Supplementary table

Table S1. The tumor biomarkers immunohistochemical staining information

Pathological signaling	Tumor markers	Antibody information	Dilutions
Migration & Invasion			
	E-Cadherin	Cell Signaling, # 4065	1:50
	N-Cadherin	Upstate, clone 13A9	1:200
	β-catenin	Millipore, MAB2081	1:100
	Snail	Abcam, ab70983	1:50
	Twist	Santa Cruz, SC-102032	1:50
	C-Met	Santa Cruz, SC-161	1:200
	nm23-H1	Santa Cruz, SC-56928	1:200
Apoptosis & autophagy			
	Beclin 1	Santa Cruz, SC-11427	1:200
	AKT 1	Cell Signaling, # 4685	1:100
	Bax	Santa Cruz, SC-7480	1:200
	Survivin	Santa Cruz, SC-47750	1:100
	Bcl-2	Santa Cruz, SC-7382	1:50
	Pontin	Cell Signaling, # 8959	1:100
Cell cycle			
	14-3-3σ	Santa Cruz, SC-100638	1:200
	CENP-H	Santa Cruz, SC-22792	1:200
	Aurora-A	Upstate, # 04-1037	1:200
	Cyclin D1	Cell Signaling, # 2978	1:200
	CDC2	Santa Cruz, SC-53	1:200
	Ki-67	Santa Cruz, SC-23900	1:50
	ERK	Santa Cruz, SC-94	1:200
	P21WAF1	Santa Cruz, SC-817	1:200
	p-ERK	Santa Cruz, SC-7383	1:100
	P27	Millipore, clone Y236	1:200
	Stathmin	Cell Signaling, # 3352	1:200
Microvessel density			
	CD34	MaiXin, MAB-0034-P	1:200
	CD31	MaiXin, MAB-0031	1:200
Tumor microenvironment			
	MMP-2	Santa Cruz, SC-53630	1:200
	MMP-9	Santa Cruz, SC-6840	1:200
	TIMP-2	Santa Cruz, SC-21753	1:200
	COX2	Santa Cruz, SC-58344	1:200
	HIF-1α	Millpore, MAB5382	1:200
Others			
	EZH2	Cell Signaling, # 4905	1:200
	LMP 1	Santa Cruz, SC-57721	1:200

Table S2. The predictive efficacy of decision tree algorithm in prediction of recurrence pattern

for locally advanced NPC

Dataset	Test scheme	PPV	NPV	Sensitivity	Specificity	AUC	OA
Overall patients	Training subset	86.0	81.0	91.0	72.0	80.0	84.5
	Overall patients	87.6	85.4	92.4	77.4	92.2	86.9
IC/CCRT subset	Training subset	100.0	88.9	92.3	100.0	95.2	95.2
	IC/CCRT patients	85.7	88.9	93.3	77.4	91.3	86.8
IC/RT subset	Training subset	87.0	82.0	93.0	69.0	86.2	85.7
	IC/RT patients	93.2	76.0	87.2	87.4	93.6	87.0

Supplementary methods

Feature subset selection

Feature subset selection algorithm usually falls into two categories, named as filter and wrapper methods [1]. Filter method select features subset at a pre-processing step, which is independent of the chosen predictor. Among the existing filter methods in feature weighting, the RELIEF algorithm is considered as one of the most powerful ones due to its simplicity and effectiveness [2]. Among of these RELIEF algorithms, the Local Linear RELIEF (LL-RELIEF) model is the pioneering technique in the RELIEF family [3]. LL-RELIEF weights the feature importance by iterative maximizing the margin between different sample subsets. The features weights are estimated under assumption of local linear, thus a given complex problem is analysed by parsing it into a set of locally linear problems. LL-RELIEF has been shown to be effective in removing redundant features and in handling many feature selection problems [4]. In comparison, wrapper method uses base classifier to score features subset according to their predictive power. Though there are many classifiers, such as Support Vector Machine (SVM), Naive Bayes, selecting of suitable classier catering to the interested dataset is still an open problem [5,6]. [3, 4]Alternatively, one may increase the performance of weak binary classifiers, as the decision tree algorithm in our case, by reinforcing training on misclassified samples through voting to combine the output of multiple models [7]. This technique, which is called boosting, can achieve comparable performance to classical classifier and is less sensitive to data characteristics [8,9]. The Adaptive boosting (AdaBoost) is a widely used boosting algorithm due to its high efficiency and implementation simplicity [10,11]. The wrapper method has the advantage of favourable performance; however, its usage in biomedical area is limited because of its high computational cost.

