UNIVERSITY OF CALIFORNIA, SAN DIEGO

Image Cytometry Classifiers: Isolation of Cancer Cells from Blood and Morphometric Characterization of Cytotoxicity in High Content Screening

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Behrad Azimi

Committee in charge:

David A. Gough, Chair
Mark Mercola, Co-Chair
Sadik C. Esener
Xiaohua Huang
Jeffrey H. Price
Shyni Varghese

2010

The dissertation of Behrad Azimi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2010

DEDICATION

***T**o my parents,*

for planting the seed of curiosity in my heart.

EPIGRAPH

Cerca Trova

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

All of the research presented in this dissertation was performed under the guidance of my research advisor Dr. Jeffrey H. Price at the Sanford-Burnham Medical Research Institute. I am forever grateful to Dr. Price for his exceptional support and guidance throughout the years. I am also thankful for the excellent counsel that my academic advisor and dissertation committee chair, Dr. David A. Gough, provided me regarding academic matters at the Department of Bioengineering. I would also like to thank my dissertation committee co-chair, Dr. Mark Mercola, and committee members, Drs. Sadik C. Esener, Xiaohua Huang and Shyni Varghese, for their constructive comments and inputs.

This dissertation would not have been possible without the help and support of all my lab members who helped me with my research and made the lab as friendly and collaborative as possible. In particular, I would like to thank Dr. Ramses Agustin for his invaluable guidance and collaboration; David Charlot for his generous assistance with hardware automation; Ryan Kast, Eman Elmi and Wenting Shih for sample preparation; Joseph Russo, Diego Calzolari, Ilaria Bruschi, Mirco Guigli and Francesco Bedogni and Drs. David Reiner, Derek Fuller, Albert Kellner, Edward Hunter and Mahsa Ranji for the great many educational, stimulating and entertaining conversations.

I am also grateful to the members of the Conrad Prebys Center for Chemical Genomics for their collaboration and generous support. I would specially like to thank Dr. Susanne Heynen-Genel for her continued support and constructive suggestions and Michelle Sauer, Loribelle Milan and Iveta Kalcheva for their help with biological preparations.

Chapter 5, in full, is a reprint of the material as it appears in Hybrid median filter background estimator for correcting distortions in microtiter plate data; Paul J. Bushway, Behrad Azimi, Susanne Heynen-Genel, Jeffrey H. Price, and Mark Mercola; ASSAY and Drug Development Technologies, 8(2), 238–250, April 2010. The dissertation author was one of the primary co-investigators and co-authors of this paper. The paper was produced in collaboration with Paul Bushway, whose tireless efforts were key to driving the Hybrid Median Filter project forward.

Other collaborators at the Sanford-Burnham Medical Research Institute who contributed to this dissertation include Dr. Ken Yip who prepared biological samples used in the preliminary toxicity studies. I would also like to thank my committee co-chair Dr. Mark Mercola and Drs. Pamela Itkin-Ansari and Fred Levine for their empowering guidance and support. I am especially grateful for the administrative support of Lisa O'Brien which made my experience at Sanford-Burnham as hassle-free as possible.

I would also like to thank all the employees of Vala Sciences Inc. for assistance and support in running instruments and software that were imperative for the success of my research. In particular, I would like to thank Casey Laris for his encouragement and his help with image analysis; Piyush Gehalot for his hardware support and Dr. Ross Whittaker for biological sample preparation.

I am very grateful to my high school teachers for their dedication to motivating and promoting students to pursue their dreams against all odds. Specifically, I would like to thank Hazem Faripour, Gholamreza Arasteh and Bahman Eslahpazir and Drs. Sherwin Vakili, Behrang Nooralishahi and Arash Afraz for their inspiration and persistent mentorship.

Lastly, I am grateful to all my friends who made my life more interesting and enjoyable. My most sincere gratitude is for Lilian Shaari, whose love, support and encouragements have been the cornerstones of my strength. Throughout the years, she consistently kept me focused on the bigger picture in life. For that, I am eternally thankful.

VITA

2004-2005	Teaching Assistant, University of California, San Diego
2005	B. S., *Warren College Honors*, Bioengineering-Biotechnology, University of California, San Diego
2007	M. S., Bioengineering, University of California, San Diego
2005-2010	Research Assistant, Conrad Prebys Center for Chemical Genomics, Sanford-Burnham Medical Research Institute
2007-2009	Graduate Teaching Assistant, University of California, San Diego
2010	Ph. D., Bioengineering, University of California, San Diego

PUBLICATIONS

Ramses M. Agustin, **Behrad Azimi** and Jeffrey H. Price, "Automated neural network classifiers for identifying micrometastases in peripheral blood via high-throughput microscopy", *Proceeds of SPIE (International Society of Optical Engineering*, 6441, 64410M, 2007.

David Charlot, Victor Campa, **Behrad Azimi**, Mark Mercola, Randall Ingermanson, Patrick McDonough, Jeffrey H. Price, "Automated Calcium Measurements in Live Cardiomyocytes", *Invited Refereed Meeting Paper, Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium*, 14(17), 316 - 319, May 2008

Paul J. Bushway*, **Behrad Azimi***, Susanne Heynen-Genel, Jeffrey H. Price, Mark Mercola, "Hybrid Median Filter Background Estimator for Correcting Distortions in Microtiter Plate Data.", *Assay and Drug Development Technology*, April 2010, 8(2): 238-250. *These authors contributed equally to this work.

Alice Kiselyuk, **Behrad Azimi**, Jeffrey H. Price, Mark Mercola, "Phenothiazine Neuroleptics Signal To The Human Insulin Promoter As Revealed By A Novel High-Throughput Screen", *Journal of Biomolecular Screening*, Manuscript Accepted.

ABSTRACT OF THE DISSERTATION

Image Cytometry Classifiers: Isolation of Cancer Cells from Blood and Morphometric Characterization of Cytotoxicity in High Content Screening

by

Behrad Azimi

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2010

David A. Gough, Chair
Mark Mercola, Co-Chair

Advances in microscopy have made the automatic acquisition of large volumes of images from biological specimens a routine. In high content screening (HCS), databases of cytometric measurements from cell images can be easily made. The transformation of such cytometric data into biological knowledge is the general aim of this dissertation.

One such transformation can help detect circulating tumor cells (CTCs) in blood based only on their morphology. CTCs travel from a primary tumor to settle in a metastatic site and form more tumors. Enabling the detection of CTCs can help characterize metastatic behavior and improve cancer diagnosis. Here, the use of neural networks in detecting CTCs based on nuclear morphology alone was explored (Chapter 1). Furthermore, various methods are proposed and tested to systematically improve the performance of such classifiers (Chapter 2). Additionally, since the method is novel, it could detect cells that may be key to the study of cancer progression but have been missed by current techniques.

Therefore, it is also important to physically isolate the resultant cells to do downstream experimentation. To accomplish this, a microscopy cell sorting instrument is developed and tested that retrieves desired cells directly from the surface of a microscopy slide with relatively high throughput (Chapter 3).

Another transformation of cytometric data to biological knowledge was explored by attempting to gauge cytotoxicity using morphology alone. In HCS, hundreds of thousands of compounds are tested on cells and a desired biological response is assayed to find drug candidates. Assessing toxicity in that stage would stop the advancement of toxic candidates into more expensive testing and clinical trials. Implementing cytotoxicity assays into screens, however, is prohibitively costly. A method for mimicking the costly cytotoxicity measurements in large screens by using morphology alone is proposed and studied (Chapter 4). The method is then applied to a small screen and its performance is measured. Further, distortions often exist in screen data because of problems with preparation of microtiter plates. A median-based method has been developed and tested here to remedy this problem and make screen data statistically more viable (Chapter 5).

Chapter 1

Neural Network Classification of Circulating Tumor Cells in Blood

1.1 Introduction

According to the National Cancer Institute (NCI), about 1 in 8 women in the United States will develop breast cancer during her lifetime. Breast cancer ranks just behind lung cancer as the leading cause of cancer death for women and in the leading cause of death in women ages 15-54. Between 1973 and 1991, the incidence of breast cancer rose 24% in the United States, due mainly to the increased use of mammography, but has steadily decreased since then. Moreover, during the same period, the mortality rate did not increase and has been steadily decreasing since the late 1990s. Early detection and prognostic screening probably played roles in lowering these statistics. In fact, according to the American Cancer Society's facts and figures from 2006-2007, mammography will detect 80%-90% of breast cancers in women without symptoms. The five-year survival rate for breast cancer is 98% for women who were diagnosed with localized breast cancer, and 89% for women who were diagnosed when the cancer was in a regionalized stage.

Current clinical methods for detecting breast cancer include palpation (either through a clinical breast exam or monthly self-examination), mammography, MRI, and ultrasound to evaluate suspicious lumps. Once a breast anomaly has

been identified, a tissue sample is biopsied to determine whether the lump is cancerous. According to Bjurstam, approximately 20% of breast biopsies turn out to be breast cancer [12]. Prognosis and proper treatment depend on a number of factors such as tumor size, location within the breast, and invasiveness (as assessed by the presence of cancer in the lymph nodes or other parts of the body).

Mammography is considered the gold standard for diagnosing breast cancer, and there is substantial active research on finding methods to detect breast cancer cells in the circulation as a possible diagnostic/prognostic tool. Research efforts to identify CTCs in the bone marrow and/or peripheral blood have utilized PCR, soluble protein assays, immunomagnetic sorting, fluorescence-activated cell sorting (or flow cytometry), and image cytometry [61, 60, 72, 83, 95, 13, 107].

Polymerase chain reaction (PCR) relies on selective expansion of a known DNA sequence, delineated by a set of olignucleotide primers, by a modified DNA polymerase. A gene may be amplified 106-109-fold by PCR, and detection specificity is determined by sequencing the amplified DNA product. Reverse transcriptase PCR (RT-PCR) allows detection of RNA by an initial conversion of RNA to DNA catalyzed by enzyme reverse transcriptase. Targets for RT-PCR detection have included the oncogene HER2/neu [28, 104], Cytokeratin 19 [78, 94, 96, 117, 71], as well as other tissue-specific biomarkers [26, 66, 118].

Cell separation and sorting methods have been used to isolate rare cancer cells from blood and bone marrow. In flow cytometry, fluorescently-labeled cells in a hydrodynamically focused fluid stream are excited by a laser beam. Forward-scatter and one or more measurements of side-scatter are collected for each cell. Suspicious cells are identified by gating their fluorescent intensities. Flow cytometry for detecting circulating tumor cells in blood tends to produce high numbers of false positives due to specimen autofluorescence, nonspecific staining, and cell aggregates. Furthermore, multiple handling steps and cell transfers through the complex capillary system can contribute to sample loss. In contrast, immunomagnetic sorting uses antibodies conjugated to paramagnetic beads. Once beads are bound to surface antigens on the cell, a magnet can be used to capture cells for sorting and enrichment. Due to the size of the magnetic beads used, immunomagnetic

sorting is limited to use against surface antigens of the cells.

While these alternative detection assays will not likely replace mammography for diagnosis, they may be useful for characterizing the cells escaping the solid tumor(s), assessing disease progression and for monitoring minimal residual disease. Furthermore, the afore-mentioned research technologies rely on tissue-specific biomarkers. A biomarker is any substance, such as proteins, antibodies, genes, as well as DNA or RNA fragments, whose detection may be used to measure the progress of a disease or its treatment. One of the earliest breast-associated biomarkers to be used was CA15.3 against MUC1 which is frequently overexpressed in malignant breast cancer. Zimmerman et al. tested 114 patient samples that were known to be either benign or malignant. Positive staining of greater than 10% of tumor cells was found in 97% of breast carcinoma cases and in 90% of adenocarcinomas overall. Four of 40 (10%) cases with benign mesotheliomas stained as false positives [119]. Cheung et al. reported that CA15.3, CEA, and ESR were the best validated combination for detection, albeit for late-stage cancer that had already been diagnosed and for which the presence of metastatic tumors had been detected by other means [22]. It was eventually found that the utility of CA15.3 for detecting early stage disease was low, and that 20-30% of women with benign tumors will have elevated expression [99]. In other words, preoperative measurements of CA15.3 concentrations can't be used for screening or diagnosis. However, elevated concentrations of CA15.3 pre- or postoperatively and at the time of initial relapse have been correlated with patient outcome [29].

These technologies have shown promising initial findings and have been translated to the clinical setting in at least one case for following and predicting the likelihood of disease progression. To date, only the CellSearch®Epithelial Cell Kit has been FDA-approved for detecting circulating tumor cells in blood [1]. This FDA-approved technology uses the most common biomarkers investigated for epithelial carcinomas: cytokeratins and EpCAM. Cytokeratins (CK) are intermediate filaments comprising the cytoskeleton of epithelial cells. In humans, twenty cytokeratins have been identified and classified roughly according to decreasing molecular weight. Cytokeratin 19 (CK19) expression is commonly asso-

ciated with breast carcinomas, with over 95% of tumor cells in all human primary and metastatic tumors staining positive for CK19 [4]. The epithelial adhesion molecule (EpCAM) is a cell surface glycoprotein involved with cell-to-cell adhesions. EpCAM is expressed at a basal level in normal adult and neonatal epithelia but is overexpressed in epithelial carcinomas, including breast cancers [82, 106]. The expression of EpCAM is down-regulated in CTCs as compared to primary tumor cells [88].

One of the obvious limitations of this technology is that circulating tumor cells are not detectable in all cancer patients–even in some with aggressive metastatic disease. For example, Kim et al. found CTCs in only 8 of 29 women with metastatic disease[60], while Witzig et al. found CK+ cells in only 21 of 75 cancer patients [107]. It is interesting to note the inconsistencies in the detection of CTCs using cytokeratins and/or EpCAM. Kasimir-Bauer et al. [55] studied samples from 28 patients using two methods in parallel: immunocytochemistry (ICC) alone and immunomagnetic sorting followed by ICC (IMS/ICC). In 6 of 28 patients, CTCs were found by both assays. Sixteen were positive for CTCs by ICC but negative by IMS/ICC. Finally, six were negative by ICC but positive by IMS/ICC. Thurm et al. analyzed Ficoll-enriched bone marrow aspirates that were either double-labeled EpCAM+CK or immunomagnetically sorted against EpCAM then CK stained. Ten of 35 samples were CK+ but none coexpressed EpCAM. Only 2 of 27 specimens enriched by IMS coexpressed cytokeratin [101]. Most recently, a prospective study of the only FDA-approved method, CellSearch®, found that among 92 breast cancer patients with metastesis, only 70% had a positive CTC count [89].This lack of sensitivity is explained by citing the heterogeneity of EpCAM expression [106, 88].

