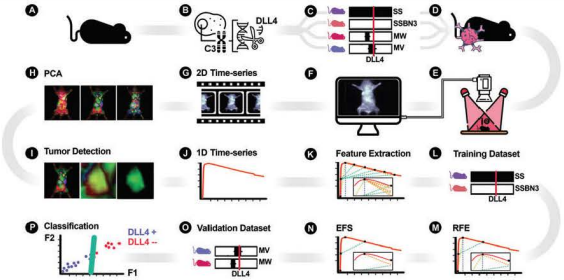
DEPARTMENT OF BIOMEDICAL ENGINEERING
MARQUETTE UNIVERSITY
MEDICAL COLLEGE OF WISCONSIN

IBM
Quantum Computing
IBM Research Europe - Dublin

A machine learning framework for stratifying high vs. low notch-DLL4 expressing host microenvironment for breast cancer bearing subjects

Shayan Shafiee, Mykhaylo Zayats, Jonathan Epperlein, Sergiy Zhuk, Amit Joshi



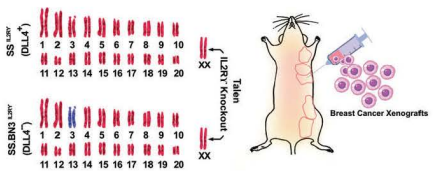


Figure 1. (A) Salt sensitive(SS) rat with high systemic DLL4 expression (B, C) Substitution of the third chromosome from Brown Norway with low DLL4 expression into the SS background to generate consomic SSBN3, which mimics clinically observed vascular and treatment response heterogeneity. We also constructed two novel SSBN3 congenic (inbred strains containing a given genomic region in their genome) xenograft host strains (MV and MW) by introducing segments of BN chromosome 3 into the genetic background of the parental SS strain, which resulted in SSBN3 tumors demonstrating significantly higher blood vessel density than SS tumors. (D) Identical triple negative breast cancer tumors were implanted in all rats. (E, F) The whole body imaging platform for rats. Animals were injected intravenously via tail vein with ICG (0.75mg/kg, MP Biomedicals) and simultaneously imaged using a PIMAX4 ICCD camera(Princeton Instruments), equipped with a 25mm lens (Navitar, DO-25, 0.95 f-stop), with an 830nm long pass filter (ThorLabs), excited by 785nm diode laser (~5mW/cm2), was used with a frame rate of 10.6 fps for 6 minutes. (G) the time series intensity of NIR signal for all four groups of animals. (H) Each pixel's fluorescence intensity kinetic curve was decomposed into principal components (PC), which enable clustering and visualization of pixels corresponding to individual organs. (I) A 2D cross-correlation mapping algorithm was used to rank the PCs with a higher chance of showing the tumor and automatically selected the tumor ROIs based on the tumor geometry template. (J) For each tumor mask, pixels were averaged, and a single time series was recorded. (K) Based on TME and therapy response differences, we characterized dynamic perfusion patterns caused by vascular TME using several easily explainable features in the Near Infrared(NIR) time-series of intravenously administered NPs (e.g., the slope of the initial uptake of NP, the time it takes to reach peak perfusion, and the rate of decay of NP after the peak) (L) The classification algorithm was trained on SS and SSBN3 strains (M, N) Best pairs of features in terms of performed sensitivity, specificity, and accuracy were selected by a two-step procedure, recursive feature elimination(RFE) to select four features, followed by an exhaustive feature selection(EFS) algorithm which uses training performance of the classifiers to select two out of four features selected in the first step. (O) The pair of features selected in the previous step was used to evaluate the performance of the classifiers. (P) The decision boundary of a naive Bayes classifier is based on the selected feature combination. The dots denote the data used to evaluate the classifiers.

