BMC Bioinformatics

Open Access

CrossMark

Automated prostate tissue referencing for cancer detection and diagnosis

Jin Tae Kwak[1], Stephen M. Hewitt[2], André Alexander Kajdacsy-Balla[3], Saurabh Sinha[4*] and Rohit Bhargava[5*]

Abstract

Background: The current practice of histopathology review is limited in speed and accuracy. The current diagnostic paradigm does not fully describe the complex and complicated patterns of cancer. To address these needs, we develop an automated and objective system that facilitates a comprehensive and easy information management and decision-making. We also develop a tissue similarity measure scheme to broaden our understanding of tissue characteristics.

Results: The system includes a database of previously evaluated prostate tissue images, clinical information and a tissue retrieval process. In the system, a tissue is characterized by its morphology. The retrieval process seeks to find the closest matching cases with the tissue of interest. Moreover, we define 9 morphologic criteria by which a pathologist arrives at a histomorphologic diagnosis. Based on the 9 criteria, true tissue similarity is determined and serves as the gold standard of tissue retrieval. Here, we found a minimum of 4 and 3 matching cases, out of 5, for ~80 % and ~60 % of the queries when a match was defined as the tissue similarity score ≥5 and ≥6, respectively. We were also able to examine the relationship between tissues beyond the Gleason grading system due to the tissue similarity scoring system.

Conclusions: Providing the closest matching cases and their clinical information with pathologists will help to conduct consistent and reliable diagnoses. Thus, we expect the system to facilitate quality maintenance and quality improvement of cancer pathology.

Keywords: Prostate cancer, Database, Tissue morphology, Tissue retrieval, Infrared imaging, Decision support

Background

Quality assurance in diagnostic histopathology plays a critical role in development of a treatment plan for patients with prostate cancer [1]. Methods to integrate quality development, maintenance, and improvement of diagnostic accuracy are, hence, critical to cancer management in any setting. In diagnostic prostate pathology, Gleason grading [2] is the most commonly used grading system that is based upon the structural patterns of the tumor. The Gleason grade is a primary determinant in

* Correspondence: sinhas@illinois.edu; rxb@illinois.edu
[4]Department of Computer Science, University of Illinois at Urbana-Champaign, 2122 Siebel Center, 201 N. Goodwin Avenue, Urbana, IL 61801, USA
[5]Beckman Institute for Advanced Science and Technology, Department of Bioengineering, Department of Mechanical Science and Engineering, Electrical and Computer Engineering, Chemical and Biomolecular Engineering and University of Illinois Cancer Center, University of Illinois at Urbana-Champaign, 4265 Beckman Institute 405 N. Mathews Avenue, Urbana, IL 61801, USA
Full list of author information is available at the end of the article

treatment planning [3]. However, it is well known that the grading of prostate tissues suffers from intra- and inter-pathologist variability [4–6]; for example, the exact intra-pathologist agreement was achieved in 43–78 % of the instances, and 36–81 % of the exact inter-pathologist agreement was reported. It is also known that the variability of the grading can be reduced with focused retraining. There could be many ways to educate pathologists such as meetings, courses, online tutorials, and etc [7], but these are not time- and cost-effective and rarely implemented. Therefore, building an automated, fast, and objective method to aid pathologists in evaluating prostate can improve prostate cancer diagnosis.

When a pathologist evaluates a tissue sample, he/she looks at a stained tissue and mentally compares it against a fund of knowledge and experience and may consult publications when needed. In essence, the pathologist is matching structural patterns with samples they have seen earlier and mentally recalling the diagnosis

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 2 of 12

made such that they can make the same diagnosis in the specific test case. Despite training, intra- and inter-observer variation and controversial areas still exist [8]. To aid and improve the diagnostic process, there have been several research efforts to develop automated systems for the detection and grading of prostate cancer. The majority of the previous methods have used morphological features [9–16] to characterize and classify tissue samples into correct classes, and others have also used Fourier Transform [17], Wavelet Transform [13, 18, 19], and Fractal Analysis [13, 20] to extract texture features. Though these methods claim to be accurate, the information that pathologists will obtain by using such methods may be limited since these only provide the predicted grade in general. The prediction also relies on the conditions of the training and testing datasets such as acquisition settings [15, 19] and staining [21].

Alternatively, content-based image retrieval (CBIR) systems [22–24] have been proposed to aid cancer pathology. The main objective is to effectively and efficiently manage an enormous amount of image data and to provide similar cases to a new test case that is examined. In addition to clinical usage, CBIR systems can help medical research, education, and training [22, 24]. The similar cases can be determined as owning the same grade [25–28] or sub-structures [29, 30]; for instance, single lumen glands, multi-lumen glands blood vessels, and lymphocytes in prostate [31]. The basic premise of such systems in diagnostic histopathology is that tissue samples that have the same grade or similar characteristics and patterns with the sample of interest will afford useful information to pathologists and improve the decision-making process. Similar to cancer detection and grading systems, tissue is represented as several quantitative features such as morphology [26, 32, 33], histogram [30], color [28, 34], and texture [27–29, 32–35]. The similar samples can be retrieved by computing distance metrics or similarity scores between a new case and the previously diagnosed or examined cases. In order to improve tissue representation and retrieval, features are often post-transformed and/or their weights are adjusted in an implicit or explicit manner; for example, kernel function [30], simplex method [32], manifold learning [26, 36], boosting [25, 27], and self-organizing map (SOM) [35].

Previous retrieval systems have been measured against a gold standard of diagnostic category and grade of tumor, defined by a pathologist. Prostate cancer is, in particular, a multifactorial disease and a mixture of heterogeneous growth patterns [37], and hence tissues belonging to the same Gleason grade may possess different cellular, nuclear, or glandular sub-patterns. A number of histological variants, in fact, exist in prostate carcinoma

and some of the variants cannot be addressed by the Gleason grading system [38]. Moreover, the Gleason grading system results in a tumor grade that correlates with overall outcomes (survival), but fails to provide information on risk of metastasis, and correlates poorly with the clinical decision making process. Further, the Gleason grading system has gone through several refinements over time [8, 39–41] and may undergo further changes [42, 43]. These changes result variations among pathologists in practice [7] and disrupt developing robust automated grading and retrieval systems.

Here, we report developing a new computer information and management and decision-support system that consists of a database of pre-defined prostate tissues and a retrieval process (Fig. 1). The database retains tissue images, clinical information, and one or more measurements of the structure of tissue. The retrieval process utilizes the structural/morphological features of the tissue sample image and provides tissue samples similar to the sample under consideration from the database. In assessing tissue morphology, we utilize infrared (IR) chemical imaging for providing cell type information in tissue [44]. We previously reported its utility in stabilizing and improving the automated cancer detection [45]. As a retrieval function, we adopt a machine learning ranking approach, called Ranking-support vector machine (Ranking-SVM) [46] in conjunction with a two-stage "feature selection" step [47]. Ranking-SVM learns a ranking function of high generalization due to maximum-margin property [48]. Feature selection step finds the most informative subsets while preserving the essential characteristics of the data. Moreover, we propose to determine the ground truth tissue similarity based on various structural properties of tissue, not solely on Gleason grade or a single structural component. Here, the structural properties are examined by pathologists. Combining different structural components of tissue ensures better characterization of tissue structure, and thus more accurate measurement of tissue similarity can be made. Thereby, the automated and computerized analysis and human experts' assessment of tissue morphology are correlated through a machine learning approach.