What complicates things further is that not all cancer cells from a patient express biomarkers at the same level. That is, cancer cells in an individual patient will have an expression profile (or distribution) that will be different from the expression profiles from other patients. Moreover, subpopulations of cancer cells from the same patient biopsy often express different proteins. The genetic instability of cancer that enables conversion to an aggressive phenotype that may

demonstrate further alterations in gene expression, and heterogeneity of the proteins expressed[32, 79, 87, 102, 113] is the reason that finding an absolute set of differential expression markers is difficult, if not impossible, for accurate detection of CTCs. Even if a subset of CTCs will always be found, seeking them based on a particular biomarker or set of biomarkers likely leave a potentially large and important set of CTCs unnoticed.

For this reason, it is important to find a method that is independent of biomarker expression for detecting CTCs. Preliminary studies were done in our lab to detect breast cancer cells by staining for cytokeratins in an *in vitro* spiked model of CTC in peripheral blood consisted of cancer cells mixed with mononuclear white blood cells (mWBCs). Expectedly, the cancer cells were heterogeneous for cytokeratins and were not all identified. Through these experiments, it was noted that some cancer cells appeared larger and more irregularly shaped than mWBCs. We thus decided to explore whether detection of CTCs could be performed using nuclear morphology alone, independent of tissue-specific biomarkers. To test this, we first used manual (or interactive) classification, which involved a user setting gates on empirically-determined size and shape features to select a subpopulation of suspicious cells and manually reviewing the resulting gallery of cell images to eliminate false positives. We then, trained neural networks to select suspicious cells and compared the sensitivity and specificity with manual gating.

Over the past decade, neural network classifiers have been put to greater use in biology and medicine, e.g. in analyzing microarray[36, 37, 48, 68], mass spectrometry[63, 68], flow cytometry[67], ultrasound[21, 52], MRI[73], and image cytometry[20, 62] data. Particularly for clinical applications related to cancer, neural networks have been applied to risk assessment in prostate cancer[2, 75, 76], survival prediction in breast cancer[15, 16] as well as the identification of microcalcification patterns in mammograms [19, 47, 49, 54, 110, 111]. Some of the neural networks evaluated included qualitative features (e.g. diet, race, and general health [2] or age, race, and symptoms[16]) or manually segmented objects (from digitized mammograms [47]). Furthermore, with the exception of the detection of activated lymphocytes [67], none of these studies performed exhaustive cell-by-cell

analysis.

1.2 Materials and Methods

1.2.1 Preliminary studies to explore nuclear morphology

While studying cytokeratin expression in cancer cells (results not shown), images were acquired to demonstrate differences in nuclear morphology among the different cell lines. Cancer cells from each line were grown as adherent cultures on #1.5 coverslips (Fisher Scientific, Tustin, CA), fixed with 4% paraformaldehyde for 1 hour, then counterstained with DAPI for fluorescent imaging. Adapted from Hamada[39], the preparation of DAPI solution consisted of 75ng/ml DAPI, 10nM Tris, 10nM EDTA, 100nM NaCl, and 2% 2-mercaptoethanol, prior to mounting and sealing to 3x1 microscope slides (Fisher).

1.2.2 *in vitro* Mixture Models

Mononucleated white blood cells were isolated from a sample of whole blood from a healthy individual by density gradient centrifugation using HISTOPAQUE-1077 (Sigma-Aldrich, St. Louis, MO).

Each of six human breast cancer cell lines were cultured as adherent monolayers and spiked into human blood. These cell lines were HCC1395, MCF7, MDA-MB-231, MDA-MB-435, MDA-MB-468, and T47D (ATCC, Manassas, VA). The cells were first stained with CellTracker Orange (Molecular Probes, Eugene, OR) as a gold standard against which to grade the performance of classification via nuclear features. After detaching from culture dishes and resuspended in PBS, cancer cells were mixed with mWBCs at ratios of 1:10 (HCC1395 only), 1:102, or 1:103 and fixed with 4% paraformaldehyde for 1 hour.

Cell mixtures were centrifuged as a monolayer onto silane-coated slides (Sigma-Aldrich) for imaging. Finally, nuclei were counterstained with DAPI solution, prior to the application and sealing of coverslips.

1.2.3 Instrumentation and data collection

A system developed in our laboratory that is functionally equivalent to a Q3DM/Beckman Coulter EIDAQ/IC 100 image cytometer includes the following components for image acquisition, automated stage movement, and autofocus. A red LED (HWPT-MH00, peak wavelength 626nm, Agilent Technologies, Palo Alto, CA) provides phase-contrast illumination, strobed under computer control, on a Nikon Eclipse TE300 inverted microscope. Strobing the LED allows for essentially instantaneous response when switching between transmitted light and epifluorescence. Primary optics include a Nikon LWD 0.52NA condenser and Nikon Plan Fluor 20X/0.50NA Ph1 DLL objective. For epifluorescence excitation, we use a 100W mercury vapor short arc lamp (OSRAM HBO 103 W/2) in a Nikon LH-M100CB-1 arc lamp housing (Nikon Instruments, Lewisville, TX) with a Uniblitz VS25S2ZM1R1 shutter (Vincent Associates, Rochester, NY).

Lateral motion control of a motorized stage (99S008-N23, Ludl Electronics Products Ltd, Hawthrown, NY) is performed under computer control by stepper motors (PCI-7324, National Instruments) at a minimum step size of 125nm. A Polytec, Inc. (Auburn, MA) 350-nm range piezoelectric objective positioner (PI-FOC) and E-S810.L0 closed-loop controller, monitored by a National Instruments PCI-6031E data acquisition board, allow for computer-controlled focus positioning. Calculation of best focus is done in hardware with a Q3DM, Inc. (San Diego, CA) Afx-3000 autofocus circuit that was originally developed in the Price lab [14].

Filter cube sets for DAPI and Spectrum Orange (for CellTracker Orange) were used for fluorescent imaging (Chroma Technology, Brattleboro, VT). Images were acquired with a Cohu 3000-6612 Progressive Scan CCD RS170 camera (San Diego, CA) and digitized to 640x480 pixels2 with a National Instruments IMAQ PCI-1424 multi-channel frame grabber (Austin, TX). Under 20X magnification, each pixel is 0.517344 μm x 0.517344 μm or 0.267645 μm^2.

Software to drive the image cytometer (CytoShop v1.6.512.648 and CytoShop v2.1 Build 0.4) was kindly donated by Q3DM and Beckman Coulter. CytoShop enables straightforward image acquisition of arbitrarily large areas of a slide. Fast image segmentation in CytoShop is performed on fluorescent images,

Table 1.1: The *a priori* feature set. This set is consisted of 7 nuclear features (morphometric and fluorimetric) from CytoShop which were used as inputs to the *a priori* neural networks.

Feature	**Description**
Area	Area of the nucleus in square microns
Perimeter	Perimeter of the nucleus
Wiggle	Perimeter divided by area of the nucleus
Integ_gs	Sum of grayscale intensities of pixels in the nucleus
StdDev	Standard deviation of intensities of nuclear pixels
CM3	Third central moment of the intensities of nuclear pixels
CM4	Fourth central moment of intensities of nuclear pixels

using a method that is largely independent of fluorescence intensity [84]. Non-linear least-squares-designed finite impulse response (FIR) filters greatly enhance contrast between object and background to enable automatic histogram thresholding. After segmentation, a set of about 75 features based on pixel intensity and object morphology are measured on each cell. Table 1.1 summarizes nuclear features expected to be pertinent for classification. The empirically-chosen features are designated with a check in the right column.

Initially, nuclear area and nuclear wiggle (a perimeter-area ratio), were examined for distinguishing between cancer and normal using size and shape. Using simple thresholds on these two features, subpopulations of cells were quickly delineated. Once a subset of cells is defined, CytoShop creates a gallery that is quickly displayed for visual review of the cells. As mentioned previously, the gold standard for assessing cancer origin for the *in vitro* models is CellTracker Orange. After gating on CellTracker Orange signal, the montage of cell images from the positive population is manually examined and artifacts (e.g. CellTracker signal not actually localized within the cell boundary) are removed.

1.2.4 Neural Network Classifiers

Cytoshop measures more than thirty morphometric and fluorometric features per object segmented from images. These measurements can be exported for analysis in third-party programs, and the framework for training and evaluating a neural network was programmed in MATLAB (The MathWorks, Inc., Natick, MA) using the Neural Network Toolbox. For the purposes of this paper, we designed neural networks that used a subset of seven features thought *a priori* useful for classifications as inputs.

The general architecture of the feed-forward neural networks described here is shown in Figure 1.1. Output of each neuron depends on a weighted sum of its inputs but does not necessarily need to be linear or continuous. All neurons in our analysis used a tan-sigmoid activation function to mimic the activation of biologic neurons. To use the Neural Network Toolbox in MATLAB, the input features were rescaled to the range. The result from the output node is also scaled [0-1], and a final decision between cancer (1) and normal (0) was made using a simple post-processing threshold of 0.5. Values below 0.5 are considered to be non-cancer cells; and values of 0.5 and above were considered to be cancer. The error between the output classification decision and the target classification is calculated as the mean squared error (MSE), given by Equation 1.1 where Q is the size of the training set, k is the index of the training vector, T is the target vector, and A is the output vector for the network prior to post-processing. The neural networks were trained via backpropagation. Essentially, training consists of iteratively running a subset of the data through the neural network, comparing the network's output to the target (desired) output, and adjusting network weights to achieve a desired minimum for the mean squared error. This process is computationally-intensive and can require a long time for convergence. At each node, a local error is calculated based on what its output should have been to achieve the desired output. Weights are adjusted to minimize the local error, and how these weights change also depends on local errors at the previous layer. Neurons connected by stronger weights have a greater effect on local errors downstream. Consequently, the learning through error minimization propagates backwards through the network. The process of backpropagation relies

on calculating the gradient of the error in the network with respect to network weights. Algorithms based directly or indirectly on this gradient are then be used to minimize the error.

$$MSE = \frac{1}{Q}\sum_{k=1}^{Q}\Big(T(k) - a(k)\Big)^2 \tag{1.1}$$

The neural networks used here were trained using a technique known as resilient backpropagation. In preliminary studies, we compared resilient backpropagation to a gradient descent method. For the latter method, weights are changed every training iteration (epoch) based on the derivative of performance with respect to each of the weights. If we call a particular weight X and the next update to that weight δX, then δX is given by Equation 1.2 where lr is a constant.

$$\Delta X = lr\frac{dMSE}{dX} \tag{1.2}$$

In contrast, when training is through resilient backpropagation, a variable learning rate matrix LR is used instead of a single fixed value. After initialization, LR is updated at each iteration based on whether or not the performance gradient has changed sign. That is, if the derivative of performance with respect to a given weight changes sign, the corresponding element in LR will be decreased by some constant. In the case of no sign change, the corresponding element is increased by a different constant in order to stimulate change. Changes in weight matrix are calculated according to 1.3.

$$\Delta X = lr\left(\frac{dMSE}{dX}\right) \div \left|\frac{dMSE}{dX}\right| \tag{1.3}$$

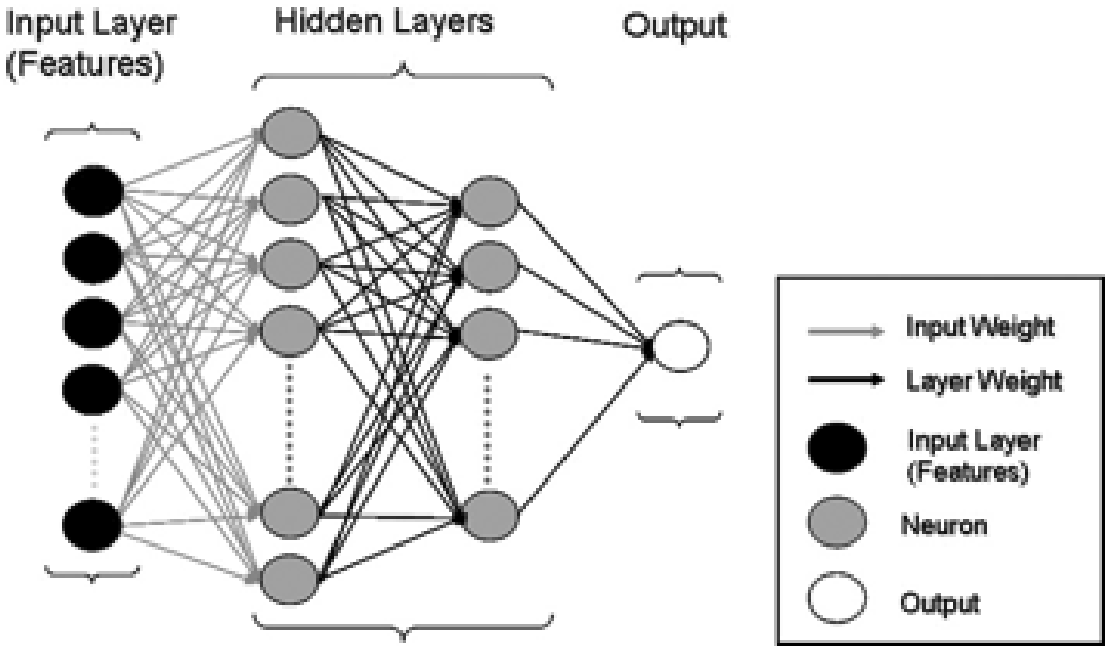


Figure 1.1: We used a seven-feature input layer, and ten and three nodes for the next two hidden layers, respectively. The output from the final single-node layer is the binary cancer/non-cancer decision.

