Predicting distant metastatic sites of cancer using perturbed correlations of miRNAs with competing endogenous RNAs

Cho et al., 2025, Computational Biology and Chemistry

**Background:**

* There are few attempts to predict distant metastasis.
  + 1) Predicting distant metastasis is more difficult than predicting lymph node metastasis because it is a multi-label (i.e. one sample can have multiple metastatic sites) and multi-class (i.e. there are multiple different sites) problem
  + 2) There are only a small number of publicly available tumor samples with distant metastasis that can be used for training a learning method.

**Hypothesis:**

* Development and metastasis of cancer is better explained by the dysregulation of gene interactions rather by individual genes alone.
* Competitive endogenous RNA (ceRNA) hypothesis suggests that RNAs with similar miRNA response elements compete to bind the same miRNA, thereby regulating each other indirectly.
* We can predict distant metastasis based on changes in the correlation of miRNAs with their competitive endogenous RNAs (ceRNAs)

**Methods:**

* TCGA data
* Four cancer types (bladder urothelial carcinoma (BLCA), N=109, breast invasive carcinoma (BRCA), N=136, esophageal carcinoma (ESCAN), N=54 and liver hepatocellular carcinoma (LIHC), N=95)
* Predicting distant metastasis based on changes in the correlation of miRNAs with their competitive endogenous RNAs (ceRNAs)
* Multi-class and multi-label problem transformed into multiple single-label binary classification problems.
* Cancer datasets were partitioned into training and test datasets with a ratio of 7:3.
* Due to the severe imbalance between the positive and negative samples, we oversampled the positive samples in the training dataset using the synthetic minority oversampling technique (SMOTE).

**Results:**

* In independent testing, MCC (Matthew’s correlation coefficient) of Mdm ranged from 0.778 to 0.894 for four cancer types. The highest performance was observed in LIHC.

MetastaSite: Predicting metastasis to different sites using deep learning with gene expression data

Albaradei et al., 2022, Frontiers in Molecular Biosciences

**Background:**

* Precision medicine is a path that could profoundly change and improve medical practices.
* Deep learning is showing promise in precision medicine as it can extract intricate structures in high-dimensional data.

**Hypothesis:**

* Develop an AI method that could translate into a tool that supports clinical decision-making with regard to identifying metastasis and pinpointing the metastasis site.

**Methods:**

* SMOTE to oversample the minority class using the imbalanced-learn python library, as the number of samples is imbalanced between the primary and metastasized group.
* AutoEncoder framework to learn the non-linear relationship between genes (i.e. reduce the dimension of the expression data), and then
* DeepLIFT was applied to calculate genes’ importance scores used to rank the genes
* Iteratively added ten top-ranked genes based on their importance score to train a DNN model.
* Trained a multi-class deep neural network (DNN) model to predict whether samples are primary or metastasized to the brain, bone, lung, or liver.

**Results:**

* Deep neural network model: “MetastaSite” based on gene expression profiles.
* The prediction performances ranged from AUC of 0.93-0.82.

Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer

Liu et al., 2020, Nature Communications

**Background:**

* Locally advanced rectal cancer (LARC) is the most common form of rectal cancer.
* Distant metastasis (DM) is the main cause of treatment failure in patients with LARC, as the incidence of DM remained 25-40%.
* Rectal cancer guidelines recommend adjuvant chemotherapy following total mesorectal excision to reduce the incidence of DM. However, this may only work in a subset of patients.
* It is crucial to detect which LARC patients could benefit from adjuvant chemotherapy.
* MRI is widely used for diagnosing and staging of rectal cancer and can detect several prognostic factors.

**Hypothesis:**

* To validate a model to predict DM after surgery from imaging features, thus identifying patients who can benefit from adjuvant chemotherapy.

**Methods:**

* 629 patients underwent MRI
* ROIs were manually delineated, and feature extraction performed using MATLAB. Image features were normalised to a Z-score. Coarse-to-fine feature selection strategy was used to reduce the risk of bias and potential overfitting.
* Univariate Cox analysis was used to detect associations between feature and patient’s DMFS (primary outcome – time from surgery to first confirmed instance of DM or death). All features were then ranked in ascending order according to Cox p values, and the top 20% of features with p<0.1 were used for further analysis.
* Pearson correlation coefficients for each feature pair was calculated. Feature pairs with R>0.6 were selected.
* LASSO algorithm with Cox analysis was used to identify the most useful prognostic features for constructing the radiomic signature.
* Kaplan-Meier survival analysis performed to assess association between radiomic signature and DMFS.

**Results:**

* In conclusion, we identified a multiparametric MRI-based radiomic signature that effectively predicted DMFS in LARC patients and improved the performance of the traditional clinicopathological prediction model. Combining the radiomic signature with pathological stage might help identify which patients are expected to benefit from adjuvant chemotherapy.