The rest of the paper is organized as follows. In Methods section, we begin with a description of the dataset and data preparation process. In the following subsections, we describe the three key components of our new system – tissue similarity measure, tissue morphological feature extraction, and tissue retrieval function. Then, feature selection and balanced training are described. In Results section, the experimental results, including tissue similarity measure and tissue retrieval, via cross-validation are demonstrated. In Discussion section,

Kwak et al. BMC Bioinformatics (2016) 17:227

Page 3 of 12
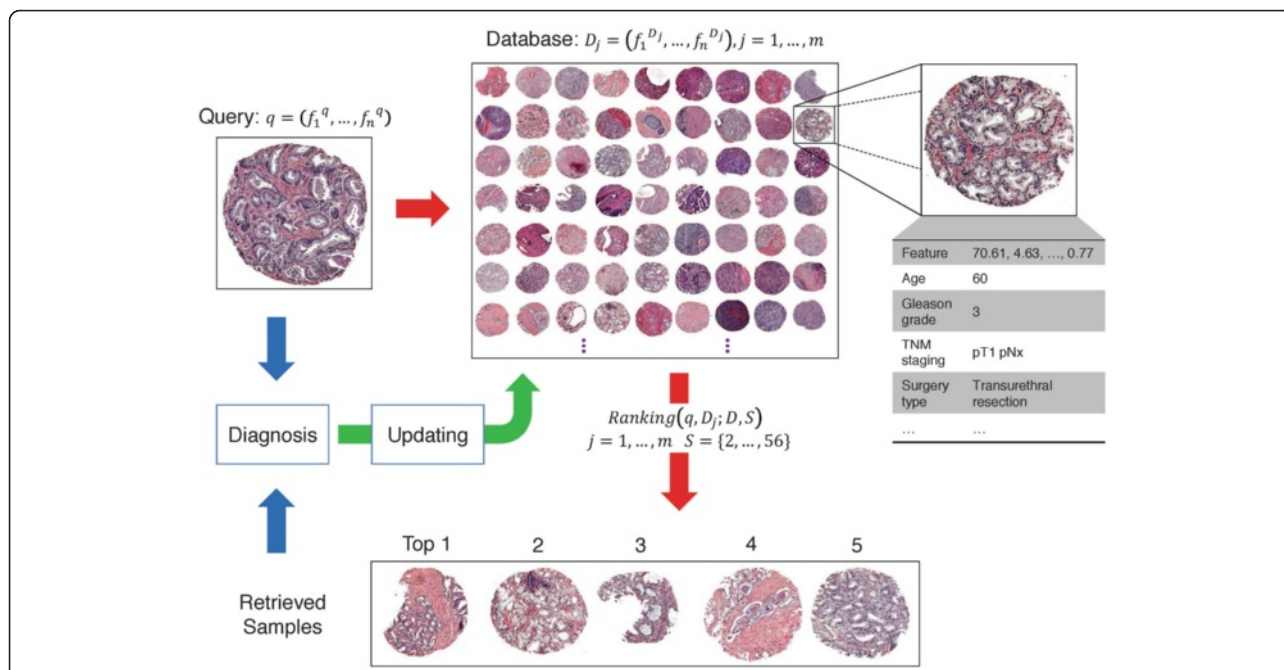
Fig. 1 Overview of System. As a query is given, the closest matches and their clinical information are retrieved from the database (*red arrows*). Provided with the matching cases, pathologists make a diagnosis (*blue arrows*), and updating may or may not be conducted (*yellow arrow*). Q, D, Ranking, f, and S denote a query, database, retrieval process, single feature, and subset of features, respectively

the implications and limitations of our study are discussed. Finally, we conclude in Conclusions section.

Methods

Samples and data preparation

This study and protocols were approved by the University of Illinois Institutional Review Board (IRB) and was conducted as per the permission of the IRB in accordance with relevant guidelines and regulations. We have obtained 114 prostate cancer tissue samples (Tissue Array Research Program, National Cancer Institute and Clinomics Inc.), composed of 19 (Gleason 6), 26 (Gleason 7; 16 Gleason 3 + 4, 10 Gleason 4 + 3), 22 (Gleason 8), 10 (Gleason 9; 1 Gleason 4 + 5, 9 Gleason 5 + 4), and 37 (Gleason 10) samples. Both hematoxylin and eosin (H&E) stained and FT-IR images are available for the samples. Tissue samples were first sectioned to ~5um thick sections, with a section being placed on a standard glass slide and a serial section on IR transparent BaF_2 slide. Stained with H&E, tissue images were acquired on a standard optical microscope at 40x magnification, and the size of a pixel is 0.963um × 0.963um. On IR transparent BaF_2 slides, FT-IR images were acquired at a spatial pixel size of 6.25um × 6.25um and a spectral resolution of 4 cm^{-1} at an undersampling ratio of 2 using Perkin-Elmer Spotlight imaging system. The spectral profile of a pixel was truncated to a spectral range of 4000-720 cm^{-1}. Detailed description of sample preparation and data acquisition for FT-IR imaging are available in Fernandez et al.

[49]. Clinical information (Gleason grade, age, surgery type, etc.) of the samples were prepared by pathologic review, and 308 morphological features were also extracted. The database we build here, therefore, contains 114 tissue images (of two different modalities) and their clinical information and 308 morphological features.

Morphologic criteria and tissue similarity measure

We define 9 criteria to describe the architectural properties of tissue: 1) Gland crowding, 2) Gland roundness, 3) Stromal reaction, 4) Nuclear grade, 5) Clefts 6) Lumen/gland ratio, 7) Gland continuity, 8) Cell separation, and 9) Gleason score. The details of the criteria are listed in Table 1. Some of the properties are the criteria used in the Gleason grading system, and others were adopted from the literature. Although some criteria are overlapped with the Gleason grading system, their importance and interpretation in our system may vary. The Gleason grading system may be also able to describe certain properties of tissues that cannot be characterized by the overlapping criteria. In the Gleason grading system, gland arrangement (Gland crowding), variations in size and shape of gland (Gland roundness), sheets of cells (Gland continuity), and single cells (Cell separation) are examined. Nuclear morphometry (Nuclear grade) [50–52], reactive stroma (Stromal reaction) [53–55], and retraction clefting (Clefts) [56] have been reported to be useful for prostate diagnosis and prognosis. In the digital and computerized tissue analyses,

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 4 of 12

Table 1 Description of 9 Morphologic criteria

Criteria	Description	Score			
		0	1	2	3
Gland crowding	Gland tightness and cohesiveness	N/A	Sparse	Moderate	Very tight
Gland roundness	Roundness of external perimeter of gland	N/A	Very round	Moderate	Serrated contours or spindle shaped contours
Stromal reaction	Swollen, plump cells in stroma and splayed collagen fibers	N/A	No reaction	Little	-
Nuclear grade	Prominent nucleoli, variation in nuclear diameter and amount of chromatin	N/A	Normal	Some prominent nucleoli, moderate variation	Many prominent nucleoli, large variation
Clefts	Cleft formation or retraction artifact around cancer glands	N/A	<30 %	≥30 %	-
Lumen/gland ratio	Ratio between lumen area and total gland area	N/A	Wide lumen	Moderate lumen	Tiny lumen
Gland continuity	Continuous sheets of cells	N/A	<30 %	≥30 %	-
Cell separation	Individual cells separated by stroma	N/A	<10 %	≥10 %	-
Gleason Score	Predominant and secondary Gleason score	6 – 10[a]			

[a]Gleason score is the sum of predominant and secondary scores. In our set, it ranges from 6 to 10

structural features describing gland arrangement [11, 36] and shape [11, 19, 36, 45] and the size of gland and lumen (Lumen/gland ratio) [12, 45, 57] have been adopted to characterize tissue. Individual cells also showed a moderate correlation with patient outcomes [58].