Training continues until the MSE reaches a specified tolerance, the gradient reaches a specified minimum, or until a certain number of epochs have passed. Depending on which *in vitro* model was trained, the minimum gradient was 10-12 or zero.

We trained and evaluated several neural networks with differing numbers of nodes, trained by gradient descent and resilient backpropagation. For networks with the same architecture, training by resilient backpropagation converged at least an order of magnitude faster and classification performance was identical. For large datasets with 100,000 or more cells, training may require up to a few day to converge, so the training speed increase realized by resilient backpropagation was important.

For our initial experiments, we trained and evaluated neural networks using data sets from HCC1395 spiked models. The data were randomly split into a training set and a testing set (to characterize classification performance), comprising 30% and 70% of the original data, respectively. The random subsets were generated such that the same ratio between cancer and normal cells in the original data was maintained. Neural networks were trained for 20,000 epochs. With this number of training iterations, the minimum gradient achieved was on the order of 10^{-12}. Afterwards, neural networks were created with data from the other spiked models.

Finally, a single neural network was designed using a pool of all the *in vitro* data. That is, the same architecture was used but trained with a combination of the HCC1395, MDA-MB-231, MDA-MB-435, MDA-MB-468, MCF7, and T47D spiked blood data. Again, a random distribution of 70% from each data set was combined and used for training, with ratios between cancer and normal maintained. In generating performance statistics, however, the neural network was evaluated on the remaining 30% of each data set separately in order to compare how well the network could classify a variety of different breast cancer cells from normal.

1.2.5 Performance Evaluation

The possible outcomes of a binary classifier are illustrated in 1.2. Cancer cells that are correctly evaluated by the classifier are the true positives. Cancer cells that are incorrectly evaluated as non-cancer are false negatives. Similarly, non-cancer cells correctly classified are true negatives, while those incorrectly classified are false positives. In quantifying these populations, the neural network was trained on one set of data and the performance was evaluated on another set not previously used for training.

Further performance measures are defined in Equation 1.4 to evaluate classification. Sensitivity and specificity are two common measures of classification performance and reliability. Sensitivity is the probability of a positive finding being correct, while specificity is a measure of the proportion of actual negatives in the entire sample. We also quantify accuracy, defined as the probability that a cell in the subpopulation classified as cancer truly is cancer.

Table 1.2: Possible outcomes of a binary classifier and definition of basic classification performance measurements.

		Gold Standard	
		+	-
Selected by Algorithm	+	$TruePositive$	$FalsePositive$
	-	$FalseNegative$	$TrueNegative$

$$
\begin{aligned}
Accuracy = & \quad \frac{TruePositive}{TruePositive + FalsePositive} \times 100\% \\
Sensitivity = & \quad \frac{TruePositive}{TruePositive + FalseNegative} \times 100\% \\
Specificity = & \quad \frac{TrueNegative}{TrueNegative + FalsePositive} \times 100\%
\end{aligned}
\tag{1.4}
$$

1.3 Results

1.3.1 Observed differences in nuclear morphology of breast cancer cell lines vs. mWBCs

We first reviewed galleries of nuclear images of the six cell lines and mWBCs. Slides with nuclei of each of the breast cancer cell lines and one with mWBCs were imaged and the morphometric features were measured. Example nuclei are shown in screenshots of the galleries from CytoShop in Figure 1.2, each of which are displayed at the same magnification with different numbers of pixels. As explained in Materials and Methods, each image pixel recorded corresponds to 0.267645 μm^2. For the image montages, CytoShop created a bounding box around each nucleus that may be different based on mean size in the population. For HCC1395, the bounding box was 119x119 $pixels^2$; for MDA-MB-231, MDA-MB-435, MCF7, and T47D, the bounding boxes were 79x79 $pixels^2$; for MDA-MB-468, the box was 109x109 $pixels^2$; and for mWBCs, the bounding box was 49x53 $pixels^2$. The mWBCs tend more to be smaller and more uniform in size and shape. Quantitative comparisons of area and wiggle are shown in Figure 1.3. HCC1395 nuclei are the largest and have the smallest wiggle (are the most irregular in shape); in Figure 1.2, they appear more lobular and generally more irregularly shaped. The MCF7

and T47D nuclei also appear larger in Figure 1.2 than mWBCs, and are larger on average in Figure 1.3, but the tails of the distributions overlap with mWBCs. The MCF7 and T47D nuclei also have smaller wiggle (are more irregularly shaped) by the measurements in Figure 1.3. The MDA-MB-468 nuclei are larger, with some overlap in the tail of the distribution, than mWBCs by Figure 1.3 and have similar wiggle. MDA-MB-231 and MDA-MB-435 cells are more comparable in both size and shape to mWBCs than the other cell lines. These results motivated us to hypothesize that cancer cells may be distinguished from mWBCs based on their nuclear morphology alone. Since HCC1395 cancer cells are dramatically different from mWBCs in their distributions, we chose them as our first model for developing our methods and testing this hypothesis.

We reviewed whether or not there may be a relationship between the clinical characteristics of these cell lines that might be associated with their morphologies. In the Boyden chamber assay for invasiveness, MDA-MB-231 cells were ranked highly invasive, MDA-MB-435 cells were moderately invasive, MDA-MB-468 cells were poorly invasive, MCF7 cells were minimally-to-moderately invasive, and T47D were not invasive [100]. In a different study, HCC1395 cells are also labeled as invasive, but they were not ranked in comparison to the first five cell lines [35]. There may be a positive correlation between invasiveness and wiggle, and an inverse correlation between invasiveness and area. That is, small round cancer cells may be more invasive than large irregular cancer cells. HCC1395 cancer cells appear may be an exception to this rule, but invasiveness was not compared relative to the other cell lines. Thus, morphology may be related to cancer metastasis and it could be important to find a method (e.g., neural networks on a larger feature set) that can detect those cancer cells that look most like mWBCs since they could be the most invasive.

1.3.2 Manual gating for classification of HCC1395 human breast cancer cells

We began to investigate the differences in size and shape between cancer and normal cells using a spiked mixture model with HCC1395 cancer cells. Clumps

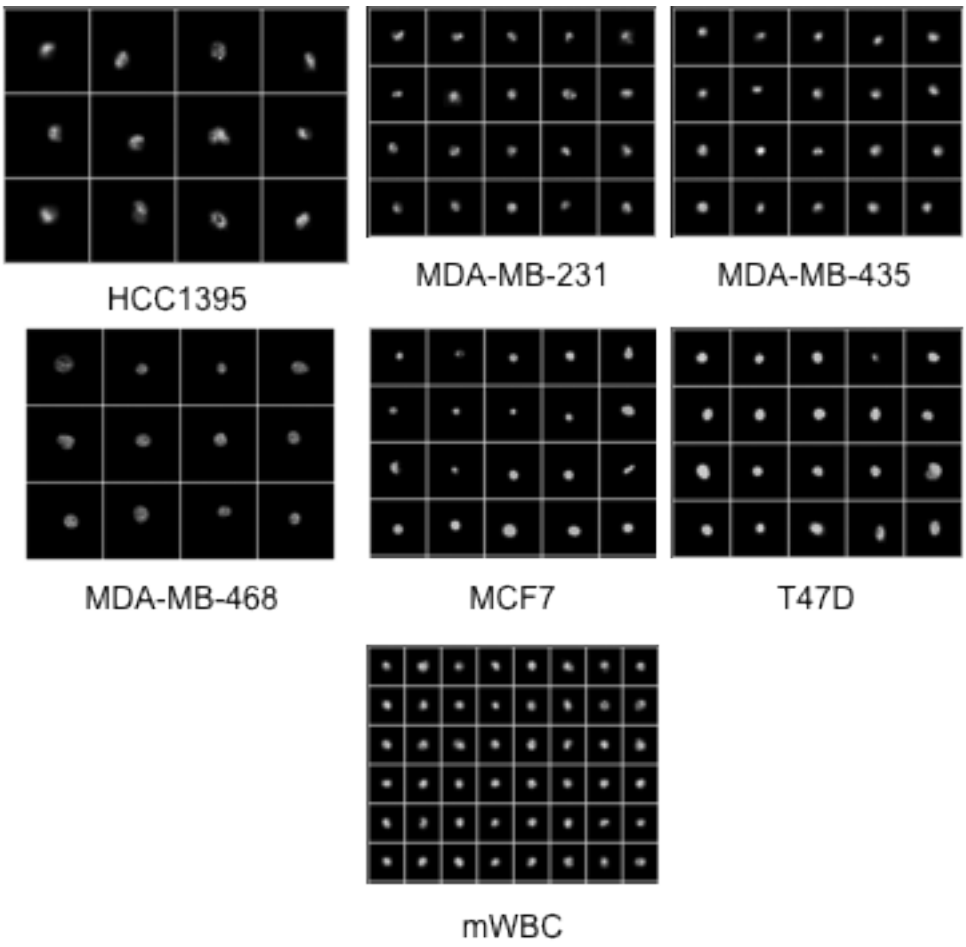


Figure 1.2: Galleries of nuclei images. Cancer cells from each cell line have distributions of size and shape that potentially could be used to distinguish them from normal mononucleated white blood cells

of cells, however, presented a challenge for study since CytoShop's measurements are meaningful only for single cells. We instituted the manual classification scheme illustrated in Figure 1.4 to eliminate clumps and continued to optimize specimen preparation in an attempt to decrease them as much as possible. By excluding objects with large areas and small wiggle, most of the clumps containing three or more nuclei could be successfully excluded. Gates (or thresholds) on nuclear area and nuclear wiggle were then manually set to classify cells as mWBC singlets, mWBC doublets and singlet cancer cells. (See Figure 1.5). Singlet white blood cells comprise most of the objects with low nuclear area. Two adjacent mWBC nuclei (incorrectly segmented as one object) often have a combined area comparable to a singlet cancer cell but with a greater perimeter. Consequently, wiggle may be used further to separate singlet cancer cells from mWBC doublets. The thresholds shown in Figure 1.4 and Figure 1.5 to define the subpopulations of cells are examples to provide an acceptable compromise between sensitivity and specificity.

Overlap in the distributions of area and wiggle for cancer cells vs. doublet mWBCs resulted in false positives (see the gallery of HCC1395 cells classified as cancer in Figure 1.6). Once a subset of suspected cancers has been found, images of the cells can be visually inspected and the doublets interactively removed. For translation to the clinical setting, classification would reduce the number of cells to be reviewed by a human from millions to hundreds. False negatives may also occur, but it's not practical to review and reclaim them. As also evidenced by the large standard deviation bars in area for HCC1395 and T47D in Figure 1.3, cancer cells are also morphologically heterogeneous. Some, like MDA-MB-231 and MDA-MB-435 are almost morphologically indistinguishable from mWBCs by the two features in Figure 1.3 . Large granulocytes (that may not have been separated during enrichment) and a very small number of normal epithelial and/or endothelial cells, thought to have frequencies of 1-2 per 10 ml of blood,[6] could be present as additional sources of false positives. Nevertheless, visual inspection of say a few hundred cells should allow most of the false positives to be identified and removed, thereby improving performance of classification by morphology alone.

Manual gating followed by visual inspection was performed in two sets of

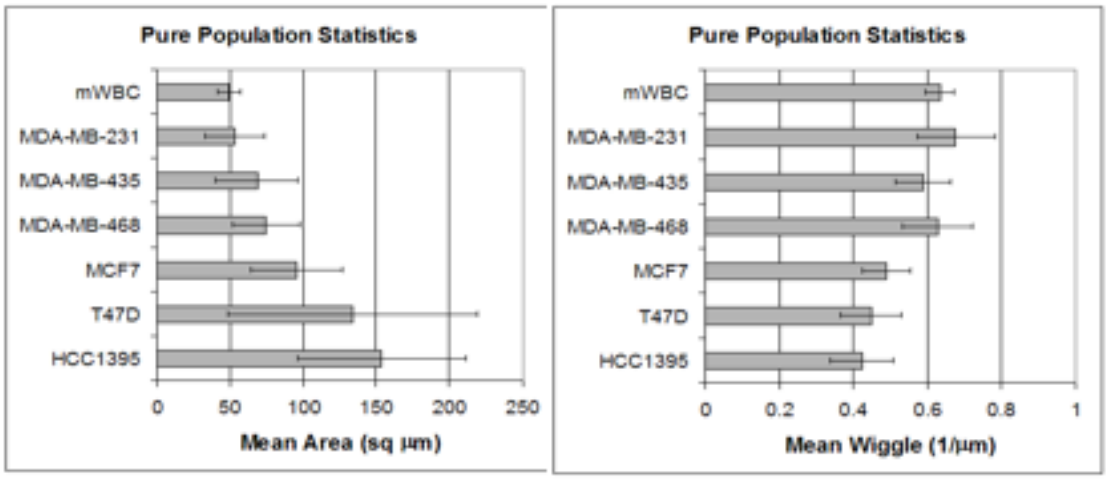


Figure 1.3: Variations in nuclear area and wiggle for the different breast cancer lines compared with mWBCs. Statistics are shown as means (bars) +/- standard deviations (error bars).