AdaBoost model

AdaBoost is a boosting algorithm, which runs a given weak learner several times on slightly altered training data, and combines the derived hypotheses to one final hypothesis in order to achieve greater accuracy than the weak learner's hypothesis would have [12]. The AdaBoost assumes that each training example of the set act in different discrimination roles at different training stages. The features that can be easily recognized are supposed to have less power in classification, while the features that are misclassified should be penalized to have good discrimination power. Therefore, the weak learner focuses on the 'difficult' features, which are believed to be rich of information. The richness of each feature is represented by a weight. The purpose of the classification is to pursue accurate weights for all the participants by adaptively adjusting, which is based on the classification results in every round. The final classification result is a combination of the results obtained from all rounds.

Measurements after receiving operating characteristics analysis

The following terms were calculated for evaluating algorithm performance. True positive (TP) and true negative (TN) means the total number of correctly predicted patients, and false negative (FN) and false positive (FP) indicates total number of misclassified patients. The overall accuracy is defined as $OA = (TP+TN)/(TP+TN+FP+FN)$, which may also be reported as a percentage. We defines the sensitivity $= TP/(TP+FN)$ and specificity $= FN/(TN+FN)$. AUC refers to area under the receiver operating characteristic (ROC) curve, which is a plot of true positive rate (i.e. sensitivity) versus false positive rate (i.e. 1-specificity) for predicted patients.

References

1. Sun Y. Iterative RELIEF for feature weighting: algorithms, theories, and applications. IEEE Trans Pattern Anal Mach Intell. 2007;29:1035-1051.

2. Sun Y, Todorovic S, Goodison S. Local-learning-based feature selection for high-dimensional data analysis. IEEE Trans Pattern Anal Mach Intell. 2010;32:1610-1626.

3. Sun Y, Goodison S, Li J, et al. Improved breast cancer prognosis through the combination of clinical and genetic markers. Bioinformatics. 2007;23:30-37.

4. Chen HL, Yang B, Liu J, et al. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Systems with Applications. 2011;38:9014–9022.

5. Aruna S, Rajagopalan SP, Nandakishore LV. Knowledge based analysis of various statistical tools in detecting breast cancer. Computer Science & Information Technology (CS&IT). 2011;2:37-45.

6. Quinlan JR. C4. 5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning). 1st ed. Morgan Kaufmann Publishers. 1993.

7. Lee JW, Lee JB, Park M, et al. An extensive comparison of recent classification tools applied to microarray data. Computational Statistics & Data Analysis. 2005;48:869–885.

8. Sun YJ, Todorovic S, Goodison S. Local-learning-based feature selection for high-dimensional data analysis. IEEE Trans Pattern Anal Mach Intell. 2010;32:1610-1620.

9. Takahashi H, Aoyagi K, Nakanishi Y, et al. Classification of intramural metastases and lymph node metastases of esophageal cancer from gene expression based on boosting and projective adaptive resonance theory. J Biosci Bioeng. 2006;102:46-52.

10. Moon H, Ahn H, Kodell RL, et al. Ensemble methods for classification of patients for personalized medicine with high-dimensional data. Artif Intell Med. 2007;41:197-207.

11. Takahashi H, Aoyagi K, Nakanishi Y, et al. Classification of intramural metastases and lymph node metastases of esophageal cancer from gene expression based on boosting and

projective adaptive resonance theory. Journal of Bioscience and Bioengineering. 2006;102:46-52.

12. Yetton BD, Niknazar M, Duggan KA, et al. Automatic detection of rapid eye movements (REMs): A machine learning approach. J Neurosci Methods. 2015;259:72-82.