For each of the criteria, a pathologist examines a tissue sample (H&E image) and assigns a score ranging from 0 to 2 (Stromal reaction, Clefts, Gland continuity, and Cell separation) or 0 to 3 (Gland crowding, Gland roundness, Nuclear grade, and Lumen/gland ratio) except Gleason score which in our set of tissues ranges from 6 to 10. The score range from 0 to 2 may be interpreted as none, low, and high, and the range from 0 to 3 may be considered as none, low, mid, and high. Due to its qualitative nature, it is difficult to highly stratify, and the impact and measurability of each criterion varies. Restricting the score range to none, low, mid, and high (or none, low, and high), in general, the scores are intended to be specific to differing morphologic patterns as well as to be reproducible by other pathologists. Using the scores of the 9 morphologic criteria, tissue morphologic similarity (*TMS*) between tissue samples is measured. Although well-defined and measured, the importance or relevance of each criterion differs. For example, the significance of Gland crowding score 1 may differ from that of Gland roundness score 1, and the difference between two samples having Gland crowding score 1 and 2 may not be identical to the difference between two samples owning Stromal reaction score 1 and 2. In these cases, the absolute values of the scores and the difference of the scores are identical. Recognizing the intrinsic relationship between scores and criteria, we utilize the distribution of each criterion score over the samples in the

database. Regardless of the absolute value of a score, two samples away from each other in the distribution of the scores of a criterion are likely dissimilar in terms of the criterion, and vice versa. In other words, tissue similarity between two samples with respect to a morphologic criterion is related to the number of samples between them as ordered by the score for the criterion. Accordingly, let $TMS(d_1, d_2)$ be the tissue morphologic similarity between two tissue samples d_1 and d_2 and computed as follows:

$$TMS(d_1, d_2) = \sum_{i=1}^{9} TMS^i(d_1, d_2)$$

where $TMS^i(d_1, d_2)$ is the morphologic tissue similarity for ith criteria. $TMS^i(d_1, d_2)$ is calculated as follows:

$$TMS^i(d_1, d_2) = 1 - \frac{\sum_{s=s_{d_1}^i+1}^{s_{d_2}^i-1} h^i(s) + \frac{1}{2}\left(h^i\left(s_{d_1}^i\right) + h^i\left(s_{d_2}^i\right)\right)}{Z}$$

where s_d^i is the ith morphologic criterion score of a tissue sample d, $h^i(s)$ is the number of samples having ith morphologic criterion score s, and Z is a normalization factor. Due to normalization, $TMS^i(d_1, d_2)$ ranges from 0 to 1, $1 \le i \le 9$, thereby $TMS(d_1, d_2)$ ranges from 0 to 9. In this study, *TMS* scores represent the true similarity between tissue samples and serve as the gold standard of tissue retrieval.

Morphological feature extraction

In prostate cancer, epithelial cells [59], which line ducts and acini in intact tissue in three-dimensional structures, are of great interest. As cancer grows, epithelial cells grow (or invade) in and out of the glands in an

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 5 of 12

uncontrolled way, and thus the structure of tissue, especially the local glandular structure, is distorted. We also note that the role of stroma cells, connective cells supporting epithelial cells, in cancer tissue has been recently recognized [53, 54]. To quantify the alterations in tissue morphology, we focus here on the nuclear and cellular morphology of epithelial and stromal cells and lumens (empty space inside a gland). In order to quantify the nuclear and cellular morphology of epithelial and stromal cells and lumens (Fig. 2a), we first segment epithelium and stroma in tissue by adopting Fourier transform infrared (FT-IR) spectroscopy imaging due to its accuracy and robustness [44]. FT-IR has been extensively validated in classifying histologic cell types in tissue [49, 60, 61] and provides a color coded cell type image of tissue. Cell type segmentation in H&E images is challenging due to limited information, color variations, etc. Rigid-body image registration overlays the epithelium and stroma segmentation from FT-IR imaging with the corresponding H&E image by using outer shape and empty space (lumens) in tissues [45]. Second, lumens and nuclei are identified from H&E images by considering their color intensities and geometric properties [45]. Using the segmented nuclei and lumens, we finally define a number of quantities measuring the morphologic changes in tissue, and the quantities include the size, number, distance, spatial distribution, and shape of epithelial nuclei and lumens (Fig. 2b). Detailed description of the quantities is available in Supplementary Information. In total, we defined 26 quantities, of which 17 quantities were previously shown to be effective in detecting prostate cancer tissue with high accuracy [45]. Computing average, standard deviation, sum total, minimum, and maximum of all or some of these quantities, 308 morphological features are extracted from a tissue
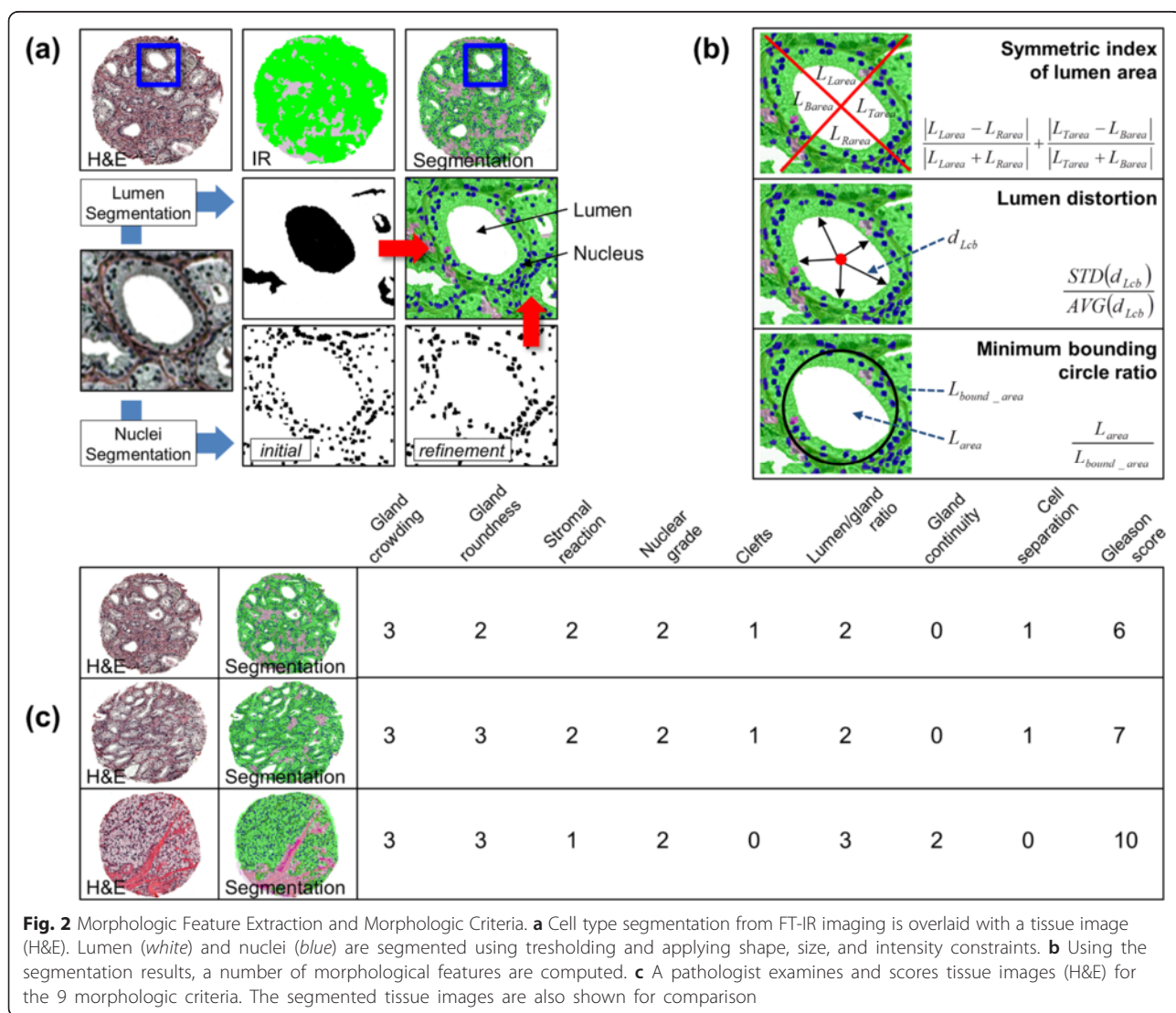
Fig. 2 Morphologic Feature Extraction and Morphologic Criteria. **a** Cell type segmentation from FT-IR imaging is overlaid with a tissue image (H&E). Lumen (*white*) and nuclei (*blue*) are segmented using tresholding and applying shape, size, and intensity constraints. **b** Using the segmentation results, a number of morphological features are computed. **c** A pathologist examines and scores tissue images (H&E) for the 9 morphologic criteria. The segmented tissue images are also shown for comparison

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 6 of 12

sample. The details of tissue segmentation and feature extraction process are described elsewhere in Kwak et al. [45].