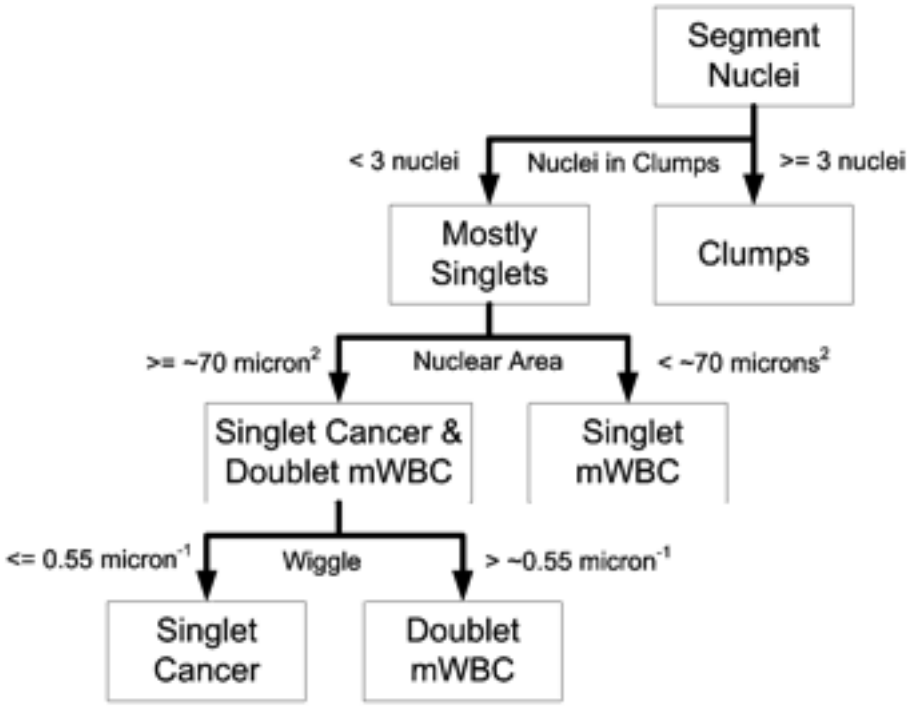


Figure 1.4: Decision tree for removing clumps and classifying cells. Values for gating on nuclear area and nuclear wiggle shown in the diagram are examples.

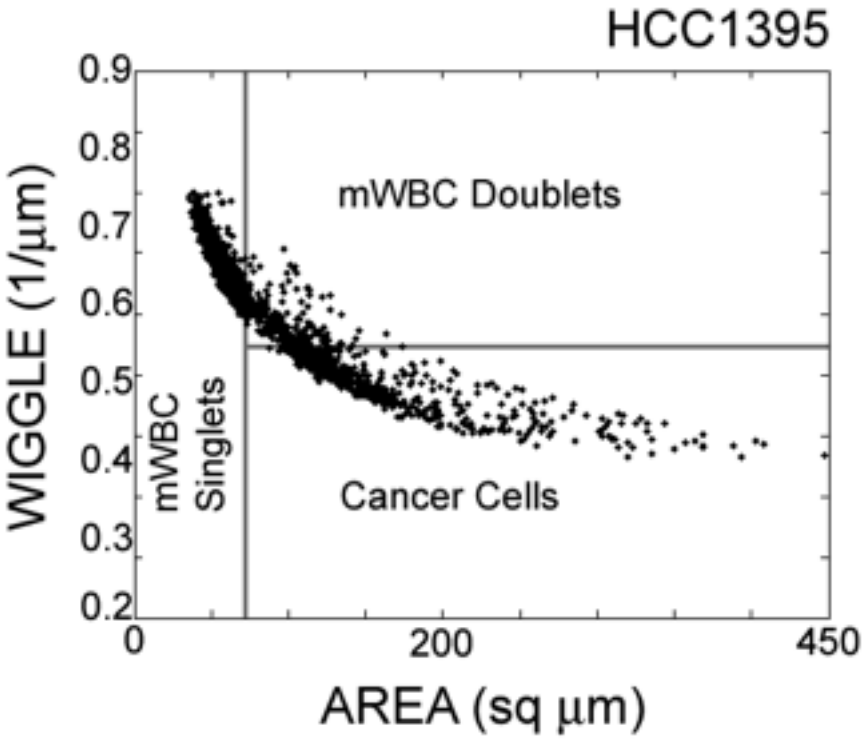


Figure 1.5: Classification of HCC1395 cancer cells using nuclear area and wiggle. This is done using gates (or thresholds) to distinguish between cancer and non-cancer (mononuclear WBC singlets and doublets).

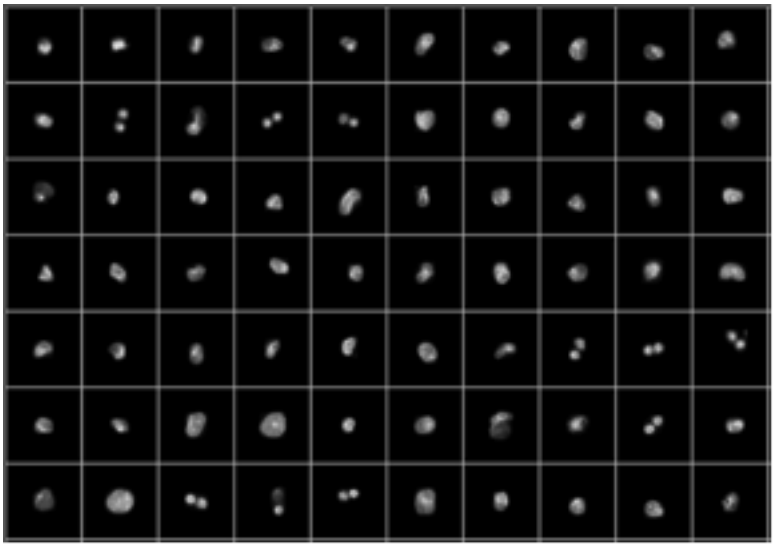


Figure 1.6: An example gallery from the subpopulation classified as cancer. This subpopulation is form the population shown in Figure 1.3 (large nuclei with low wiggle). Visual inspection to eliminate doublets (e.g., row-column: 1-2, 1-4, 1-5, 5-8, 5-9, 5-10, 6-9, 7-3, 7-4 and 7-5) reduces the number of false positives and increases accuracy and sensitivity.

experiments and results are summarized in Table 1.3. The left side of the table is for low-density samples where cells were seeded onto slides at a density or approximately 50,000 $cells/cm^2$. In the right side of the table, medium-density samples were prepared with approximately 250,000 $cells/cm^2$. The numbers of true positives, false positives, true negatives, and false negative classifications and the resulting performance statistics are shown. The results for manual gating followed by visual inspection are reported first, with the numbers for manual gating alone in parentheses. The arrows indicate whether the performance improved or worsened after visual inspection. To reiterate, only the cells positively classified as cancer were visually inspected. Finally, the performance was calculated using the presence of the CellTracker Orange as ground truth label.

Although the number of true positives sometimes decreased with visual inspection, the number of false positives usually decreased significantly. This has the biggest impact on increasing overall classification performance: accuracy, sensitivity, and specificity are 90% or greater after visual inspection. Moreover, only a small fraction of the total number of cells in the blood samples were visually inspected to achieve these results. With this successful manual classification in HCC1395 cells, we next applied similar gating strategies on area and wiggle to classify cancer cells in mixture models using the five other breast cancer cell lines.

1.3.3 Manual gating for classification of cell from five other breast cancer cell lines

The results in the previous section confirmed our observation that HCC1395 breast cancer cells are mostly larger and more misshapen than mWBCs. We next evaluated the performance of manual gating without visual inspection on spiked models using breast cancer cell lines with morphologies more similar to mWBCs. For manual gating applied to spiked models using MDA-MB-231, MDA-MB-435, MCF7, and T47D, gating was adjusted slightly to compensate for their smaller nuclear area and higher nuclear wiggle (rounder shape). For these experiments, the threshold for area was lowered to 55 μm^2, and the threshold for wiggle was raised to 0.60 μm^{-1}. Table 1.4 summarizes performance for several experiments

Table 1.3: Performance of manual gating followed by visual inspection. Manual gating alone is shown in parentheses. There were three experiments each with low-density (50,000 cells/cm2) and with medium-density (250,000 cells/cm2) seeding. The arrows indicate improvement or worsening of performance relative to manual gating alone.

	Low Density Experiments			**Medium Density Experiments**		
Experiment	**1**	**2**	**3**	**4**	**5**	**6**
Cells Inspected	342	399	316	507	713	811
Total # Cells	6,924	6,067	5,994	14,327	17,902	17,112
% Inspected	4.94	6.58	5.27	3.54	3.98	4.74
True Positive	252 (254) ▽	307 (308) ▽	212 (213) ▽	473 (473)	582 (583) ▽	666
False Positive	20 (88) △	15 (91) △	19 (103) △	23 (34) △	28 (130) △	39 (145) △
False Negative	6 (4) ▽	5 (4) ▽	9 (8) ▽	24 (24)	61 (60) ▽	62 (62)
True Negative	6,646 (6,578) △	5,740 (5,664) △	5,754 (5,670) △	13,805 (13,794) △	17,231 (17,129) △	16,345 (16,239) △
Accuracy (%)	92.6 (74.3) △	95.3 (77.2) △	91.8 (67.4) △	93.5 (93.3) △	95.4 (81.8) △	94.5 (82.1) △
Sensitivity (%)	97.7 (98.4) ▽	98.4 (98.7) ▽	95.9 (96.2) ▽	95.2 (95.2)	90.5 (90.6) ▽	91.5 (91.5)
Specificity (%)	99.7 (98.7) △	99.7 (99.2) △	99.7 (98.2) △	99.8 (99.7) △	99.8 (99.2) △	99.8 (99.1) △

with these four additional cell lines. Except for the experiment with MDA-MB-468, accuracies and sensitivities tended to be very poor, indicating many false positives and false negatives. The specificities, or true negative rates, were high as expected for rare classification. Manual gating performance was thus much poorer than for HCC1395 cells and it was not followed by visual inspection due to the inordinately high number of cells that would have needed to be reviewed.

In the case of the MDA-MB-468 model, sensitivity (97.70%) and specificity (99.31%) were very good, although there were many false positives, leading to low accuracy (54.23%). Since MDA-MB-231 and MDA-MB-435 cancer cells are small and round, i.e. with distributions of nuclear are and wiggle very similar to mWBCs, we expected the poorest results from the mixture models using these two cell lines. For the experiments with MDA-MB-231, accuracy ranged 5.37%-11.59%, sensitivity ranged 19.60%-26.87%, and specificity ranged 95.44%-99.39%. For MDA-MB-435, accuracy ranged 4.04%-5.00%, sensitivity ranged 15.18%-46.82%, and specificity ranged 97.13%-98.07%. The results are poor, with the low accuracies due to many false positives, and low sensitivities due to many false negatives.

Unfortunately, performance was even worse with MCF7 and T47D. For the experiments with MCF7, accuracy ranged 0.791%-77.78%, sensitivity ranged 2.74%-44.68%, and specificity ranged 98.54%-99.97%. For T47D, accuracy ranged 0.274%-0.506%, sensitivity ranged 14.63%-21.14%, and specificity ranged 94.07%-95.59%. The low classification performance with these spiked models may be due to poor choices for the area and wiggle gates. However, since these cells lines supposedly had distributions of area and wiggle less like mWBCs, the gates should have worked better than for MDA-MB-231 and MDA-MB-435.

A possible explanation for this inconsistency may be in how the preliminary morphology studies were done. As mentioned in Section 2.1, breast cancer cells were grown as adherent cultures on coverslips prior to fixing. When slides of the mixture models were prepared, cells were fixed in suspension, then centrifuged as a monolayer onto slides. When grown on coverslips, MCF7 and T47D cells may spread out much more so than the other cell lines. When fixed in suspension and subsequently centrifuged onto slides, they remained spherical and aren't as flat.

Table 1.4: Classification performance of manual gating based strictly on nuclear morphology (area and wiggle). This is done in five other *in vitro* spiked models. These cell lines were selected because they were expected to have morphologies more similar to mWBCs.

Experiment	**Cell Line**	**Total**	**Cancer**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
1	MDA-MB-231	149,698	670	5.35	26.87	97.86
2	MDA-MB-231	155,366	653	10.83	19.60	99.32
3	MDA-MB-231	191,920	759	11.59	20.16	99.39
4	MDA-MB-435	425,253	2,827	5.00	15.18	98.07
5	MDA-MB-435	475,732	1,258	4.04	46.82	97.13
6	MDA-MB-468	84,507	696	54.23	97.70	99.31
7	MCF7	14,729	27	50	12.50	99.97
8	MCF7	19,604	64	77.78	44.68	99.97
9	MCF7	39,337	8	57.5	25.56	99.93
10	MCF7	92,933	401	1.97	2.74	99.41
11	MCF7	103,199	314	0.791	3.82	98.54
12	MCF7	106,819	353	7.88	5.38	99.79
13	T47D	101,947	164	0.506	14.63	95.37
14	T47D	107,239	88	0.274	14.77	95.59
15	T47D	107,580	123	0.405	21.14	94.07

In Figure 1.7, nuclear wiggle is plotted against nuclear area for the other five *in vitro* mixture models. With the exception of MDA-MB-468, these cancer cells are more like mWBCs (lower nuclear area, higher wiggle) than HCC1395 breast cancer cells.

Interestingly, the graphs for MDA-MB-231 and T47D now appear rather similar, although T47D cells seem to have a tighter distribution for both area and wiggle. MDA-MB-435 cells appear to have the smallest standard deviation in area but have a large standard deviation in wiggle. MCF7 cells have much variation in their statistical distributions for area and wiggle. In contrast to HCC1395 and MDA-MB-468, there are no obvious demarcations among singlet mWBCs, doublet mWBCs, and singlet cancer cells. Since the cells were prepared by cytocentrifugation for these results and were cultured on glass coverslips for Figure 1.2 and Figure 1.3, this shows that cancer cell morphology depends on how the cells were prepared.

To summarize this section, manual gating based on nuclear area and nuclear wiggle classified HCC1395 cancer cells well, but not for some of the other breast cancer cell lines. While there are still subtle differences in nuclear area and wiggle between mWBCs and the other breast cancer cell lines, overlap in the tails of the distributions makes accurate classification based only on linear thresholds of these two features unlikely. The next step, then, is to explore additional features in multidimensional space to classify between normal and cancer cells. Moreover, the classification may be nonlinear and nonlinear, multidimensional classification is more difficult to visualize. With the wealth of features available, we were motivated to perform nonlinear classification using additional features.