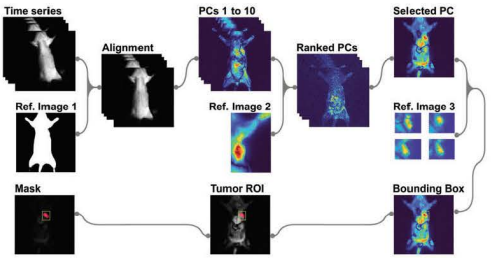


Figure 2. Diagram of Tumor Detection and ROI generation.
The ROI detection module consisted of three main steps: image registration, PCA ranking and selection, ROI selection and masking.
NIR Time series were registered to a reference image using a rigid body transformation to align animals in a consistent direction. The principal components were ranked based on the 2D cross-correlation (2DCC) of a reference image showing the tumor clearly. For this, a stack of the first ten components was generated, and each component went through a 2DCC function and was ranked based on the correlation score accordingly. The suitable component for identifying the tumor, was propose based on the ranking from the earlier step. Tumor's location in the frame and a bounding box were estimated and generate based on a stack of four reference images of different tumors. A combination of basic morphological dilation and image arithmetic operations was applied to the selected PC to emphasize the tumor's boundaries followed by applying a threshold to separate the tumor from the background. The ROI was used to mask the NIR time series.

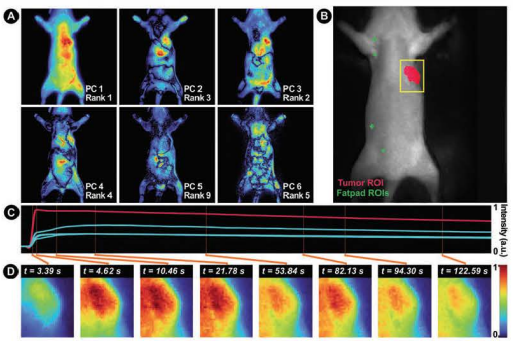
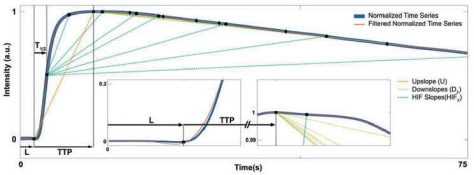


Figure 3. (A) First six principal components reconstruction of the NIR time series and their ranks based on the probability of containing a clear tumor image. (B)The image of the rat overlayed with the proposed tumor location(yellow box), tumor ROI (pink), and fat pad ROIs (green). (C) Mean NIR intensity of ROIs showing intensity distribution from the tumor(red) and fat pads(green). (D) The NIR intensity of the tumor over time.



Abbreviation	Feature	Description
L	Latency	The point in time when fluorescence starts rising beyond the initial background fluorescence
$T_{1/2}$	Time to Half Peak Intensity	The time difference between the half peak intensity time point and the latency
TTP	Time to Peak	The time difference between the peak intensity time point and the latency
U	Upslope	The average slope between initial ICG arrival(latency) and the peak
D_0	Downslopes	The downslope between the peak and X time step further. X = {20,50,80,100,160,250,400}
HIF_X	Half Intensity Forward slopes	The slope between the half peak intensity and X time step further. X = {20,50,80,100,160,250,400}

Figure 4. Normalized mean NIR intensity time series and the designed features.

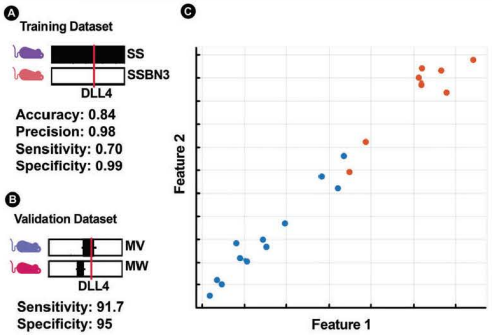


Figure 5. (A) The Training dataset and the primary performance of the classifier. (B) The Secondary validation dataset and the performance of the classifier. (C) The decision boundary of a naive Bayes classifier based on the feature combination D160and HIF160. The dots denote the data used to test the classifiers, MW(red), MV(blue)