Tissue retrieval

Given a query (unknown tissue sample image), its morphological features are extracted and used to search for similar pre-examined samples from the database. To retrieve the most similar samples, we adopt Ranking-SVM [46], which learns a function mapping onto a ranking in a pair-wise fashion (see Supplementary information for details). That is, Ranking-SVM provides a complete ranking of the entire samples in the database for the query. Since *TMS* score serves as the gold standard of the tissue similarity (or ranking), Ranking-SVM attempts to learn and reproduce the human experts' interpretation of the tissue samples. The feature vector difference between a query and the samples in the database is used for retrieval. We note that if a sample in the database is highly ranked to the query, then the query should be highly ranked for the sample (if we switch the highly ranked sample with the query). Ranking-SVM is an asymmetric measure, i.e., the ranking of a sample to the query would not be equal to the ranking of the query to the sample. Combining the two rankings, we seek to attain the more symmetric rankings between the query and the samples and to achieve the more accurate and specific retrieval (the samples that are similar to both the query and other samples in the database will be penalized, and the samples that are similar to the query and dissimilar to others will be boosted). We define the ranking of a sample to the query as

$$Ranking(q, d_i; D) = Ranking\text{-}SVM(q, d_i; D) + Ranking\text{-}SVM(d_i, q; D \ d_i \cup q), \ i = 1, ..., m$$

where $Ranking\text{-}SVM(q, d_i; D)$ denotes the ranking of the sample d_i in the database D to the query q and $Ranking\text{-}SVM(d_i, q; D \backslash d_i \cup q)$ is the ranking of the query q to the sample d_i in the database D when the query q is switched with the sample d_i. Based on the ranking, Top-T samples are retrieved. Since it is the sum of two rankings, it is likely that several rankings are tied. In such cases, the final ranking is determined by the ranking of the sample to the query, i.e., $Ranking\text{-}SVM(q, d_i; D)$, which is intuitive because the retrieval is done for the query.

Feature selection

Feature selection is the step where the retrieval algorithm examines all available features (308 in our case) with respect to the training samples, and selects a subset to use on test data. This selection is generally based on the criterion of high accuracy on training data, but also strives to ensure generalizability beyond the training data. We adopt a two-stage feature selection approach here. In the first stage, we order the features by their individual retrieval performance and sequentially measure the retrieval performance of a feature set by adding a new feature one at a time according to the order. In the second stage, feature selection continues with the feature set resulting the best retrieval performance in the first stage as the starting point, following the sequential floating forward selection (SFFS) method [62]. This method sequentially adds new features followed by conditional deletion(s) of already selected features.

Throughout the feature selection procedure, the retrieval capability of a feature set is measured by normalized discounted cumulative gain (NDCG) [63, 64], which is a popular measure to evaluate ranking algorithms with multiple levels of relevance. NDCG utilizes the relevance (*TMS* score in our study) and ranking of the retrieved samples and is based on two assumptions: 1) highly relevant samples are more valuable when they are retrieved earlier 2) highly relevant samples are more valuable than marginally relevant samples to the query. Given a database D and *TMS* scores, the performance of the retrieval function f for a query q at rank position T is computed as follows:

$$NDCG(q, f; D, TMS) = \frac{DCG}{IDCG}$$

$$DCG(q, f; D, TMS) = \sum_{t=1}^{T} \frac{2^{TMS(q, r_t)} - 1}{\log_2(1 + t)}$$

where r_t indicates the tth closest sample to the query q, retrieved by the retrieval function f, from the database D, and $IDCG$ denotes a normalization factor that is computed with the ideal (or optimal) rank of the retrieved samples, scaling the optimal retrieval to 1.

Balanced training

Ranking-SVM tries to learn an overall ranking of the training dataset. When trained on biased or unbalanced training dataset, Ranking-SVM may be biased towards dominant dataset, and thus its retrieval capability may be limited. To prevent this, we sought to take roughly balanced sub-samples of the training dataset and trained Ranking-SVM on the sub-samples. To obtain the roughly balanced training dataset, we first divide the total *TMS* score range into P equal-width partitions. Then, N_P number of pairs of samples from each partition was randomly selected. We set N_P to the smallest number of pairs of samples in a partition.

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 7 of 12

Results

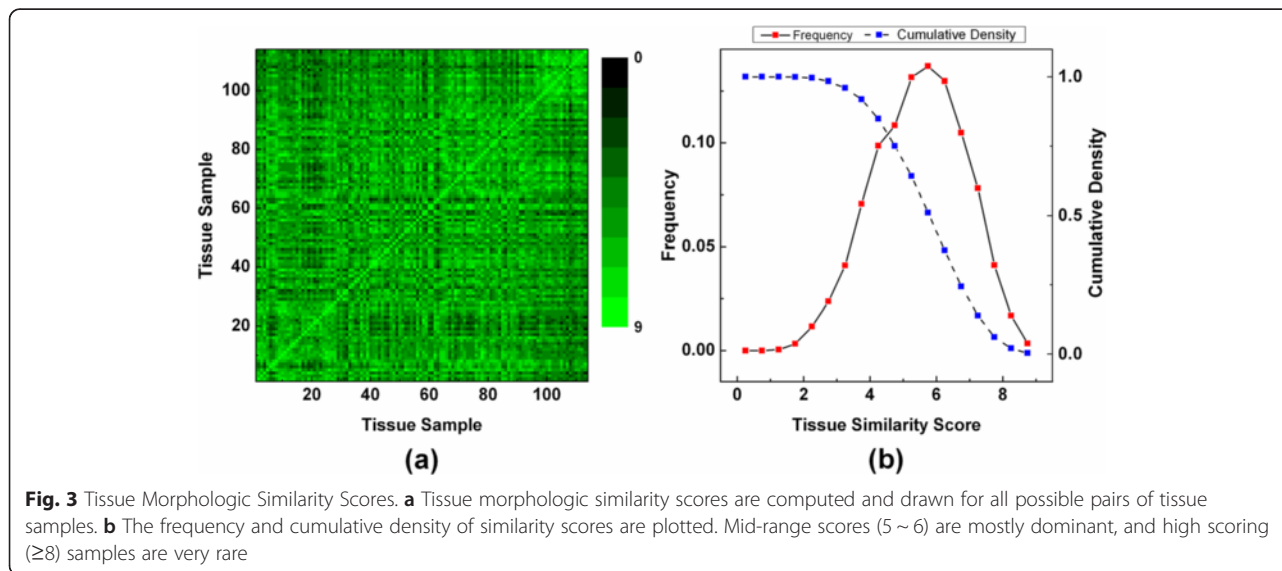### Tissue morphologic similarity measure

For 114 prostate cancer samples, we asked a pathologist (A.K.-B) to score them according to the 9 morphologic criteria. The pathologist was not involved in preparing the tissue samples and kept blind to the previous diagnosis and clinical information of the samples. Provided with the scores for the 9 morphologic criteria, tissue morphologic similarity (TMS) was measured for all possible pairs of 114 tissue samples (Fig. 2c and Fig. 3a) and used as the gold standard for training and validating our approach. We noted that TMS score, ranging from 0 to 9, is not evenly distributed, and mid-range score (5 ~ 6) is mostly dominant. Notably, only small number of pairs of samples gained a high TMS score, e.g., ~2 % of pairs of samples scores ≥8 (Fig. 3b).

Tissue retrieval system provides good matching cases

To evaluate the tissue retrieval system, we performed K-fold cross-validation ($K = 10$; maintaining a sufficient number of tissues in the database). The entire dataset was divided into K roughly equal-sized partitions, one partition was left out as "test data" (or queries), the union of the remaining $K - 1$ partitions (the "training data") was used to build the database where top-T similar samples are retrieved for each query ($T = 5$). This was repeated K times with different choices of the left-out partition. In each repetition, the 2-stage feature selection was carried out on the training data via a cross-validation (5-fold). The average NDCG at rank position T of the tissue retrievals for the queries, across all K repetitions, was computed to measure the performance of the retrieval. To handle the imbalance of TMS scores in

the dataset, a roughly balanced training dataset was formed by dividing the entire score range into P equal-width partitions ($P = 10$; allocating a sufficient number of tissues per partition in regard to the number of retrieved samples) and randomly taking equal number of samples from each partition. The method was implemented in IDL (tissue segmentation and morphological feature extraction) on 1 1.67GHz Intel Core Duo machine running Windows 7 with 2GB memory and C++ (feature selection and tissue retrieval) on a 2.5GHz Intel Core 2 Duo machine running Redhat Linux 4 with 2GB memory. The average processing time for tissue segmentation and morphological feature extraction is ~8 min per sample, and the tissue retrieval time is ~1 s. The Ranking-SVM training and the feature selection took ~3 s and ~90 min, respectively.