1.3.4 Automated neural network

Network development with HCC1395 spiked model

To test nonlinear classification, we selected a subset of seven features we thought *a priori* might separate the classes. We then trained and tested neural networks on the HCC1395 spiked model. The seven nuclear features are shown in Table 1.1 : 1) area, 2) perimeter, 3) wiggle (perimeter/area), 4) integrated intensity

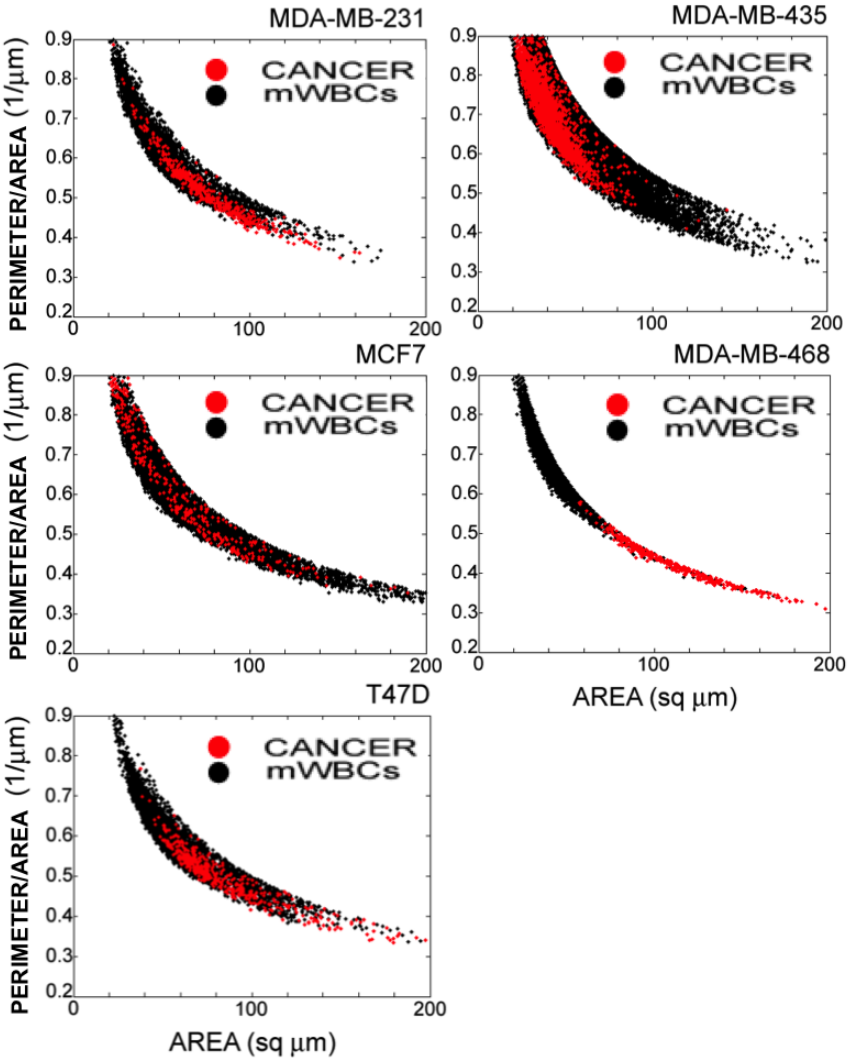


Figure 1.7: Nuclear area and wiggle for *in vitro* spiked models. The models used are consisted of MDA-MB-231, MDA-MB-435, MDA-MB-468, MCF7, and T47D breast cancer cells in mWBCs and demonstrate much more overlap than for the HCC1395 cell line (see Figure 1.3). Cancer cells are plotted in grey and mWBCs are black.

(which is proportional to the DNA content), 5) standard deviation of the intensity (which is a measure of contrast), 6) the third central moment of the intensity (which is a measure of asymmetry, or skewness, of the intensity distribution) and 7) the fourth central moment of intensity (a measure of "peakedness"of the distribution). Based on the literature[11, 30] and initial experiments, we chose a feedforward network topology consisting of four layers: an input layer with seven features, two hidden layers (ten nodes and three nodes, respectively), and an output layer with a single neuron.

Performance statistics for the HCC1395 model are summarized in the first three rows of Table 1.5 . As explained in Materials and Methods, the neural networks were trained on 30% of the original data and tested on the remaining 70%. The numbers given for Total Cells and Cancer Cells are thus from that 70% of the cells. For Data Set #1, performance was mixed. There were 1,013 true positives, 284 false positives, 5,063 true negatives, and 209 false negatives. It is interesting to note that performance is worst for the smallest data set (#2) and best for the largest data set (#3). The poor performance for Data Set #2 may be due to rarity of cancer cells, i.e. there may simply not have been enough cells (118) to train the network adequately to distinguish between cancer cells and mWBCs. Overall, the accuracy range was 43.22%-93.07% (vs. 67.4-93.3 in Table 1.3 manual gating), the sensitivity range was 30.18%-82.90% (vs. 90.6-98.4 in Table 1.3 manual gating), and the specificity range was 98.34%-99.79% (vs. 99.7-99.8 in Table 1.3 manual gating). Thus, the sensitivities and specificities (or true positive and true negative rates, respectively) were comparable to manual gating, with the neural network classification being substantially less labor-intensive.

Results from Data Set #1 are shown on 2D scatterplots of wiggle vs. area in Figure 1.8. The top panel is the actual data, with mWBCs in black and HCC1395 breast cancer cells in grey. The middle panels shows the cancer and mWBC subpopulations the neural network identifies, projected onto wiggle vs. area. The third panel plots true positives, false positive, true negative, and false positive classifications in four colors. Such nonlinear separation is impossible with manual gating using simple linear thresholds.

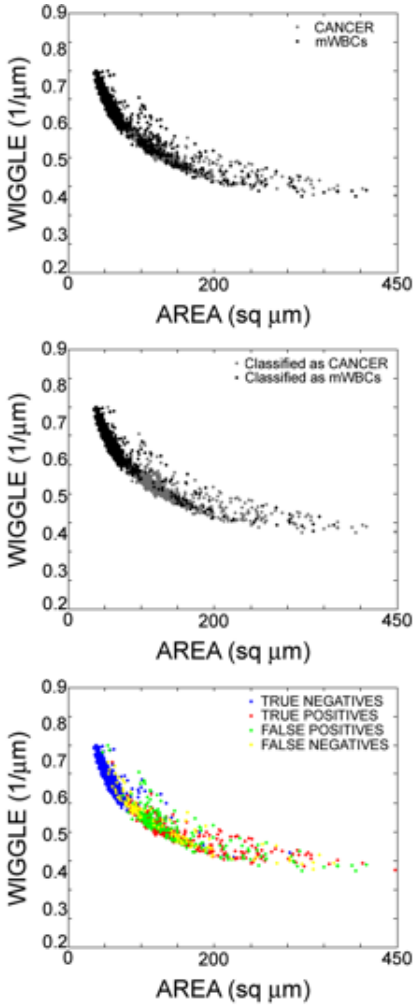


Figure 1.8: Wiggle and Area distribution of neural network classified cells. These results are from the neural network in Experiment #1, Table 1.5. 30% of the data was used for training. The training set consisted of 524 HCC1395 cancer cells with 2,291 mWBCs (2,815 cells total). The network was then evaluated with the remaining 70% of the data, consisting of 1,222 HCC1395 cancer cells and 5,347 mWBCs (6,569 cells total). The top panel shows the actual distribution of cells in the data with HCC1395 cancer cells as grey and mWBCs as black. The middle panel shows how the network classifies cells (HCC1395 grey, mWBCs black). In the bottom panel, performance of classifying cancer cells is given: true positives (red), false positives (green), true negatives (blue), and false negatives (yellow).

Network performance with other spiked models

Given the improvement in performance of HCC1395 performance with the neural network, we next applied the same methodology to train and evaluate neural networks using data from the *in vitro* spiked models of the other five cancer cell lines. Results using these data sets are also shown in Table 1.5 (Data Sets 4-11). For the two experiments using MDA-MB-231, the accuracy range was 40.50%-60.33% (vs. 5.35-11.59 in Table 1.4 manual gating), the sensitivity range was 34.48%-36.20% (vs. 19.6-26.87 in Table 1.4 manual gating), and specificity range was 99.16%-99.51% (vs. 97.86-99.39 in Table 1.4 manual gating). Thus, all three performance measurements improved with the neural network. With the MDA-MB-435 model, the accuracy was 20.0% (vs. 4.04-5.0 in Table 1.4 manual gating), the sensitivity was 21.05% (vs. 15.18-46.82 in Table 1.4 manual gating), and the specificity was 99.99% (vs. 97.13-98.07 in Table 1.4 manual gating). Thus, the accuracy and specificity improved and there was not enough data to evaluate the change in sensitivity. With the MDA-MB-468 model, the accuracy was 85.49% (vs. 54.23 in Table 1.4 manual gating), the sensitivity was 55.65% (vs. 97.7 in Table 1.4 manual gating), and the specificity was 99.92% (vs. 99.31 in Table 1.4 manual gating. Thus, the accuracy and specificity improved and the sensitivity worsened with the neural network; manual gating had many more false positives, while the neural network had significantly fewer false positives and more false negatives. With MCF7, the accuracy was 0%-23.31% (vs. 0.8-77.78 in Table 1.4 manual gating), sensitivity was 0%-6.75% (vs. 2.74-44.68 in Table 1.4 manual gating), and specificity was 100% (vs. 99.54-99.57 in Table 1.4 manual gating). With T47D, accuracy was 0-50.0% (vs. 0.27-0.5 in Table 1.4 manual gating), sensitivity was 0-0.18% (vs. 14.63-31.14 in Table 1.4 manual gating), and specificity was 100% (vs. 94.1-95.59 in Table 1.4 manual gating). Thus, while the specificity improved for both MCF7 and T47D cell lines, the accuracies worsened for MCF7 and improved for T47D and the sensitivities worsened for both MCF7 and T47D. Again, the classification performance was very poor for the MCF7 and T47D cells.

These neural networks may have been overtrained, especially with Data Sets 9 and 11, where there were too few cancer cells present in the training sets (in

terms of absolute numbers). The neural network may have trained on intracellular variations among the predominantly mWBCs instead of the cancer cells. As a result, the network might have learned better to correctly classify mWBCs, rather than to differentiate between cancer and non-cancer cells.

Neural network trained on a mixture of all six cancer cell lines

To test the hypothesis that the neural network was over training on mWBCs, we created a single, unified neural network trained using a pool of 30% of each of the *in vitro* mixture model experiments from all six cell lines (11 in all, as in Table 1.5). After training was complete, performance statistics were gathered by testing the neural network on the remaining 70% of each of the individual spiked data sets. Table 1.6 shows a comparison of the results of this more robust network with the results from the individual neural networks shown in parentheses. While performance was not equal for each of the breast cancer cell lines, it improved in many cases. With MCF7, the accuracy was 5.06%-22.22% (vs. 0%-23.31% in Table 1.5 with neural networks trained only with MCF7 spiked data), sensitivity was 0.09%-0.98% (vs. 0%-6.75% previously), and specificity was 99.93%-100% (vs. 99.91%-100% previously). For T47D, the accuracy was 27.03%-100% (vs. 0%-50.00% with neural networks trained only with T47D spiked data), sensitivity was 0.18%-1.01% (vs. 0%-0.18% previously), and specificity was 99.99%-100% (vs. 100% previously).

For MDA-MB-435, the unified network performed significantly better. Accuracy was 62.74% (vs. 20.00% previously), sensitivity was 84.65% (vs. 21.05% previously), and specificity was 100% (vs. 99.99% previously). For MDA-MB-231, accuracy was 48.21-67.19% (vs. 40.50%-60.33% previously), sensitivity was 3.37%-4.12% (vs. 34.48%-36.20%), and specificity was 99.46%-99.96% (vs. 99.16%-99.51% previously). The network has fewer false positives but many more false negatives, explaining the poor sensitivity.

The unified network's performance for HCC1395 showed fewer false positives but more false negatives. Accuracy was 80.20%-96.88% (vs. 43.22%-93.07% with networks trained only with HCC1395 spiked data), sensitivity was 18.34%-

77.39% (vs. 30.18%-73.93% previously), and specificity was 96.63%-99.98% (vs. 94.69%-99.79% previously). The single, unified neural network appeared to have sacrificed some performance for the cells at the extremes small (MDA-MB-231) and large (HCC1395) cancer cells. On the other hand, since it had a more diverse training set, the network didn't have a problem with overtraining. Where previous neural networks had 0% accuracy and sensitivity (Data Sets #9 and #11), the unified network showed improvement.

1.4 Discussion

Pathologists routinely use nuclear morphology while analyzing histological tissue sections to diagnose cancer. For example, pathologists often look at nuclear-to-cytoplasmic volume ratio and changes in cellular texture and shape as indicators of malignant transformation. Here, we tested whether nuclear morphology might be applied on a cell-by-cell basis to classify cancer cells in fluorescent microscopy images.

In observing images of cell nuclei from different breast cancer cell lines and comparing them with nuclei from mononucleated white blood cells, there were differences in the distributions of nuclear area and nuclear wiggle. It was hypothesized that these differences could be exploited to classify between cancer and normal using nuclear morphology alone. A simple model system using HCC1395 breast cancer cells spiked among normal mWBCs was used to develop classification methods since HCC1395 cells were very dissimilar to mWBCs and relatively easy to identify. Manual gating with linear thresholds on area and wiggle provided good initial results with HCC1395, but when similar gating strategies were used for on spiked model with the other breast cancer cell lines, performance was worse. Since these other cell lines were chosen for morphologies more similar to mWBCs, they were expected to provide a greater challenge.

To further test the ability to CTCs based on nuclear morphology, additional morphometric and fluorometric features were chosen and it was hypothesized that these additional features and nonlinear supervised classification using neural net-

works would improve performance. We trained and evaluated a fully-connected, feed-forward neural network architecture that used seven features thought *a priori* to be useful for classification. We tried neural networks trained specifically on data from each spiked model, as well as a unified neural network trained with data pooled from model experiments using all six cell lines. The individual neural networks had poor performance, possibly due to overtraining or over-fitting of the data. That is, because most of the spiked data consisted of mWBCs, the networks were learning to identify the variations among mWBCs rather than the differences between normal and cancer cells. The unified network on pooled cells from all six cell lines may have partially overcome this problem because its training data was more diverse. The downside is that classification performance decreased for the largest (HCC1395) and smallest (MDA-MB-231) breast cancer cell lines.

