The Interaction Factor: Quantifying Protein-Protein Interactions by Stochastic Modeling of Super-Resolution Fluorescence Microscopy Images

Keria Bermudez-Hernandez[1,2,3*], Sarah Keegan[1,2,3*], Dylan Reid[2], Jennifer Zagelbaum[2], Donna Whelan[2], Yandong Yin[2], Sisi Ma[3,4], Eli Rothenberg[2] and David Fenyo[1,2,3]

[1.] Institute for System Genetics, New York University School of Medicine, New York, NY, USA
[2.] Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, USA
[3.] Center for Health Informatics and Bioinformatics, New York University School of Medicine, New York, NY, USA
[4.] Department of Medicine, New York University School of Medicine, New York, NY, USA
[*] These authors contributed equally to this work.

With the development of super-resolution microscopy (SRM) techniques, it is possible to study biological samples at the sub-diffraction level, allowing researchers to observe proteins and their interactions on the tens of nanometers scale. One area of study is the degree to which proteins labeled with differently colored fluorophores can be found to overlap, or 'co-localize' in an image, which may correspond to interaction between the proteins. In order to measure co-localization, techniques developed for diffraction-limited microscopy may be used, which focus on correlating intensity levels in the color channels of the image or segmenting the image by thresholding and measuring distances from intensity centers of objects [1]. However, these methods can be subject to distortions based on noise or differing levels of intensity of the color channels [1]. Additionally, for single molecule localization microscopy (SMLM) [2], a type of SRM that localizes individual light sources corresponding to labeled proteins, methods have been developed based on spatial analysis techniques, such as Ripley's functions and the pair-correlation function adapted for multiple species [3,4]. These measures are of use and give an overall picture of the degree of correlation between the localized particles, although they require a certain amount of computing effort to implement, and are only applicable where individual particle position tables are available. An intuitive method is to focus on examining the co-localization of identified protein 'clusters' in an image. Once the boundaries of protein clusters are determined, either by image segmentation or spatial clustering techniques, it is possible to measure the amount of overlap. However, problems arise when attempting to compare overlap amounts across images and data sets, and there is no standard by which to quantify the degree of co-localization of an image. In addition, the question may arise as to how to determine whether the overlap observed is merely due to random occurrence. We present a method based on stochastic modeling to quantify in a simple, unbiased manner the degree of co-localization of protein clusters in an image. This measure, which we call 'Interaction Factor', can be compared across data sets and can also be used to distinguish co-localization resulting from a random arrangement of clusters in an image. It is a number from 0 to 1, 0 indicating co-localization due to random occurrence and 1 indicating co-localization of all protein clusters.

The technique is designed for a set of two-color fluorescent microscopy images or particle coordinate tables and corresponding ROIs (regions of interest) over which the protein clusters will be detected. The color channels are separated and the protein clusters are segmented using either thresholding or any other segmentation or clustering method that is acceptable to the researcher. Next, a series of simulated images are produced by drawing from the distributions of clusters present in the original image. The results of these simulations allow estimation of the Interaction Factor (IF) for an image (Figure 1a). The

IF is estimated by producing simulations where clusters are placed at both random positions and at positions corresponding to an increasing probability of co-localization between clusters and observing where the level of co-localization of the real image falls within the distributions of these simulations (Figure 1b). The most likely IF for an image can be estimated by fitting a model to the data produced from creating simulated images over a range of IFs and plotting IF vs. percentage of overlap for the simulations. The center curve is fit to the data points resulting from 20 simulations at each chosen IF, and error ranges are calculated from curves fitted to the 5th and 95th percentile data points (Figure 1c).

References:

[1] S. Bolte and F.P. Cordelieres, J Microsc 224(Pt 3) (2006), p. 213-32.
[2] G. Patterson, et al, Annu Rev Phys Chem 61 (2010), p. 345-67.
[3] C. Coltharp, X. Yang and J. Xiao, Curr Opin Struct Biol 28 (2014), p. 112-21.
[4] H. Deschout, et al, Histochem Cell Biol 142(1) (2014), p. 5-17.