Although we have computed TMS scores and used them to train and test the retrieval process, it is unclear what similarity score is sufficient to provide useful information with pathologists when evaluating unknown samples. Setting a threshold TMS too high score is unrealistic because there are not enough samples available; as mentioned above, only ~2 % of the training samples have similarity score ≥8 for a query (Fig. 3b). Setting the TMS threshold lower is not beneficial to pathologists. We therefore adopted a new data management approach: In order to examine the retrieval performance in a broad sense, we changed a threshold similarity score th_s from 0 to 8, and designated a sample as a good match (or relevant sample) to a query if their similarity score is ≥ th_s. Then, we counted the number of good matches (N_G) among the retrieved samples for each query and plotted the fraction of the

Fig. 3 Tissue Morphologic Similarity Scores. **a** Tissue morphologic similarity scores are computed and drawn for all possible pairs of tissue samples. **b** The frequency and cumulative density of similarity scores are plotted. Mid-range scores (5 ~ 6) are mostly dominant, and high scoring (≥8) samples are very rare

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 8 of 12

queries retrieving $\geq N_G(N_G = 1, ..., T)$. N_G among the retrieved samples is equivalent to the fraction of the retrieved samples that are relevant to the query ("precision"). That is, Fig. 4a shows the fraction of the queries achieving a precision level equal to or higher than 0.2, 0.4, 0.6, 0.8, and 1. It is noticeable that ~80 % and ~60 % of the queries retrieving ≥4 and ≥3 good matching cases (or ≥0.8 and ≥0.6 precision) as setting th_s to 5 and 6, respectively. Compared to the random chance of retrieving ≥4 and ≥3 good matches, both were increased by two-fold, and the retrievals

were statistically significant (*p-value* <1.0e-10) by a binomial test (Table 2). As shown in Fig. 4b, it was obvious that *TMS* scores of pairs of the query and its top-*T* matching samples are higher than those of pairs of the query and all the samples in the database, especially *TMS* scores are 5 or greater.

Moreover, we performed the tissue retrieval by using the k-Nearest Neighbor (kNN) algorithm (k = 5), instead of Ranking-SVM. Examining the number of good matches, Ranking-SVM consistently outperformed kNN; for instance, setting th_s to 5 and 6, Ranking-SVM
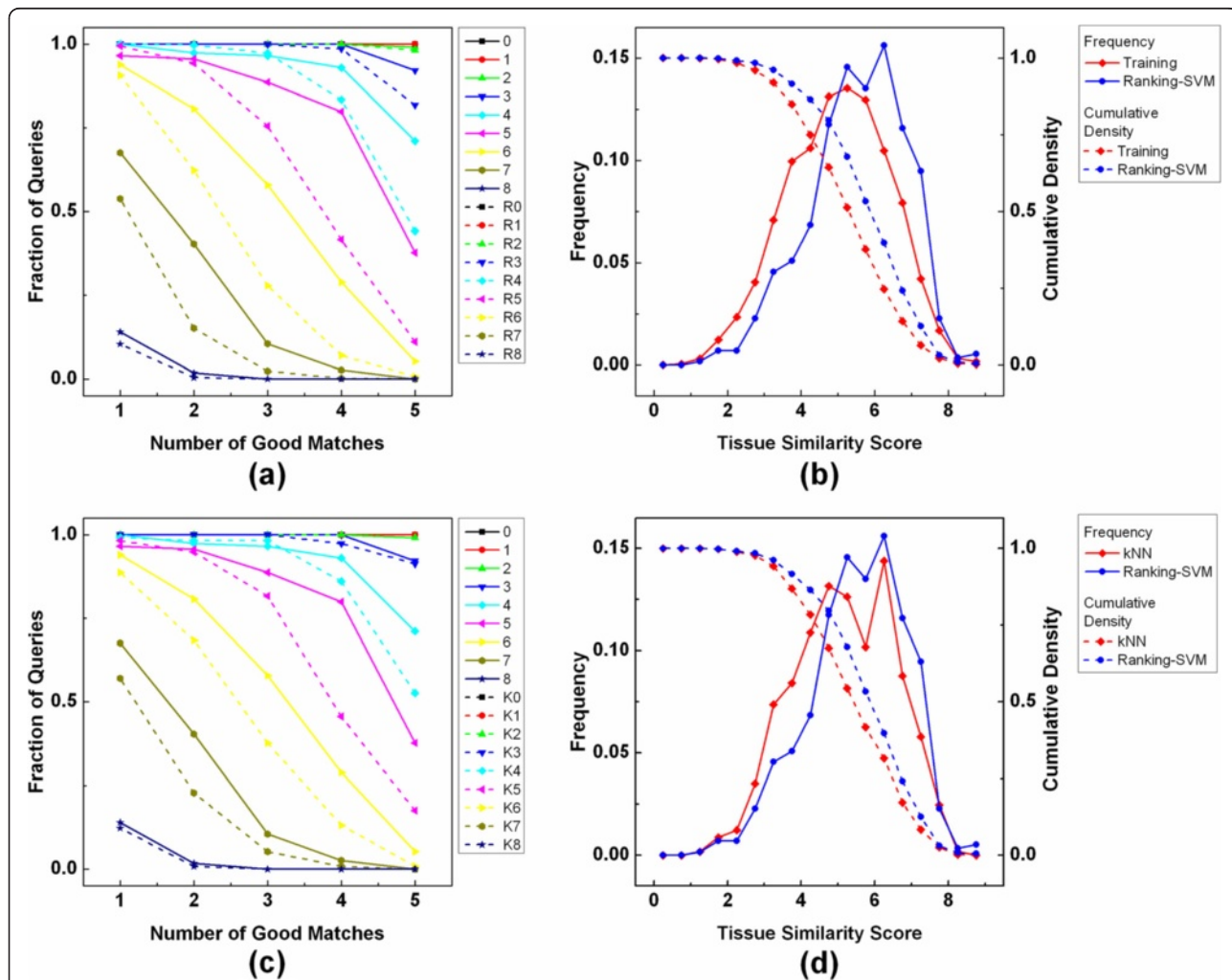
Fig. 4 Tissue Retrieval Performance. The number of queries retrieving at least N_G number of good matches by our system (Ranking-SVM), out of *T* retrieved samples, is computed ($N_G = 1, ..., T$), and compared to **a** the random chance (R0 ~ R9) and **c** kNN retrieval (K0 ~ K9) obtaining that number of good matching cases. The frequency and cumulative density of similarity scores are plotted for **b** the entire training samples and *T* matching samples by our system, respectively. **d** The frequency and cumulative density are also plotted for kNN retrieval. A good matching case is defined as a pair of samples whose similarity score is ≥ th_s, $th_s = 0, ..., 8$. Random chance of retrieving

$\geq N_G$ good matching cases is computed as $\Pr(X \geq N_G) = \sum_{x \geq N_G} \dfrac{\dbinom{N_{ss}}{x}\dbinom{N_s - N_{ss}}{T - x}}{\dbinom{m}{T}}$ where N_S and N_{SS} denote the number of samples in

the database and the number of samples whose *TMS* with the query ≥ th_s, respectively

Kwak et al. BMC Bioinformatics (2016) 17:227

Page 9 of 12

Table 2 Statistical significance of tissue retrieval

		th_s		
		5	6	7
N_G	3	101 <0.001	66 <1.e-10	12 <0.0001
	4	91 <1.e-16	33 <1.e-11	3 <0.01
	5	43 <1.e-12	6 <0.001	0 1