Detecting and classifying rare cancer cells in blood is a challenge. The results presented here, however, are promising and lead us to believe that classification using morphology is worth exploring further. Additional insights may be gained by applying more advanced and sophisticated methods. We assumed our input feature set would be effective for classification. Systematic feature selection should be applied to a large set of nuclear features to determine whether performance can be improved by better feature selection. Artificial enrichment of the training data (since there are far many more mWBCs than cancer cells) or increased error-weighting of false negatives vs. false positives may also improve results. Other unsupervised methods such as principal components analysis (PCA) and support-vector machines (SVM) may also improve performance. At the very least, morphology may be useful as an adjunct to gene expression biomarkers for detection of rare cancer cells in blood.

Table 1.5: Neural network performance (*a priori* feature set) trained and evaluated for *in vitro* spiked models. Each experiment represents a different neural network. Training was done on 30% of the data set. Total Cells and Cancer Cells refer to the remaining 70% of the data comprising the testing sets.

Dataset	**Cell Line**	**Total Cells**	**Cancer Cells**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
1	HCC1395	6,569	1,222	78.10	82.90	94.69
2	HCC1395	4,211	169	43.22	30.18	98.34
3	HCC1395	50,651	1,327	93.07	73.93	99.79
4	MDA-MB-231	51,515	801	40.50	36.20	99.16
5	MDA-MB-231	347,889	7,302	60.33	34.48	99.51
6	MDA-MB-435	115,612	19	20.00	21.05	99.99
7	MDA-MB-468	59,155	487	85.49	55.65	99.92
8	MCF7	212,066	815	23.31	6.75	99.91
9	MCF7	278,729	2,198	0	0	100
10	T47D	47,723	554	50.00	0.18	100
11	T47D	221,736	988	0	0	100

Table 1.6: Performance of a single, unified neural network for *in vitro* spiked models. A single neural network was trained and tested using a pool of all the data. Similar to before, the training set consisted of 30% of the cells from each data set, while the testing set consisted of the remaining 70%. Results in parentheses are for the neural network trained on that data set alone. Up and down arrows indicate relative change in performance.

Dataset	**Cell Line**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
1	HCC1395	96.88 (43.22) △	18.34 (30.18) ▽	99.98 (98.34) △
2	HCC1395	80.20 (78.10) △	59.66 (82.90) ▽	96.63 (94.69) △
3	HCC1395	91.29 (93.07) ▽	77.39 (73.93) △	99.71 (99.79) ▽
4	MDA-MB-231	48.21 (40.50) △	3.37 (36.20) ▽	99.94 (99.16) △
5	MDA-MB-231	67.19 (60.33) △	4.12 (34.48) ▽	99.96 (99.51) △
6	MDA-MB-435	62.74 (20.00) △	84.65 (21.05) △	100 (99.99) △
7	MDA-MB-468	84.65 (85.49) ▽	77.00 (55.65) △	99.88 (99.92) ▽
8	MCF7	5.06 (23.31) ▽	0.98 (6.75) ▽	99.93 (99.91) △
9	MCF7	22.22 (0) △	0.09 (0) △	100 (100)
10	T47D	100.00 (50.00) △	0.18 (0.18)	100 (100)
11	T47D	27.03 (0) △	1.01 (0) △	99.99 (100) ▽

Chapter 2

Improving Classification by Enriching the Cytometric Feature Space

2.1 Introduction

A classifier makes decisions based only on the information it is fed. If the data inputed into the classifier contains relatively few information about the class of the cells to be sorted, the classifier fails to detect and utilize the class information as it remains hidden in the plethora of other irrelevant data. A typical cytometry dataset can be defined as a two-dimensional matrix of numerical data, with features in the columns and cells in the rows. We hypothesize that this problem can be solved by enriching the class information content in both of these dimensions.

First, not all features fed into the classifier contain useful class information. Therefore, it is useful to systematically identify and eliminate those features that do not contain significant class information and to retain those that do. We propose three feature selection methods to perform this task and aim to briefly compare the resultant feature sets.

Second, if a dataset contains very few cells that belong to one of the two classes, the error due to their misclassification will not be large enough to affect the

training. Therefore, it is useful to train the classifier on a set that has a significant number of cells belonging to all classes to be identified. However, the problem of detecting CTCs in blood has one class that is inherently rare. To remedy this, the traing set is enriched in a supervised manner by removing WBCs, which lack the gold standard stain. Further Principal Component Analysis (PCA) is explored as a classifier that assumes the variance in feature space of an enriched class is chiefly due to the presence of multiple classes.

2.2 Features Selection

Not all features fed into the classifier carry significant information regarding the class of the cells. Such features introduce unnecessary noise into any classifier and can further deter classification by confusing the classifier. These features should be systematically identified and removed from the final feature set that is used for training of the classifier.

2.2.1 Features Selection Methods

Leave One Out Cross Validation (LOOCV)

A network with 17 inputs was created and a feature set with all 17 available features (features in Table 4.2 with percentile distribution measures excluded) were used to train the neural network with %30 of the cells to 20000 epochs or minimum error gradient of -1×10^{-20}, whichever occurred first. The classification performance was measured on the unseen rest (%70) and this performance was recorded as the reference performance. Next, a series of 17 feature sets were prepared in a way that each set lacked one of the available 17 features. A neural network was constructed with 16 inputs and was trained with each of the feature sets using the same training parameters as the reference training. To compare the performance of the network trained on data lacking each feature with the reference performance a new measure J was devised as shown in Equation (2.1). We described this measure as one minus the ratio of total false detections of the subject network to that of the reference network which allows J to approach +1

when performance improves by the removal of the feature and grows increasingly negative when the performance is compromised by the removal of the feature. Therefore, a negative J value indicates the removed feature was important in the classification–the performance was worsened by its removal–and a zero or positive J value indicates a feature that is redundant or unnecessary in the classification decision making–its removal led to improved performance.

$$J = 1 - \frac{(FP + FN)_{FeatureEliminated}}{(FP + FN)_{AllFeatures}} \tag{2.1}$$

Traditional Weight Observation

In a neural network such as the one used here, the input to a neuron is the linear combination of the outputs of the previous layer of the input feature for the first layer. After training the weight for the input significant to the classification will therefore have a large magnitude. During the training this weight should only stably increase rather than fluctuating uncontrollably. Hence we hypothesize that a weights corresponding to a significant input feature should exhibit a high magnitude and low rate of change.

To observe the weight changes during training, two methods were devised. First, we constructed a network with the four layers common to the other networks used here Figure 2.1. The network was trained with the the given feature set for 10,000 epochs. The weight of the connections between the input and first hidden layer was observed and recorded during training. The rate of change of these values was subsequently calculated. After the training was done, the sum of the absolute value of the weights (Equation (2.2)) and the sum of the absolute value of the rate of change of the weights (Equation (2.3)) were calculated and recorded for each of the connections that correspond to each feature. The weights change abruptly in the beginning of the training (from random initialization values towards trained values). Therefore, the gradients are summed only from iteration 1000 to the end of training and the beginning of training is excluded form the sum. Features with high sum absolute value of weight and low sum weight gradient were chosen to construct a new feature set.

$$\sum_{t=1}^{T}\sum_{i=1}^{m} \mid X_{i,t} \mid \tag{2.2}$$

$$\sum_{t=1000}^{T}\sum_{i=1}^{m} \mid \frac{dX_i}{dt} \mid \tag{2.3}$$

Pseudo-Layer Weight Observation

A second network architecture, shown in Figure 2.2, was made to create a single connection weight per input. In this architecture, a pseudo-layer was introduced and connections were made on one-to-one bases between the pseudo layer and the input layer. The weights of these connections were observed and recorded as the network was trained on the full feature set. Sum of the absolute values of the weights and the sum of the absolute value of the rate of change of the weights were calculated and reported according to Equations (2.4) and (2.5). The weights change abruptly in the beginning of the training (from random initialization values towards trained values). Therefore, the gradients are summed only from iteration 1000 to the end of training and the beginning of training is excluded form the sum. As before, features with high sum absolute value of weight and low sum weight gradient were chosen to construct a new feature set.

$$\sum_{t=1}^{T} \mid X_t \mid \tag{2.4}$$

$$\sum_{t=1000}^{T} \mid \frac{dX_t}{dt} \mid \tag{2.5}$$

2.2.2 Features Selection Results and Discussion

Leave One Out Cross Validation (LOOCV)

Table 2.1 illustrates J values found for each of the features. Features with a negative J were selected and a new feature set using these features was constructed.

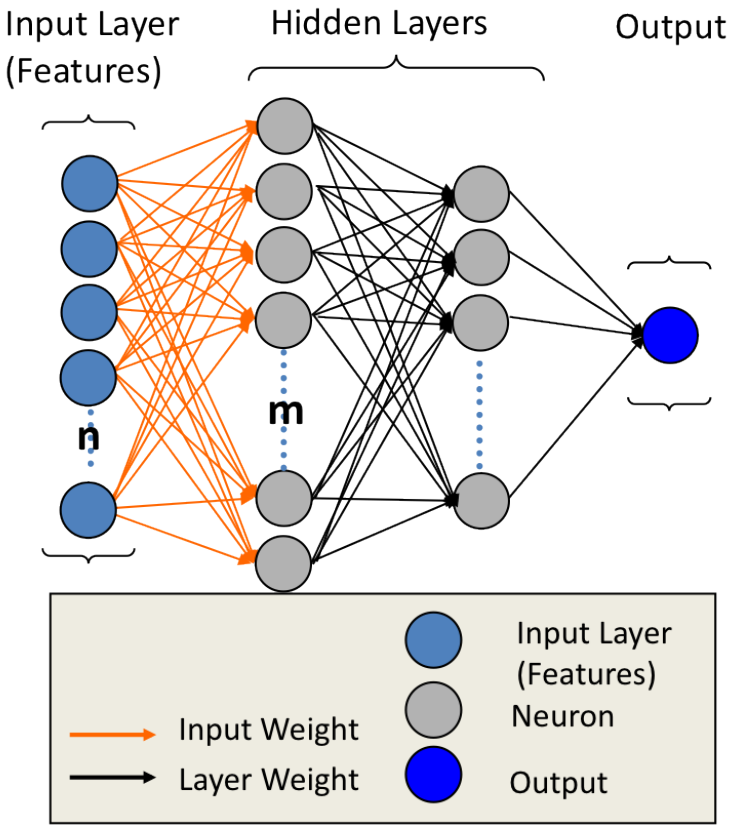


Figure 2.1: Network architecture used for Traditional Weight Observation. The network is consisted of n input neurons, each connected to a feature and m neurons in the second layer. Note the Input Weight connections shown in orange.

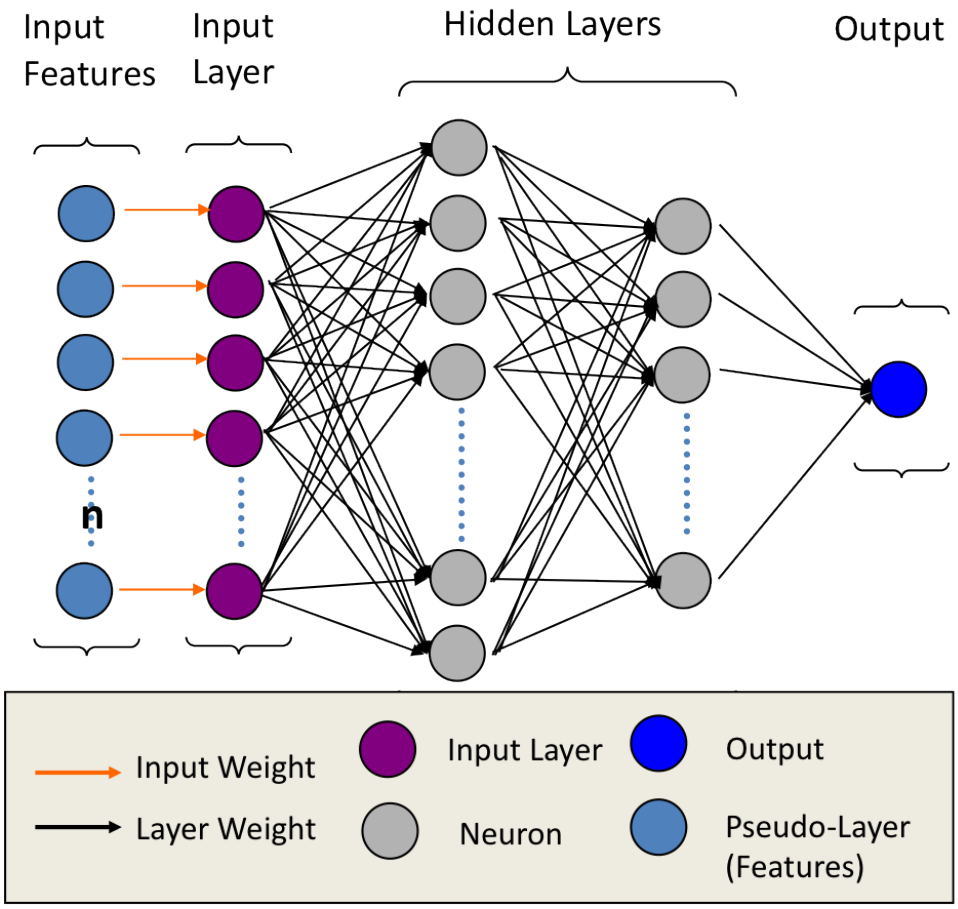


Figure 2.2: Network architecture used for Pseudo-Layer Weight Observation. The network is consisted of n input neurons. Note the additional input layer and the Input Weight connections shown in orange.

Table 2.1: Relative Performance (J) for features tested. Negative values indicate significant features for which the removal worsened the performance. A complete description of each feature is provided in Table 4.2.

