Multivariate characterization of membrane proteins

Ju Han, Hang Chang,
Mary Helen Barcellos-Hoff, and Bahram Parvin
Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Cell membrane proteins play an important role in tissue architecture and cell-cell communication. We hypothesize that segmentation and multivariate characterization of these molecules, on a cell-by-cell basis, enable an improved differentiation of treatment groups and identify important characteristics that can otherwise be hidden. We have developed a series of computational steps, as shown in Figure 1, to (i) delineate cell membrane protein signals in the context of the nuclear morphology, (ii) compute a coupled representation of the multiplexed DNA content with membrane proteins and other end points, (iii) rank computed features associated with such a multivariate representation, (iv) visualize selected features for comparative evaluation, and (v) discriminate between treatment groups in an optimal fashion. The novelty of our method is in the segmentation of the membrane signal and the multivariate representation of the phenotype on a cell-by-cell basis.

A critical component in this multivariate analysis is segmentation of fluorescence signals. The main issues with the segmentation of the membrane protein signal are heterogeneity, scale, noise, and discontinuity in the signal, which leads to perceptual gaps. We have extended iterative voting along the tangential direction for enhancing membrane proteins. Ultimately, the regularized membrane protein signal serves as an external force for an evolving front. The front is initialized by a region-based Voronoi tessellation of the segmented nuclear regions and attracted to the regularized membrane energy. Proposed method can be applied to a wide range of surface proteins and subcellular compartments.

As an example, the proposed computational steps are applied to samples that have been irradiated with different radiation qualities in the presence and absence of TGFb treatment. These samples are labeled for their DNA content and E-cadherin membrane protein. We demonstrate that multivariate representation of cell-by-cell phenotypes improves predictive and visualization capabilities between different treatment groups, and increases quantitative sensitivity of cellular responses. For example, Figure 2 illustrates that combining representations based on quantification of the labeled probe and texture features results in an improved predictor for separating treatment groups. Furthermore, the dose response curves are significantly different between different radiation qualities, as shown in Figure 3. This quantitative insight can only be revealed through cell-by-cell analysis as global florescence analysis hides these differences. Additionally, visualization is augmented with heatmap to provide a compact summary of multivariate data.