The number of queries (N_q) retrieving at least N_G number of good matching cases as a good match is defined as $TMS \geq th_s$ and its statistical significance. Assuming the number of good matches follows a binomial distribution, p-value is computed as $\Pr(X \geq N_q) = \sum_{x \geq N_q} \binom{m}{x} p^x (1-p)^{m-x}$ where p is a random chance of retrieving $\geq N_G$ good matches ($TMS \geq th_s$). (top) the number of queries and (bottom) its statistical significance

demonstrated a 1.5-fold increase in the fraction of the queries retrieving ≥4 and ≥3 good matches, respectively (Fig. 4c). We investigated the distribution of *TMS* scores of pairs of the query and top-*T* matching samples by Ranking-SVM and kNN (Fig. 4d). Ranking-SVM showed higher *TMS* scores than kNN (*TMS* score ≥ 5). Further, the retrieval results were evaluated by using NDCG (Table 3). Considering top-*T* matching samples, Ranking-SVM achieved the average NDCG of 0.35, and 0.29 NDCG was obtained by kNN on average. NDCG was computed for the ranking of the entire samples in the database; Ranking-SVM and kNN showed the average NDCG of 0.75 and 0.68 NDCG, respectively.

TMS score reveals the complicated relationship between tissues

We examined the utility of *TMS* scores in retrieving similar tissue samples by a visual comparison between tissue H&E images. The relationship between *TMS* score and Gleason sum score was also investigated since Gleason sum score is the only definite information available in prostate pathology today. In Fig. 5, the examples of queries and their matching cases are presented. A pair of samples belonging to the same grade tends to have a (relatively) high *TMS* score, for example, in the second row of Fig. 5, three retrieved samples with Gleason sum score 7 have >6.5 *TMS* score for the query whose Gleason sum score is 7. Other two samples have different Gleason sum score as well as lower *TMS* scores (<5.6). However, high *TMS* scoring sample pairs are not necessarily to be the same grade. In the last row of Fig. 5, none of the retrieved samples are diagnosed with the same Gleason sum score with the query, but their *TMS*

Table 3 Tissue retrieval performance

	Top-*T*	All samples
Ranking-SVM	0.35 ± 0.13	0.75 ± 0.06
kNN	0.29 ± 0.14	0.68 ± 0.06
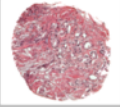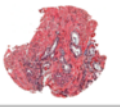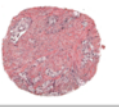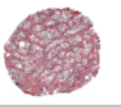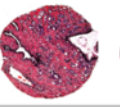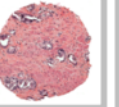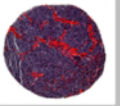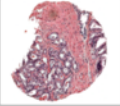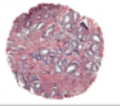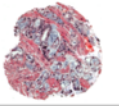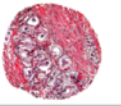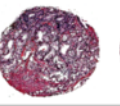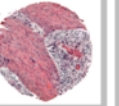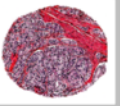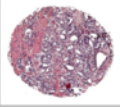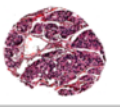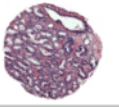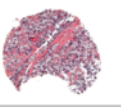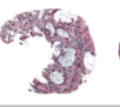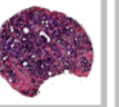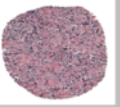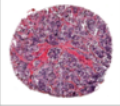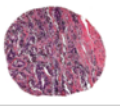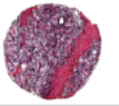
Data represent average ± standard deviation of NDCG

scores are generally high. Four of them have >6.6 *TMS* score, of which each has identical scores with the query for at least 4 morphologic criteria except Gleason score, demonstrating the capability of *TMS* scoring system in examining the relationship between tissues beyond the Gleason grading system. These types of relationships between tissue samples can never be retrieved or assessed if an automated system is built solely on the Gleason grading system. Thus, *TMS* scoring system may help to analyze the complicated and complex tissue morphology and to broaden our understanding.

Discussion

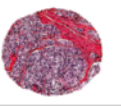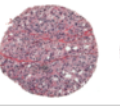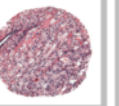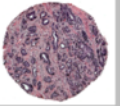Herein, a tissue retrieval system has been developed and tested for prostate cancer. This approach is particularly well suited for cancer and other diagnostic situations where there are multiple parameters applied to defining a grade. In the system, a database allows pathologists to easily manage and maintain the previous cases and outcomes, and immediate access to them is available due to efficient retrieval algorithm. Accordingly, the performance of tissue retrieval is reliant on both a database and a retrieval process. Hence, further study on matching algorithm, performance measure, and data handling, e.g., data normalization, would be necessary, and a large-scale validation study should be conducted to optimize and stabilize the system for various queries, tasks and users' demands.

The size of the database may substantially affect the performance of the retrieval system. In tissue retrieval, it is assumed that the database contains enough number of similar samples to any kind of query. That is, the retrieval system will benefit from the large-scale database, including a variety of patterns of tissue samples from multiple institutions. The retrieval system with the large-scale database will not only serve for various queries and tasks but also improve and stabilize *TMS* scores. The similarity score for a criterion between two samples is dependent on the number of samples between them according to the criterion. The distribution of the samples will become more realistic, leading to the more accurate and reliable similarity measure. Moreover, scoring tissue samples by multiple pathologists will further aid in improving *TMS* scores. However, with the limited size of the database, the distribution of *TMS* score for one query differs from another (Fig. 3a). Some may have many high scoring sample pairs, but some may have few of them. In the latter cases, the retrieval system may return the most similar samples, i.e., the retrieval is valid and useful, but it is a seemingly bad retrieval due to relatively lower *TMS* score. The overall distribution of *TMS* score also affects the retrieval. In our study, a limited number of tissue sample pairs show a high or low *TMS* score (Fig. 3b), i.e., it is likely that the system retrieves

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 10 of 12

Fig. 5 Examples of queries and their matching cases. For each query (*left column*), 5 closest matches are retrieved. The least similar sample is also vshown (*right column*). *TMS* denotes tissue morphologic similarity score for a pair of samples. GS indicates a Gleason sum score which is a sum of predominant and secondary Gleason scores

tissue samples owning mid-range *TMS* scores. In fact, as we trained Ranking-SVM on the entire training dataset, i.e., without balanced training, less number of samples owning higher *TMS* scores was retrieved for the query (Additional file 1: Figure S1), for example, *TMS* score ≥6. Accordingly, taking a roughly balanced subset of the training dataset is a valid decision and helps to provide a more effective and robust retrieval process.

Gleason grades in the dataset are not evenly distributed. A lack of a sufficient number of samples per grade may result in a loss of information of certain patterns in prostate cancer. However, the imbalance of the distribution in this study is not likely to have a significant impact on the retrieval system. The system is still able to retrieve matching cases from the database. A high *TMS* score does not indicate that a sample pair has the same grade. The effect of each grade on the retrieval system may be further studied to improve and stabilize the retrieval system.

We only retrieved the 5 closest samples to a query. The more samples we retrieve, the higher probability the system provides well matched cases with pathologists. However, retrieving many samples (e.g., >10) will be burden to pathologists due to additional time and effort to decide what samples are relevant and useful. Hence, providing the most similar samples would be more helpful and effective. It necessitates little time and work from

pathologists to judge on the retrieved samples, however deliver good matches. We note that if a pathologist would like to retrieve more or fewer samples from the database, then the retrieval system (Ranking-SVM) should be re-trained by adjusting the number of retrievals. If more samples are added to the database, then the whole system should be re-trained (or updated) by computing TMS scores and morphological features and constructing a new Ranking-SVM. Moreover, as one or more morphological properties are of interest to a pathologist, the similarity score can be re-computed and used to train the retrieval system. The pathologist may indicate that certain matches were better than others, resulting in an updating of the database (e.g., changing TMS score) and matching algorithms as needed. The updating may be conducted in real-time. Therefore, the system is potentially adaptable to users' demand and purpose.