Feature Eliminated	J
Integ_gs	0.10
CM3**1/3	0.03
Width	0.02
Area	0.02
CM4**1/4	0.00
Height	-0.02
IQ_range	-0.02
Variance	-0.02
Wiggle	-0.07
CM4	-0.11
AbsDev	-0.16
Perimeter	-0.20
CM3	-0.25
Area**1/2	-0.28
Aver_gs	-0.28
Wiggle_Nrm	-0.39
StdDev	-0.46

Traditional Weight Observation

Figure 2.3 illustrates the sum of absolute value of the training weights in the first layer as described in Equation (2.2) for 10,000 epochs of training. While the values change abruptly in the beginning of the training, they soon settle to their respective levels. This illustrates the training behavior of the neural network and its apparent convergence to the point where classification error is left relatively unchanged. Also noteworthy is the large value corresponding to the CM4 feature. At the end of the training, most features seem to have values in the range of 0-500,

CM4 has a relatively large value of 5000.

Figure 2.4 illustrates the sum of absolute value of the gradient of gradient of weights in the input layer as described in Equation (2.3). The abrupt change in the weights in the first 1000 epochs is visible and indicative of weights changing from their random initialized values to values that accommodate low classification error. This is the reason why we chose to exclude this region from the summation in Equation (2.3).

Once both of the weights and their gradient is summed for the training period, the resultant values are used to select significant features. Figure 2.5 illustrates the plot of the two measures given in Equations (2.2) and (2.3). A line threshold is applied to select a subset of features with high training weight magnitude and a stable weight value throughout the training.

Pseudo-Layer Weight Observation

Figure 2.6 illustrates the absolute value of the training weights in the pseudo layer structure (left) as described in Equation (2.4) and its cumulative sum (right). The values are found for training the network for 10,000 epochs. Conveyance of each weight to its corresponding trained value is apparent in the left panel of the figure. The cumulative sum of those values are then plotted for each epoch and are illustrated on the right. The final cumulative sum (at epoch 10,000) is used as one of the criteria for feature selection.

Figure 2.7 illustrates the absolute value of the weight gradient of each feature node in the pseudo layer (left) as described in Equation (2.5) and its cumulative sum throughout training (right). The large values in the beginning of the training are indicative of the rapid changes in the weights from their randomly initialized values towards their trained target. As mentioned before, this is why the first 1000 epochs are excluded form summation in Equation (2.5).

The results at the end of training are recorded and plotted in Figure 2.8. This figure illustrates the plot of the two measures given in Equations (2.4) and (2.5). A line threshold is applied to select a subset of features that with corresponding nodes that have high weight magnitude and a stable weight value throughout

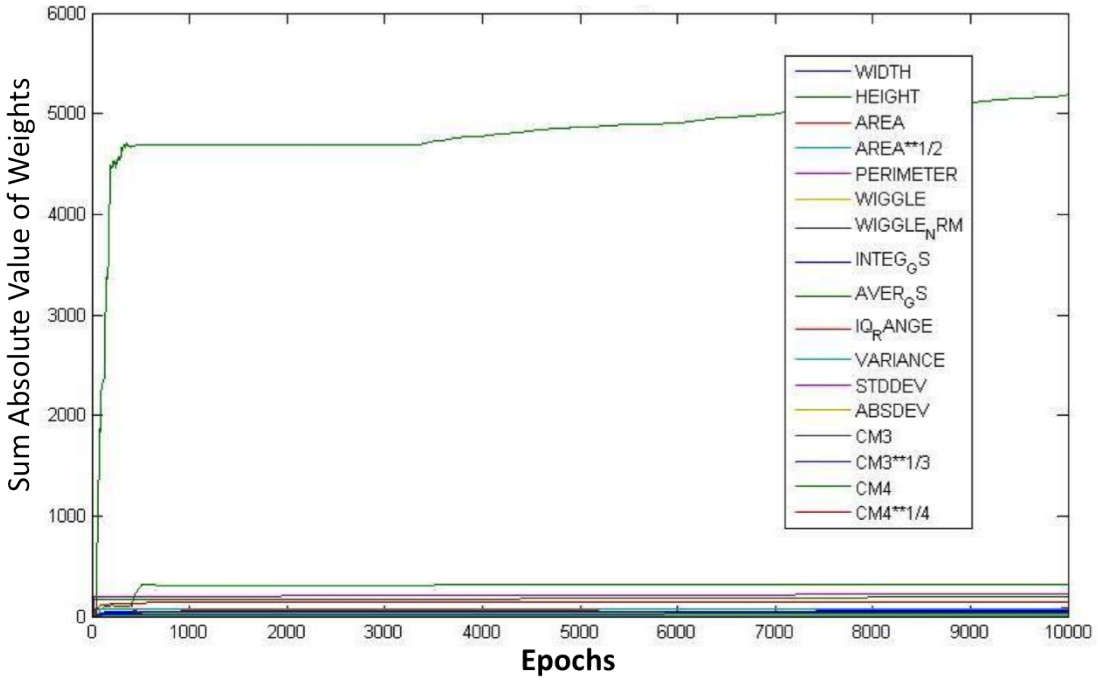


Figure 2.3: Plot of sum absolute value of training weights in the input layer of the traditional network structure. This value, as defined in Equation (2.2) is plotted for each feature node and as a function of the epoch (iteration) of training.

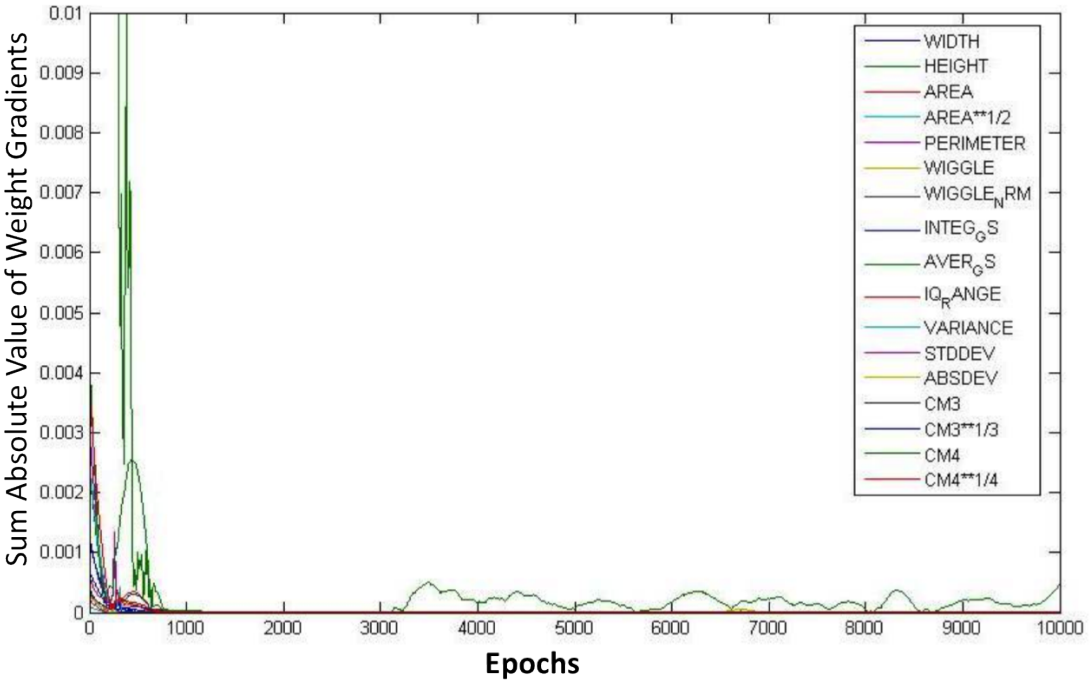


Figure 2.4: Plot of sum absolute value of the gradient of training weights in the input layer of the traditional network structure.This value, as defined in Equation (2.3) is plotted for each feature node and as a function of the epoch (iteration) of training.

the training.

Summery of Feature Selection Results

From the three methods above, three feature sets were constructed as illustrated in Table 2.2. In confirmation of our preliminary studies, a form of area and wiggle (ratio of perimeter/area) appeared in all three feature sets. Interestingly, area was replaced by square root of area in all three feature sets. A significant number of features repeated in two feature sets. The performance improved significantly for most of the cell lines when these reduced feature sets were used in lieu of the full feature set (Figure 2.9, pseudo-layer weight observation not shown).

Table 2.2: Feature sets constructed by weight observation and leave one out cross validation methods. The number in the parentheses indicate the number of sets in which the Feature is represented.A complete description of each feature is provided in Table 4.2.

Traditional Weight Observation	**Pseudo-Layer Weight Observation**	**LOOCV**
AbsDev **(2)**	Area**1/2 **(3)**	AbsDev **(2)**
Area**1/2 **(3)**	CM3**1/3 **(2)**	Area**1/2 **(3)**
Aver_gs **(2)**	Ineg_gs **(2)**	Aver_gs **(2)**
CM3**1/3 **(2)**	IQ_range **(1)**	CM3 **(1)**
CM4**1/4 **(1)**	Variance **(2)**	Perimeter **(1)**
Integ_gs **(3)**	Wiggle_Nrm **(3)**	StdDev **(1)**
Variance **(3)**	Wiggle **(1)**	Wiggle_Nrm **(3)**
Wiggle_Nrm **(3)**		

2.3 Training Set Enrichment

As mentioned in the introduction of this chapter, one of the aims of the study is to enrich the training feature set in both the features dimension and the

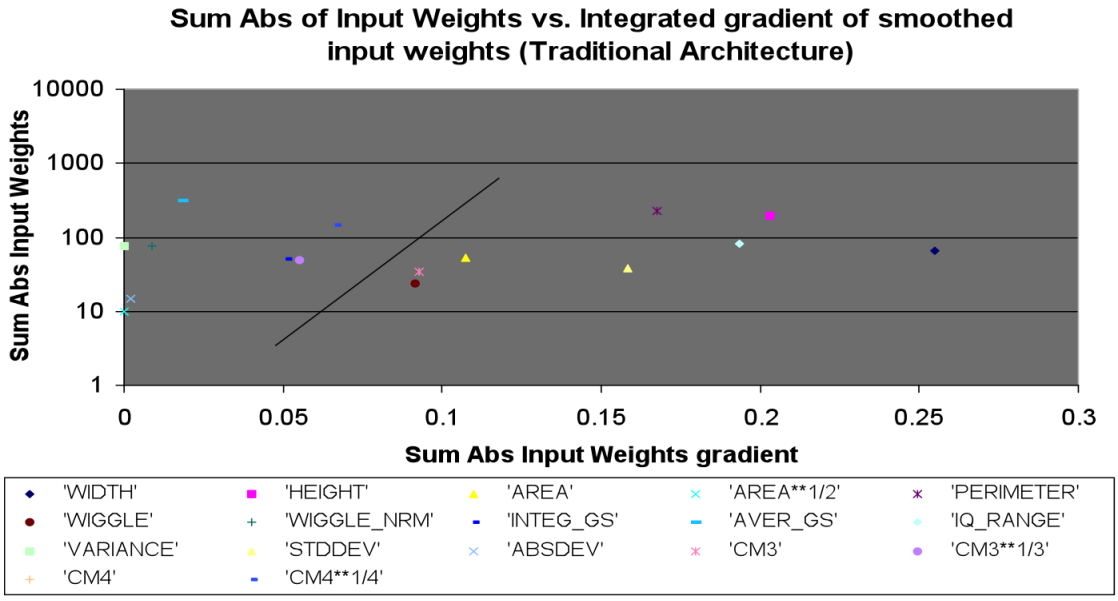


Figure 2.5: Scatter plot of feature evaluation results based on the traditional network structure.

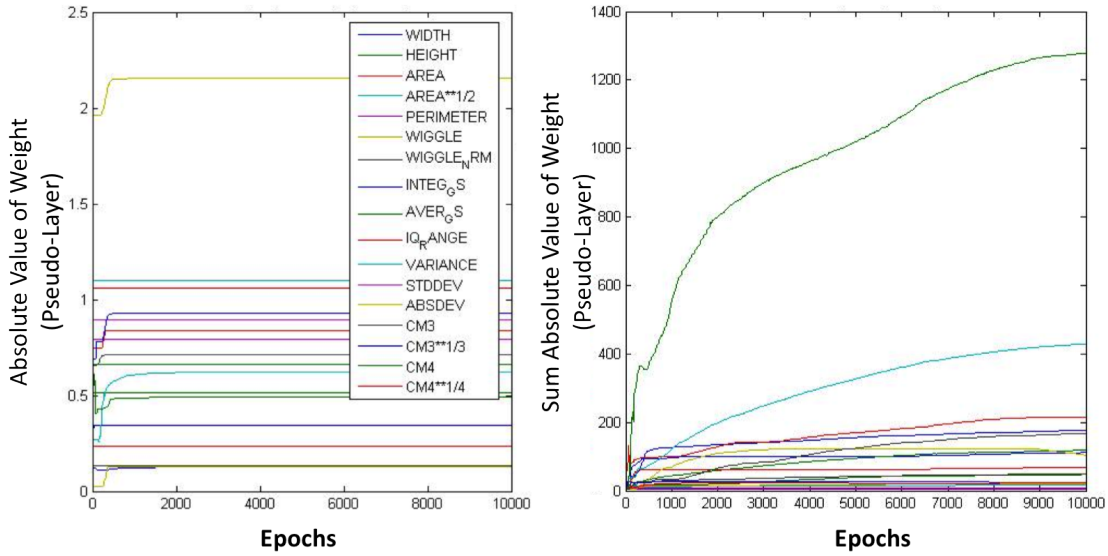


Figure 2.6: Plot of the absolute value of the weight of each feature node in the pseudo layer structure (left) and its cumulative sum throughout training (right). The values plotted in the right panel are defined in Equation (2.4) from which X_t is plotted on the left. Both values are plotted as a function of epoch (iteration) of training.