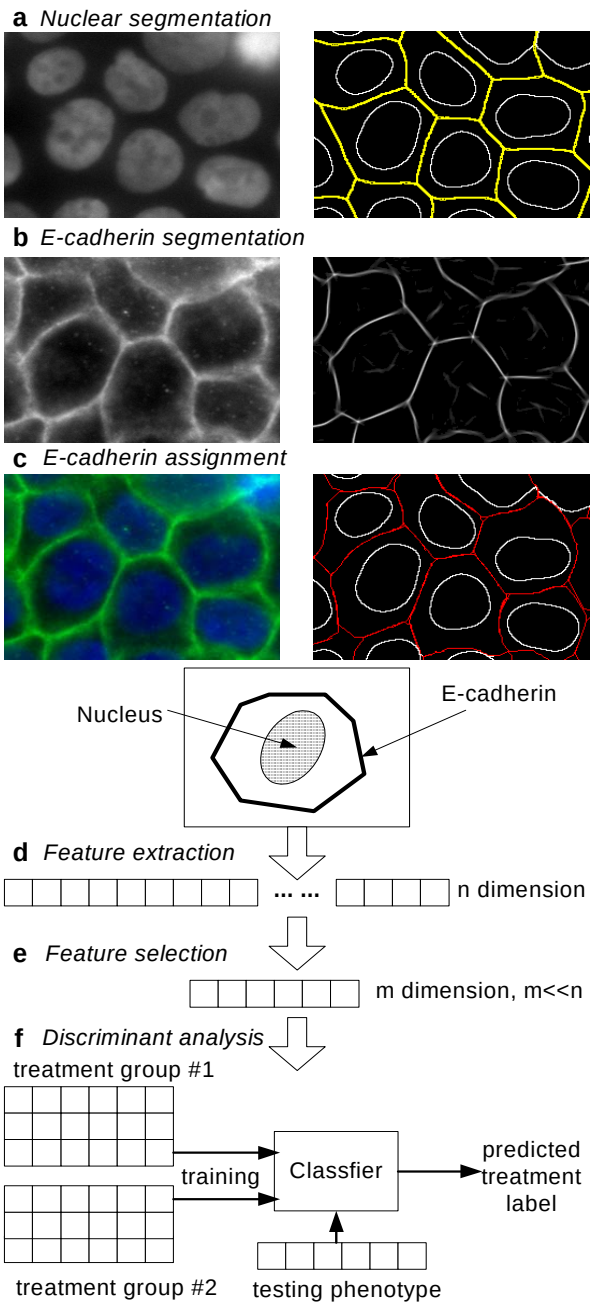

Fig. 1. Multivariate representation of nuclear and E-cadherin responses for discriminant analysis: (a) Each nucleus is localized using a combination of gradient features and a set of geometric constraints and region-based Voronoi tessellation of each cell is then established (yellow boundary). (b) E-cadherin signal is enhances through iterative tangential voting. (c) E-cadherin signal is assigned to the corresponding cell through evolving fronts. (d) Each cell is represented with morphological, structural, and localization features. Furthermore, spatial organization of these features is also captured and registered. (e) An optimal subset of features is selected for maximizing class reparability. (f) The classifier is trained to predict the treatment condition based on selected features.

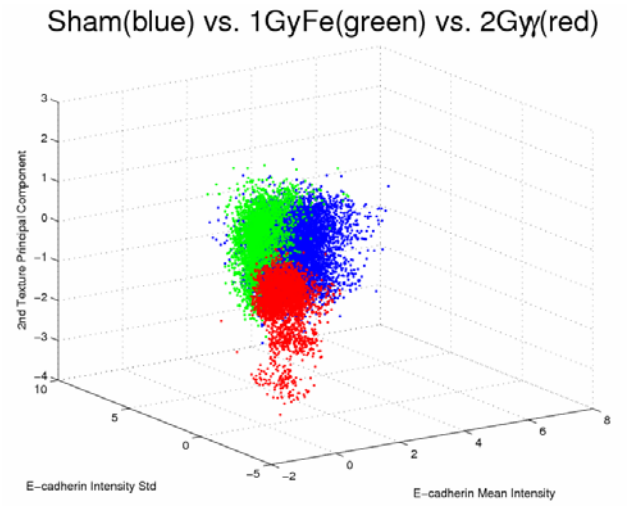

Fig. 2. Scatter plots of features associated with two radiation qualities following feature selection and ranking.

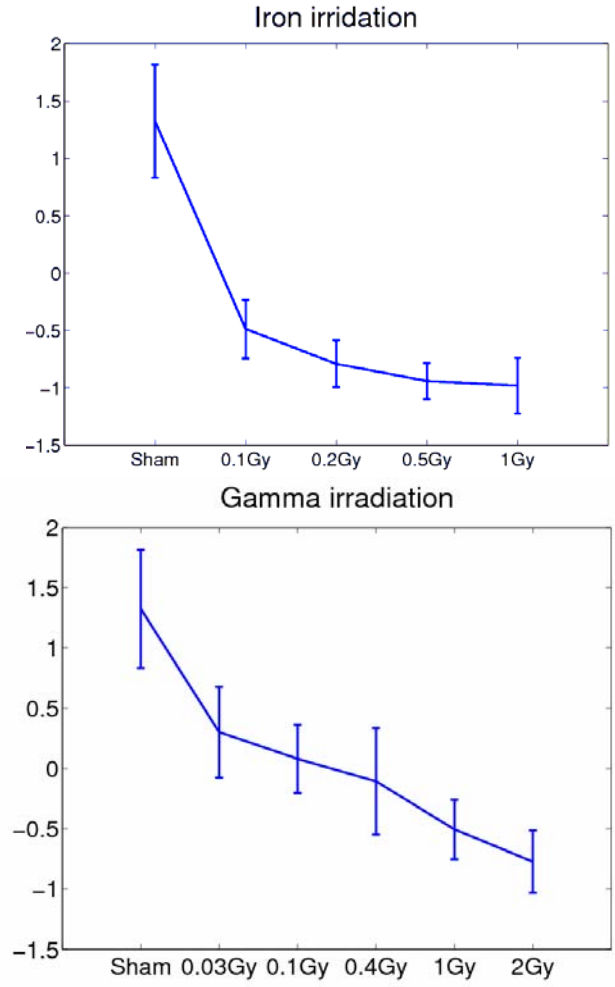

Fig. 3. Dose response of E-cadherin on a cell by cell basis indicates a sharper drop in the membrane protein in low dosage as a result of iron irradiation. Comparison is at equal doses of relative biological effects.