43. Siglidis, G., Nikolentzos, G., Limnios, S., Giatsidis, C., Skianis, K., Vazirgiannis, M.: Grakel: A graph kernel library in python. arXiv preprint arXiv:1806.02193 (2018)
44. Smith, K., Gunaratnam, L., Morley, M., Franovic, A., Mekhail, K., Lee, S.: Silencing of epidermal growth factor receptor suppresses hypoxia-inducible factor-2–DrivenVHL-/-renal cancer. Cancer Research **65**(12), 5221–5230 (Jun 2005). https://doi.org/10.1158/0008-5472.can-05-0169, https://doi.org/10.1158/0008-5472.can-05-0169
45. Speicher, N.K., Pfeifer, N.: Integrating different data types by regularized unsupervised multiple kernel learning with application to cancer subtype discovery. Bioinformatics **31**(12), i268–i275 (2015)
46. Togninalli, M., Ghisu, E., Llinares-López, F., Rieck, B., Borgwardt, K.: Wasserstein weisfeiler-lehman graph kernels. In: Advances in Neural Information Processing Systems 32, pp. 6439–6449. Curran Associates, Inc. (2019), http://papers.nips.cc/paper/8872-wasserstein-weisfeiler-lehman-graph-kernels.pdf
47. Togninalli, M., Ghisu, E., Llinares-López, F., Rieck, B., Borgwardt, K.: Wasserstein weisfeiler–lehman graph kernels. In: Advances in Neural Information Processing Systems 32 (2019)
48. Toss, A., Cristofanilli, M.: Molecular characterization and targeted therapeutic approaches in breast cancer. Breast Cancer Research **17**(1), 60 (2015)
49. Ünal, A.B., Tastan, O.: Identification of cancer patient subgroups via smoothed shortest path graph kernel. In: NeurIPS 2016 Workshop on Machine Learning in Computational Biology. vol. abs/1612.04431 (2016), http://arxiv.org/abs/1612.04431
50. Vandin, F., Papoutsaki, A., Raphael, B.J., Upfal, E.: Accurate computation of survival statistics in genome-wide studies. PLOS Computational Biology **11**(5), e1004071 (May 2015). https://doi.org/10.1371/journal.pcbi.1004071, https://doi.org/10.1371/journal.pcbi.1004071
51. Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B.A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., Feiler, H.S., Hodgson, J.G., James, C.D., Sarkaria, J.N., Brennan, C., Kahn, A., Spellman, P.T., Wilson, R.K., Speed, T.P., Gray, J.W., Meyerson, M., Getz, G., Perou, C.M., Hayes, D.N.: Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell **17**(1), 98–110 (Jan 2010). https://doi.org/10.1016/j.ccr.2009.12.020, https://doi.org/10.1016/j.ccr.2009.12.020
52. Vishwanathan, S.V.N., Borgwardt, K.M., Kondor, I.R., Schraudolph, N.N.: Graph kernels. CoRR **abs/0807.0093** (2008), http://arxiv.org/abs/0807.0093
53. Vogelstein, B., Lane, D., Levine, A.J.: Surfing the p53 network. Nature **408**(6810), 307–310 (Nov 2000). https://doi.org/10.1038/35042675, https://doi.org/10.1038/35042675
54. Von Luxburg, U.: A tutorial on spectral clustering. Statistics and computing **17**(4), 395–416 (2007)
55. Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., Goldenberg, A.: Similarity network fusion for aggregating data types on a genomic scale. Nature methods **11**(3), 333 (2014)
56. Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., Goldenberg, A.: Similarity network fusion for aggregating data types on a genomic scale. Nature Methods **11**(3), 333–337 (Jan 2014). https://doi.org/10.1038/nmeth.2810, https://doi.org/10.1038/nmeth.2810
57. Weinstein, J.N., , Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature Genetics **45**(10), 1113–1120 (Sep 2013). https://doi.org/10.1038/ng.2764, https://doi.org/10.1038/ng.2764
58. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Network, C.G.A.R., et al.: The cancer genome atlas pan-cancer analysis project. Nature genetics **45**(10), 1113 (2013)
59. Witten, D.M., Tibshirani, R.J.: Extensions of sparse canonical correlation analysis with applications to genomic data. Statistical Applications in Genetics and Molecular Biology **8**(1), 1–27 (Jan 2009). https://doi.org/10.2202/1544-6115.1470, https://doi.org/10.2202/1544-6115.1470
60. Wu, D., Wang, D., Zhang, M.Q., Gu, J.: Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification. BMC Genomics **16**(1) (Dec 2015). https://doi.org/10.1186/s12864-015-2223-8, https://doi.org/10.1186/s12864-015-2223-8
61. Yu, S., Tranchevent, L., Liu, X., Glanzel, W., Suykens, J.A., De Moor, B., Moreau, Y.: Optimized data fusion for kernel k-means clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(5), 1031–1039 (2011)

62. Zhao, J., Xijiong, X., Xu, X., Sun, S.: Multi-view learning overview: Recent progress and new challenges. Information Fusion **38** (02 2017). https://doi.org/10.1016/j.inffus.2017.02.007
63. Zhou, D., Burges, C.J.: Spectral clustering and transductive learning with multiple views. In: Proceedings of the 24th international conference on Machine learning. pp. 1159–1166. ACM (2007)