Introduction

Nanomedicine and macromolecule drug delivery rely on the enhanced permeation and retention effect in solid tumors, and crosstalk between malignant tumor cells and the non-malignant TME contributes to tumor growth, drug delivery, and therapy efficacy (PMID: 25540894). Delta Like Canonical Notch Ligand4(DLL4) is a protein-coding gene that is responsible for developing blood vessels and plays a role in tumor angiogenesis. With Consomic rat strains differing in inherited levels of DLL4 in host stroma, we have shown that the degree of DLL4-dependent dysfunctional angiogenesis affects tumor growth and metastasis, drug delivery and therapy response (PMID: 32373218). Herein, we propose a machine learning framework to identify and classify hosts with low and high levels of DLL4 expression on tumor endothelium based on the kinetic NIR fluorescence imaging with Indocyanine green dye. Overall objective is to identify tumour bearing animals likely to respond to DLL4 directed therapies.

Materials and Methods

We generated two rat strains, Salt sensitive (SS) with high systemic DLL4 expression and consomic(inbred strains containing a whole chromosome from another strain in their genome) SSBN3 with third chromosome substitution from Brown Norway with low DLL4 expression. We also constructed two novel SSBN3 congenic(inbred strains containing a given genomic region in their genome) xenograft host strains (MV, MW) by introducing segments of BN chromosome 3 into the genetic background of the parental SS strain, such that MV inherited DLL4 locus from SS and MW from BN.

A whole body imaging platform for rats was developed, and all four strains bearing identical tumors were imaged. MDA-MB-231 (231), triple negative breast cancer, was used as a Xenograft cell line. Animals were injected intravenously via tail vein with ICG (0.75mg/kg, MP Biomedicals) and simultaneously imaged using a PIMAX4 ICCD camera(Princeton Instruments), equipped with a 25mm lens (Navitar, DO-25, 0.95 f-stop), with an 830nm long pass filter (ThorLabs), excited by 785nm diode laser (~5mW/cm2), was used with a frame rate of 10.6 fps for 6 minutes. Respiratory motion correction was performed with Fourier-Wavelet methods, MATLAB (MathWorks Inc.). Tumor ROIs were automatically drawn on principle component decomposed images using a 2D cross-correlation mapping algorithm. Pixels' intensity was averaged from the ROIs, and a single time series was recorded. Data from tumor bearing SS and SSBN3 rats were used to refine the features and train the classifier. The framework's performance was validated on MV and MW strains (Figure 1. C, L, O). We characterized dynamic perfusion patterns using several features in the NIR time-series. Best pairs of features in terms of performed sensitivity, specificity, and accuracy were selected by a two-step procedure, recursive feature elimination(RFE), and Exhaustive Feature Selection(EFS) by training classifiers for selecting two out of four feature selected in the first step (PMID: 35308965). After feature selection, the classification algorithm becomes a standard binary classification problem, with feature selection as a subproblem. Finally, we implemented an experimental framework to validate the features by combining tumor detection, feature extraction, and classification algorithms.

Results

The performance of the tumor detection algorithm was visually validated, and all the ROIs passed the inspection. The classifier's accuracy, precision, sensitivity, and specificity over the training run in classifying MW and MV as belonging to their respective DLL4 parent strain were 0.84, 0.98, 0.70, and 0.99, respectively, and the sensitivity and specificity were 91.7 and 95, respectively, over the validation run.

Conclusion

We have demonstrated that whole body non-invasive NIR imaging can discriminate identical tumors based on the differential DLL4 expression in TME using an ML framework with high sensitivity and specificity. If developed for dynamic MR imaging used clinically, similar methods will enable patient stratification for DLL4 targeted therapies.

Funding: NIH/NCI 2R01CA193343 (PI Amit Joshi)