Dremi produces a line of fit based on binned conditional probabilities of P(Y|range of X). We could use a Kolmagrov-Smirnove test of this line against the line produced under the null hypothesis to determine if there is an interesting relationship between X and Y.

I like the density-dependant down-sampling employed by SPADE more than that employed by DREMI. It uses local neighborhoods to define the likelihood that every point might be down sampled rather than a bin of X values over which the joint distribution is normalized. However, this local uniformity my complicate trend finding if the resulting down-sampled data is uniform. Could we do a weighted combination of these down sampling probabilities per sample point? This would make for easy preliminary work.

Data cleaning is an important step. You could be detecting non-biologically relevant trends between X and Y like RNA degradation, the rate of which has been shown to be both gene and sample dependant (<http://www.nature.com/ncomms/2015/150803/ncomms8816/fig_tab/ncomms8816_F2.html>)

Check out self ordering maps (SOM) for clustering data. Mangiameli et al. say it’s worth it.

Look at the ‘fuzzified’ High, Medium, and low values of gene triplets (woolf and wang, 2000) through pseudo time (progress through a dynamic biological process as inferred by monocle). Could dynamic time warping be useful in determining relationships across pseudotime?

Options for transitioning DREMI into fuzzy logic direction inference:

1. Per gene pair, per quadrant, sum up the fine grid of points’ (G) Ic values which is an estimation of P(Y|X). This is less general and can’t really be applied to other dependency metrics directly.
2. Apply Yates’ workflow to sub-sampled data.
3. Could also take a frequent’s statistical approach and find p-value per quadrant. And p-values per Boolean function described by Yates.

Bump-shaped or double sigmoidal relationships may arise as a result of 3 or more cell states being present in a single batch of samples. Potentially this type of relationship may be successfully broken up by applying our workflow to pairs of clustered samples.

Transcription factors vs enzymes. Why are transcription factors picked up by Boolean implication more easily?