The 9 morphological criteria were manually scored by a pathologist and used to measure *TMS* score. Like Gleason grading, it is still a qualitative measure. Based on the qualitative measure, the pathologist categorizes (or scores) tissue samples per criterion. It is well known that such qualitative measure is subject to inter- and intra-observer variability, i.e., likely mis-score (or mis-classify) tissue samples, in particular for the borderline cases. Poor scoring (or mis-scoring), in our study, will

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 11 of 12

disrupt the similarity measure. However, the impact of mis-scoring on the retrieval system may not be as significant as that of Gleason grading. Mis-scoring in Gleason grading may give rise to a totally different pattern and outcome prediction. Unlikely, *TMS* score is a combined measure of the 9 different properties and varies in a continuous fashion. Some mis-scorings of the 9 criteria clearly affect the similarity measure but may not cause a complete change in the tissue similarity. Nevertheless, a follow-up study is desirable to examine the influence of mis-scorings among the 9 criteria on the similarity measure and the tissue retrieval performance.

Conclusions

We have presented an efficient and effective tissue management and decision-support system. *TMS* score offers an alternate means of assessing tissue characteristics and similarities as well as developing and testing computerized methods. Next steps in development would be the validation and application of this system with additional users. The system can be applied to a diversity of diagnostic entities in histopathology. The approach is adaptable in scale, including reference dataset, scoring metrics and matches presented to the pathologist. We anticipate that this approach will open a new direction for the development of automated methods for cancer pathology.

Additional file

Additional file 1: Supplementary material. (PDF 437 kb)

Availability of data and material
The source code, datasets, and supplementary information are available through the following link: chemimage.illinois.edu.

Authors' contributions
Conception and design: JK, SMH, SS, RB. Development of methodology: JK, AK-B, SS, RB. Acquisition of data: JK, AK-B, SMH. Analysis and interpretation of data: JK, AKB, SS, RB. Writing, review, and/or revision of the manuscript: JK, AK-B, SMH, SS, RB. Study supervision: SS, RB. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Ethics statement
This study was performed on diagnostic specimens with information that neither identified the subjects directly nor indirectly through identifiers linked to the subjects. It was approved by and performed in accordance with the University of Illinois at Urbana-Champaign Institutional Review Board. The approved project is entitled "Optical spectroscopy and imaging of archival fixed tissue," case number 06684, and consisted only of secondary analysis performed on anonymized archival tissue and, as such, according to the University of Illinois at Urbana-Champaign IRB policy, is exempt from written, informed consent.

Author details
[1]Department of Computer Science and Engineering, Sejong University, Seoul 05006, Korea. [2]Tissue Array Research Program, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20850, USA. [3]Department of Pathology, University of Illinois at Chicago, Chicago, IL 60612, USA. [4]Department of Computer Science, University of Illinois at Urbana-Champaign, 2122 Siebel Center, 201 N. Goodwin Avenue, Urbana, IL 61801, USA. [5]Beckman Institute for Advanced Science and Technology, Department of Bioengineering, Department of Mechanical Science and Engineering, Electrical and Computer Engineering, Chemical and Biomolecular Engineering and University of Illinois Cancer Center, University of Illinois at Urbana-Champaign, 4265 Beckman Institute 405 N. Mathews Avenue, Urbana, IL 61801, USA.

References
1. Humphrey PA. Prostate pathology. Chicago: American Society for Clinical Pathology; 2003.
2. Gleason DF. Classification of prostatic carcinomas. Cancer chemotherapy reports Part 1. 1966;50(3):125–8.
3. Simmons MN, Berglund RK, Jones JS. A practical guide to prostate cancer diagnosis and management. Clev Clin J Med. 2011;78(5):321–31.
4. Montironi R, Mazzucchelli R, Scarpelli M, Lopez-Beltran A, Fellegara G, Algaba F. Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies. Bju Int. 2005;95(8):1146–52.
5. Cintra M, Billis A. Histologic grading of prostatic adenocarcinoma: Intraobserver reproducibility of the Mostofi, Gleason and Böcking grading systems. International urology and nephrology. 1991;23(5):449–54.
6. Ozdamar SO, Sarikaya S, Yildiz L, Atilla MK, Kandemir B, Yildiz S. Intraobserver and interobserver reproducibility of WHO and Gleason histologic grading systems in prostatic adenocarcinomas. International urology and nephrology. 1996;28(1):73–7.
7. Egevad L, Allsbrook WC, Epstein JI. Current practice of Gleason grading among genitourinary pathologists. Hum Pathol. 2005;36(1):5–9.
8. Epstein JI, Allsbrook Jr WC, Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol. 2005;29(9):1228–42.
9. Stotzka R, Manner R, Bartels PH, Thompson D. A hybrid neural and statistical classifier system for histopathologic grading of prostatic lesions. Anal Quant Cytol Histol. 1995;17(3):204–18.
10. Wetzel AW, Crowley R, Kim SJ, Dawson R, Zheng L, Joo YM, Yagi Y, Gilbertson J, Gadd C, Deerfield DW, et al. Evaluation of prostate tumor grades by content based image retrieval. P Soc Photo-Opt Ins. 1999;3584:244–52.
11. Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M, Tomaszewski J. Automated grading of prostate cancer using architectural and textural image features, I S Biomed Imaging. 2007. p. 1284–7.
12. Naik S, Doyle S, Feldman M, Tomaszewski J, Madabhushi A. Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. In: MIAAB workshop. 2007. p. 1–8.
13. Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, Saidi O. Multifeature prostate cancer diagnosis and Gleason grading of histological images. Ieee T Med Imaging. 2007;26(10):1366–78.
14. Arif M, Rajpoot N. Classification of potential nuclei in prostate histology images using shape manifold learning, International Conference on Machine Vision 2007, Proceedings. 2007. p. 113–8.
15. Farjam R, Soltanian-Zadeh H, Jafari-Khouzani K, Zoroofi RA. An image analysis approach for automatic malignancy determination of prostate pathological images. Cytom Part B-Clin Cy. 2007;72B(4):227–40.
16. Sparks R, Madabhushi A. Explicit shape descriptors: Novel morphologic features for histopathology classification. Med Image Anal. 2013;17(8):997–1009.
17. Smith Y, Zajicek G, Werman M, Pizov G, Sherman Y. Similarity measurement method for the classification of architecturally differentiated images. Comput Biomed Res. 1999;32(1):1–12.
18. Jafari-Khouzani K, Soltanian-Zadeh H. Multiwavelet grading of pathological images of prostate. Ieee T Bio-Med Eng. 2003;50(6):697–704.

Kwak *et al. BMC Bioinformatics* (2016) 17:227

Page 12 of 12

19. Farjam R, Soltanian-Zadeh H, Zoroofi RA, Jafari-Khouzani K. Tree-structured grading of pathological images of prostate. Medical Imaging 2005: Image Processing, Pt 1-3. 2005;5747:840–51.

20. Huang PW, Lee CH. Automatic Classification for Pathological Prostate Images Based on Fractal Analysis. Ieee T Med Imaging. 2009;28(7):1037–50.

21. Schulte EKW. Standardization of Biological Dyes and Stains - Pitfalls and Possibilities. Histochemistry. 1991;95(4):319–28.

22. Muller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. Int J Med Inform. 2004;73(1):1–23.

23. Caicedo JC, Conzalez FA, Triana E, Romero E. Design of a medical image database with content-based retrieval capabilities. Lect Notes Comput Sc. 2007;4872:919–31.

24. Wei C-H, Li C-T, Wilson R. A content-based approach to medical image database retrieval, Database Modeling for Industrial Data Management: Emerging Technologies and Applications. 2005. p. 258–90.

25. Naik J, Doyle S, Basavanhally A, Ganesan S, Feldman MD, Tomaszewski JE, Madabhushi A. A boosted distance metric: application to content based image retrieval and classification of digitized histopathology. In: SPIE Medical Imaging. Lake Buena Vista, USA: International Society for Optics and Photonics: 72603F-72603F-72612; 2009.

