Spatial Correlations and Robust Statistical Analysis for Combinatorial Methodologies

Amis, EJ, Kennedy, SB, Forster, AR, Washburn, NR
Polymers Division, National Institute of Standards and Technology, Gaithersburg, MD, 20899 USA

Introduction

Gradient combinatorial approaches to investigating cell-material interactions add a level of complexity because a range of biological data must be gathered and preserved within the tight spatial correlations of the variation of material properties. We can consider two assay methods are capable of meeting stringent spatial requirements: Fluorescence Microscopy (FM) and Laser Capture Microdissection (LCM). Of these two methods, LCM offers the most direct comparisons with traditional biological assays in that cells can be harvested individually, or in small clusters with a spatial precision on the order of microns and the cells can be analyzed with a range of assay methods. LCM is a relatively new technology and its development has been pushed towards precision histological sectioning, but it is ideal for library assay. On the other hand, fully automated fluorescence microscopy is within the scope of many laboratories.

On the scale of complexity, synthetic or processed polymeric materials are rather simple, nevertheless assessing the range of complex interplay that characterizes the interactions of cells with synthetic materials presents a challenge for tissue engineering. The gradient library approach to combinatorial and high-throughput analysis of these interactions provides a powerful tool. Just as the combinatorial methodology has changed the paradigm of pharmaceutical assessment, and it is moving to have a similar impact on traditional materials science, we are working at the interface. Unlike the combinatorial synthetic approach, the NIST combinatorial approach emphasizes high-throughput methods to generate experimental data over the multi-parameter space.

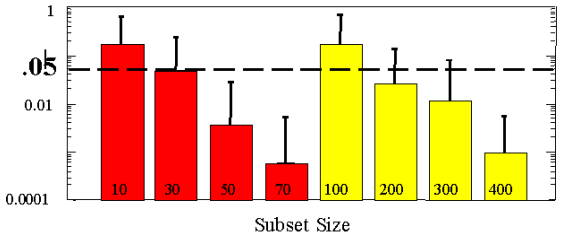
Experiment and Results

This paper will consider the analysis of microscopy data. In our approach, protocols were established for collecting focused images across an x-y grid that spans a gradient film sample in either a contiguous or non-contiguous manner. Various layers of data were collected at each x-y location, for example, by imaging multiple fluorescent channels. For a typical library the number of images collected escalates into thousands, and by necessity image analysis for quantification was automated. In addition automation reduces sampling bias by removing human judgment from the selection of representative assay fields and automation allows more fields of view to be analyzed, which further reduces sampling bias and strengthens the statistical significance of the data. Finally, automation provides spatially resolved maps of single cell metrics. This is critical on two levels: providing correlation of cell metrics one with another, and correlating all cell metrics within a spatially resolved map of material parameters.

In order to address the issue of sampling bias the simplest quantitative metric, counting cells, was employed to analyze four control samples. Our samples were cultured with MC3T3 osteoblasts under standard conditions. Cells were fixed in 3.7% paraformaldehyde and stained with DAPI. Experiments sought to quantify the extent of data collection necessary to limit sampling errors.

Initially we considered:

- Four samples: 2 at relatively high cell density, but still very different, 2 at low cell density. We ask, given that these samples have a different number of cells, how many images must be randomly sampled to make the statement with confidence $p < 0.05$?
- Discreet subsets were sampled from the larger data sets and compared using a T-test. This process was repeated 35 times for each discreet subset and p values were analyzed as the dependent variable. Results are shown below.



This exercise identifies sampling errors inherent to any microscopy technique. The most reliable experimental designs will take this into account and follow proper measures to ensure that sampling errors are minimal and do not compound with experimental errors.

The next level of complexity in analyzing fluorescence microscopy data emerges when multiple stains and fluorescent channels are used. For example, fixed cells can be stained using Texas Red maleimide, which stains the entire cell membrane, and counterstained with DAPI, which stains only the nucleus. Images are collected at every x-y position on blue (DAPI) and red fluorescent channels, thus each channel correlates perfectly with the other. Analysis of images from the red channel provides information pertaining to cell morphology. Alone, this data would not be accurate because analysis software cannot distinguish single cells from overlapping cells. By analysing the two channels together, morphology data was corrected and normalized per cell to allow capture of single-cell metrics across large samples.

As we implement the screening of materials libraries via single-cell metrologies, we face unique statistical challenges. Distributions of single cell metrics are not analogous to standard error measurements of uncertainty. Each data point represents the unique response of an individual cell, and thus the distribution itself can contain information as specific cells process information. In terms of a robust mathematical analysis, we demonstrate that parametric descriptors, suitable for standard statistical treatments, can fail to characterize response functions of cell populations, often revealing contradictory null hypotheses, and nonparametric tests must also be considered.