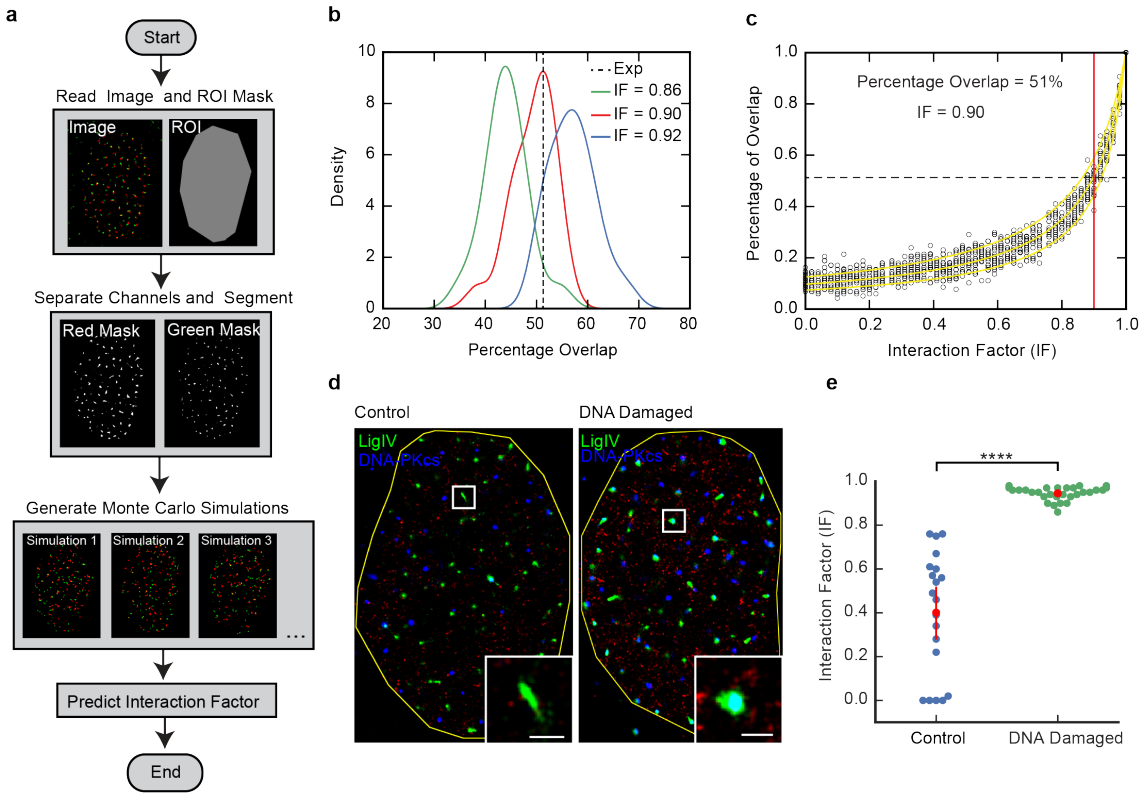

Figure 1. (a) Workflow of algorithm. (b) Density plots showing the percentage of overlap for simulations generated with three different IFs, n = 20 simulations per IF. (c) Percentage of overlap for simulations generated with different IFs. Models were fitted to measurements generated by simulations (yellow lines) to predict the IF of the experimental image (dashed line; IF = 0.90). (d) Examples of SRM images of a U2OS cell nuclei (yellow) in control (left) and after neocarzinostatin treatment (DNA damaged; right) labeled for LigIV (green) and DNA-PKcs (blue). Boxed areas are magnified. Calibration = 500nm. (e) IF between LigIV/DNA-PKcs in control (n = 20) and DNA damaged group (n = 29). The IF between LigIV/DNA-PKcs was greater in DNA damaged (0.95±0.03; mean±SD) compared to control group (0.40±0.28; plot error bars: 95 CI; ****p<0.0001; Welch's t-test).