26. Sparks R, Madabhushi A. Out-of-Sample Extrapolation Using Semi-Supervised Manifold Learning (Ose-Ssl): Content-Based Image Retrieval for Prostate Histology Grading, 2011 8th Ieee International Symposium on Biomedical Imaging: From Nano to Macro. 2011. p. 734–7.

27. Sridhar A, Doyle S, Madabhushi A. Boosted Spectral Embedding (Bose): Applications to Content-Based Image Retrieval of Histopathology, 2011 8th Ieee International Symposium on Biomedical Imaging: From Nano to Macro. 2011. p. 1897–900.

28. Akakin HC, Gurcan MN. Content-Based Microscopic Image Retrieval System for Multi-Image Queries. Ieee T Inf Technol B. 2012;16(4):758–69.

29. Yu FY, Ip HHS. Semantic content analysis and annotation of histological images. Comput Biol Med. 2008;38(6):635–49.

30. Caicedo JC, Gonzalez FA, Romero E. Content-based histopathology image retrieval using a kernel-based semantic annotation framework. J Biomed Inform. 2011;44(4):519–28.

31. Mehta N, Alomari RS, Chaudhary V. Content Based Sub-Image Retrieval System for High Resolution Pathology Images Using Salient Interest Points. In: Engineering in Medicine and Biology Society. Minneapolis, USA: Annual International Conference of the IEEE; 2009. p. 3719-3722.

32. Comaniciu D, Meer P, Foran DJ. Image-guided decision support system for pathology. Mach Vision Appl. 1999;11(4):213–24.

33. Yang L, Tuzel O, Chen WJ, Meer P, Salaru G, Goodell LA, Foran DJ. PathMiner: A Web-Based Tool for Computer-Assisted Diagnostics in Pathology. Ieee T Inf Technol B. 2009;13(3):291–9.

34. Zheng L, Wetzel AW, Gilbertson J, Becich MJ. Design and analysis of a content-based pathology image retrieval system. Ieee T Inf Technol B. 2003;7(4):249–55.

35. Lessmann B, Nattkemper TW, Hans VH, Degenhard A. A method for linking computed image features to histological semantics in neuropathology. J Biomed Inform. 2007;40(6):631–41.

36. Doyle S, Hwang M, Naik S, Feldman M, Tomaszewski J, Madabhushi A. Using manifold learning for content-based image retrieval of prostate histopathology. In: MICCAI 2007 Workshop on Content-based Image Retrieval for Biomedical Image Archives: Achievements, Problems, and Prospects. Heidelberg, Germany: Citeseer; 2007. p. 53-62.

37. Nwosu V, Carpten J, Trent JM, Sheridan R. Heterogeneity of genetic alterations in prostate cancer: evidence of the complex nature of the disease. Hum Mol Genet. 2001;10(20):2313–8.

38. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Modern Pathol. 2004;17(3):292–306.

39. Gleason DF, Mellinge G. Prediction of Prognosis for Prostatic Adenocarcinoma by Combined Histological Grading and Clinical Staging. J Urology. 1974;111(1):58–64.

40. Mellinger GT. Prognosis of prostatic carcinoma. Recent Results Cancer Res. 1977;60:61–72.

41. Gleason DF. Histologic Grading of Prostate-Cancer - a Perspective. Hum Pathol. 1992;23(3):273–9.

42. Harnden P, Shelley MD, Coles B, Staffurth J, Mason MD. Should the Gleason grading system for prostate cancer be modified to account for high-grade tertiary components? A systematic review and meta-analysis. Lancet Oncol. 2007;8(5):411–9.

43. Iczkowski KA, Lucia MS. Current perspectives on Gleason grading of prostate cancer. Current urology reports. 2011;12(3):216–22.

44. Bhargava R. Infrared Spectroscopic Imaging: The Next Generation. Appl Spectrosc. 2012;66(10):1091–120.

45. Kwak JT, Hewitt SM, Sinha S, Bhargava R. Multimodal microscopy for automated histologic analysis of prostate cancer. Bmc Cancer. 2011;11:62.

46. Joachims T. Training linear SVMs in linear time. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining: 2006. ACM: 217-226.

47. Guyon I, Elisseeff A. An introduction to variable and feature selection. The Journal of Machine Learning Research. 2003;3:1157–82.

48. Yu H, Kim S. SVM Tutorial—Classification, Regression and Ranking. In: Handbook of Natural Computing. Heidelberg, Germany: Springer; 2012. p. 479-506.

49. Fernandez DC, Bhargava R, Hewitt SM, Levin IW. Infrared spectroscopic imaging for histopathologic recognition. Nat Biotechnol. 2005;23(4):469–74.

50. Veltri RW, Partin AW, Miller MC. Quantitative nuclear grade (QNG): A new image analysis-based biomarker of clinically relevant nuclear structure alterations. J Cell Biochem. 2000;79:151-57.

51. Kavantzas N, Agapitos E, Lazaris AC, Pavlopoulos RM, Sofikitis N, Davaris P. Nuclear/nucleolar morphometry and DNA image cytometry as a combined diagnostic tool in pathology of prostatic carcinoma. J Exp Clin Canc Res. 2001;20(4):537–42.

52. Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. Nat Rev Cancer. 2004;4(9):677–87.

53. Ayala G, Tuxhorn JA, Wheeler TM, Frolov A, Scardino PT, Ohori M, Wheeler M, Spitler J, Rowley DR. Reactive stroma as a predictor of biochemical-free recurrence in prostate cancer. Clin Cancer Res. 2003;9(13):4792–801.

54. Cordon-Cardo C, Kotsianti A, Verbel DA, Teverovskiy M, Capodieci P, Hamann S, Jeffers Y, Clayton M, Elkhettabi F, Khan FM, et al. Improved prediction of prostate cancer recurrence through systems pathology. J Clin Invest. 2007;117(7):1876–83.

55. Khamis ZI, Sahab ZJ, Byers SW, Sang QXA. Novel Stromal Biomarkers in Human Breast Cancer Tissues Provide Evidence for the More Malignant Phenotype of Estrogen Receptor-Negative Tumors. J Biomed Biotechnol. 2011;2011:1-7.

56. Tomas D, Spajic B, Milosevic M, Demirovic A, Marusic Z, Kruslin B. Extensive retraction artefact predicts biochemical recurrence-free survival in prostatic carcinoma. Histopathology. 2011;58(3):447–54.

57. Iczkowski KA, Torkko KC, Kotnis GR, Wilson RS, Huang W, Wheeler TM, Abeyta AM, Lucia MS. Pseudolumen size and perimeter in prostate cancer: correlation with patient outcome. Prostate Cancer. 2011;2011:693853.

58. Iczkowski KA, Torkko KC, Kotnis GR, Wilson RS, Huang W, Wheeler TM, Abeyta AM, La Rosa FG, Cook S, Werahera PN, et al. Digital Quantification of Five High-Grade Prostate Cancer Patterns, Including the Cribriform Pattern, and Their Association With Adverse Outcome. Am J Clin Pathol. 2011;136(1):98–107.

59. Epstein JI, Netto GJ. Biopsy interpretation of the prostate. Philadelphia, USA: Lippincott Williams & Wilkins; 2008.

60. Bhargava R, Fernandez DC, Hewitt SM, Levin IW. High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data. Bba-Biomembranes. 2006;1758(7):830–45.

61. Kwak JT, Sinha S, Bhargava R. A New Segmentation Framework for Infrared Spectroscopic Imaging Using Frequent Pattern Mining, 2011 8th Ieee International Symposium on Biomedical Imaging: From Nano to Macro. 2011. p. 452–5.

62. Pudil P, Novovicova J, Kittler J. Floating Search Methods in Feature-Selection. Pattern Recogn Lett. 1994;15(11):1119–25.

63. Järvelin K, Kekäläinen J. IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval: 2000. ACM: 41-48.

64. Scheel C, Lommatzsch A, Albayrak S. Performance Measures for Multi-Graded Relevance. In: SPIM. 2011. p. 54–65.