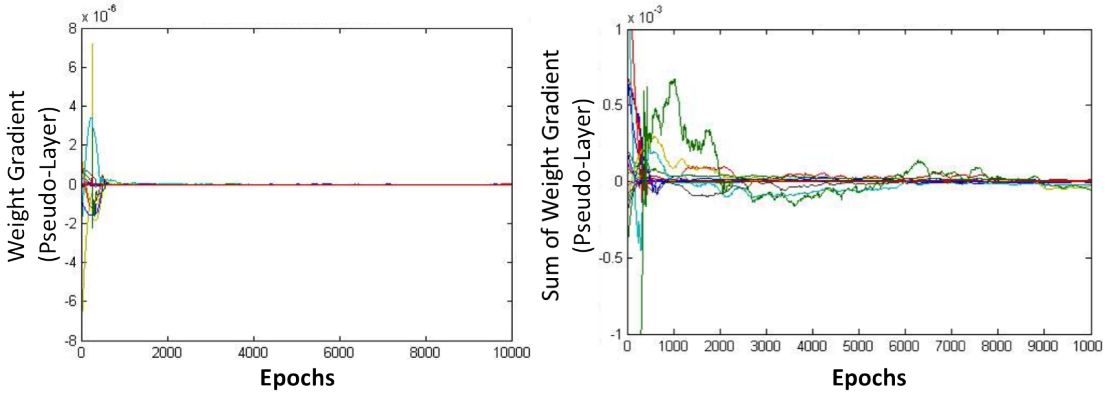


Figure 2.7: Plot of the weight gradient of each feature node in the pseudo layer structure (left) and its cumulative sum throughout training (right). The values plotted in the right panel are defined in Equation (2.5) from which dX_t/dt is plotted on the left. Both values are plotted as a function of epoch (iteration) of training.

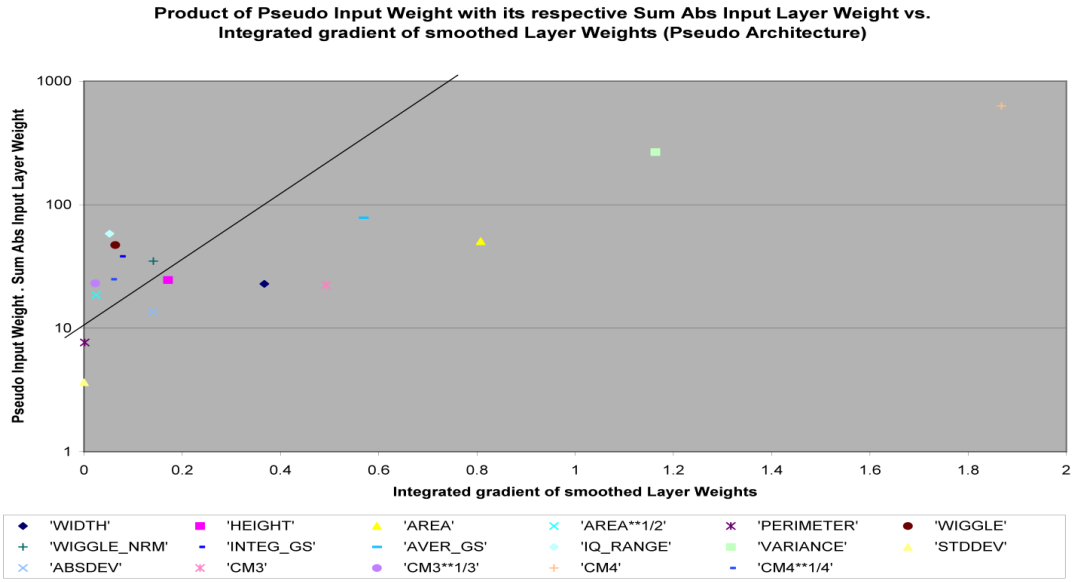


Figure 2.8: Scatter plot of feature evaluation results based on the pseudo-layer network structure.

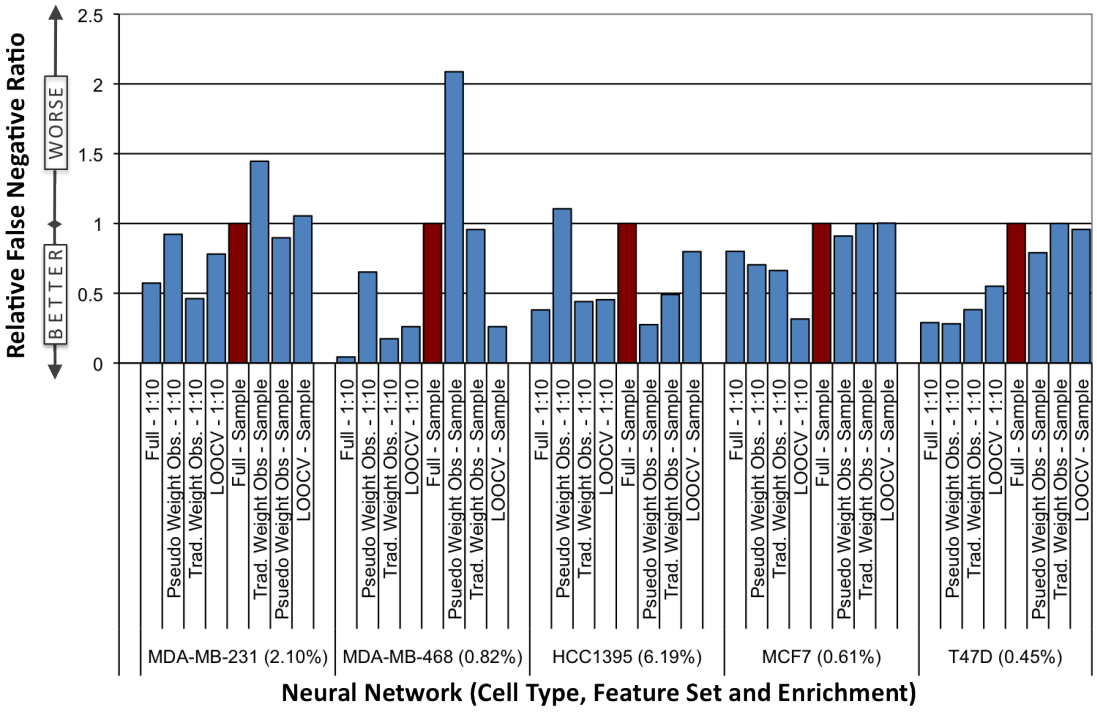


Figure 2.9: Bar graph of relative performance based on numbers of false negatives. Results normalized with respect to full network without enrichment. Generally, without enrichment (right side of each red bar), feature selection improved the performance or left it unchanged; with the exception of only three experiments out of fifteen. Regardless of feature set, however, enrichment of training set to a 1:10 ratio (left side of each red bar) decreased false negatives; with the exception of only one experiment out of twenty. The original concentration of cancer cells in the training set is shown in the parenthesis next to the name of each cell line.

cells dimension. In our early studies of neural networks, we observed that the performance of the classifier is directly dependent to the concentration of cancer cells in the training set. This led us to hypothesize that if the training set represented the cancer cells better, the trained network would be able to detect them with better performance. Hence we designed an experiment in which WBCs are taken out of the training set data to artificially increase the concentration of cancer cells to 1:10. The neural network trained on this set was then applied to unseen mixture data with the original cancer cell concentration and the performance was recorded (Figure 2.9). As expected the enriched training set greatly improved the classification performance for most cell lines and experiments. This method, in combination with feature selection described above improve the performance of the neural network classification by enriching the feature space with class data in both the features and cells dimensions.

2.4 Classification with PCA

Designing the studies related to training feature space enrichment led us to further delve into the study of the statistical richness of the training feature space. For example, an interesting question that rose was whether the variance of the feature space is dominated by the class data, or contrary by the internal variation in the features of each class. In other words, is the variance of features in each class larger than that of the combined classes? The significance of this question lies in the fact that if the variance is chiefly due to the existence of multiple classes, thresholding the high variance dimensions of the space could be used as a classification mean.

Briefly, PCA can be used to find the high variance components of the feature space. A threshold can then be applied to those high variance components to separate the two classes. Further, unlike neural networks that are supervised and require ground truth data for training, such a method is unsupervised and only requires a set for which the variance is mainly due to the existence of two classes.

2.4.1 PCA Methods

To test the application of PCA for the classification of cancer cells, we designed an *in vitro* breast cancer and an *in vivo* melanoma mouse model.

in vitro Breast Cancer Model

A spike model was created using a mixture of HCC1395 breast cancer cells in human WBCs as defined in Section 1.2.2 and cytometric data was obtained as described. Matlab (Mathworks, MA, USA) was used as a the platform to apply PCA on the cytometric data. The resultant components were thresholded using an empirical hyperplane and performances were calculated with respect to the ground truth staining.

in vivo Melanoma Mouse Model

The B16F10 human melanoma is a cell line used extensively in animal models of aggressively metastasizing melanoma. [9, 8, 103] SW1 melanoma cells grow less aggressively than B16F10 and contain an N-ras mutation within codon 13 (Gly 13 Asp due to GGT to GAT or ACC to ATC mutations). Dr. Ze'ev Ronai's group transfected both of these cell lines to stably express GFP for in vivo animal studies. After subcutaneous injection of 105 cells into mice, tumors are palpable after two weeks. Mice were sacrificed at roughly one week intervals and blood collected by direct cardiac puncture. During pilot experiments presented here, a few mice were euthanized by inhalation 16 of 100% carbon dioxide gas or by an overdose of tribromoethanol (Avertin®) injected into the peritoneal cavity. Total blood volume in a mouse is approximately 6-8% of its body weight [44]. The mice used in our studies were typically 10-15g, so we expected to collect between 0.6-1.2 ml blood. The collected blood was then centrifuged and the layer containing the WBCs and cancer cells were extracted, scanned and analyzed as described in Section 1.2.2. The cytometric data obtained was subjected to PCA in Matlab (Mathworks, MA, USA). After applying an empirical hyperplane threshold the performances were calculated with respect to the GFP ground truth. For this experiment, the selected subpopulation was again subjected to PCA and thresholded

and performances were once again calculated.

2.4.2 PCA Results and Discussion

in vitro Breast Cancer Model

The first three components of PCA applied to HCC1395 cells spiked into human blood are shown in Figure 2.10. The classification performance was evaluated using the decision plane (yellow) in the right panel of Figure 2.10 and the performance results are shown in Table 2.3. For preliminary data, this is very encouraging as it indicates the likelihood that unsupervised PCA will potentially perform comparably to neural networks. The PCA method can be converted to a fully unsupervised method by implementing clustering algorithms (e.g. k-means) to the PCA transformation of enriched data. A decision plane calculated based on the resultant clusters can then be applied to PCA transformation of unseen data for class assignemnt.

Although a spiked model may be less realistic than an *in vivo* model, its simplicity is an advantage. Preparing samples from an animal model costs more and takes significantly longer than an *in vitro* model. With the latter, we were able to improve our sample preparation techniques, test and evaluate our image processing and develop and experiment with our classification algorithms before moving to *in vivo* models. After developing the current classification techniques for an *in vitro* model, it is straightforward to test and modify it to work for *in vivo* models.

in vivo Melanoma Mouse Model

The first three components of the PCA analysis form the mouse melanoma model is shown in Figure 2.11. A reasonable separation may be created by a threshold along just the first principal component. See Views (B) and (C) in Figure 2.11. Thresholding at zero gives 13.67% accuracy, 68.87% sensitivity, and 98.75% specificity. Applying a second PCA to this subpopulation to further purify this set (e.g., distinguish between melanoma cells and WBC doublets), gives

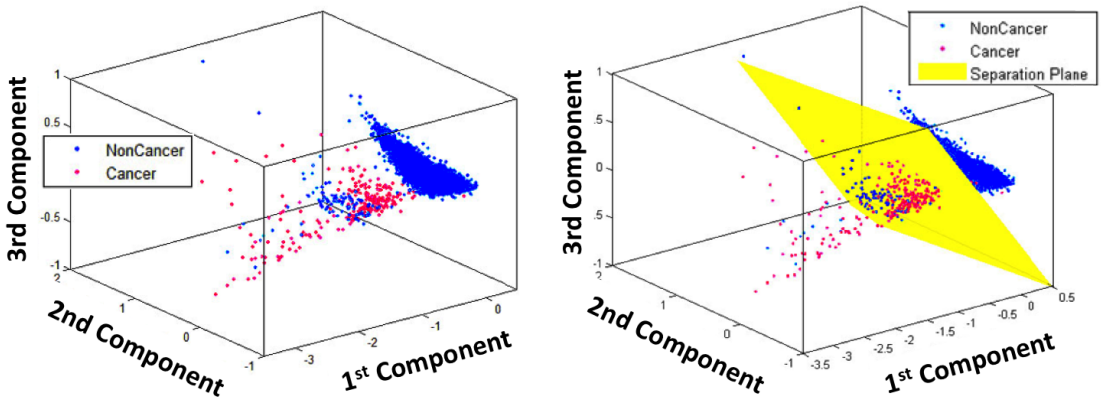


Figure 2.10: PCA components of the *in vitro* breast cancer model.. PCA linearly maps features of the spike model into components that have maximum variance (left). A decision hyperplane is applied to assign class to each datapoint (right).

Table 2.3: Performance of PCA classification on the *in vitro* breast cancer model containing hWBCs and HCC1395 cells.

Accuracy (%)	72.9%
Sensitivity (%)	92.1%
Specificity (%)	98.6%
True Positives	223
True Negatives	5691
False Positives	83
False Negatives	19

25.87% accuracy, 65.84% sensitivity, and 99.46% specificity. Detailed performance statistics are given in Table 2.4.

These preliminary animal studies highlight important considerations in order to move beyond the *in vitro* models. We see that cancer cells have different morphology when in the circulation. As the cancer cells travel through blood, they are subjected significant mechanical disruption due to the shear forces of blood flow and the strains of the tiny vessels in the microvasculature. Such forces eliminate the very large cancer cells. Those that remain will have size characteristics that are similar to WBCs, as WBCs too survive circulation. This toughens the task of classification of *in vivo* CTCs using morphology. With the above example, by applying PCA on an *in vivo* model and finding performances, we were able to illustrate that morphology alone could still be used for classification of *in vivo* cancer cells. With the given methods developed, we are now able to investigate interesting questions regarding the process of metastasis. For example, we can study whether there exists a correlation between numbers of CTCs detected and disease progression.

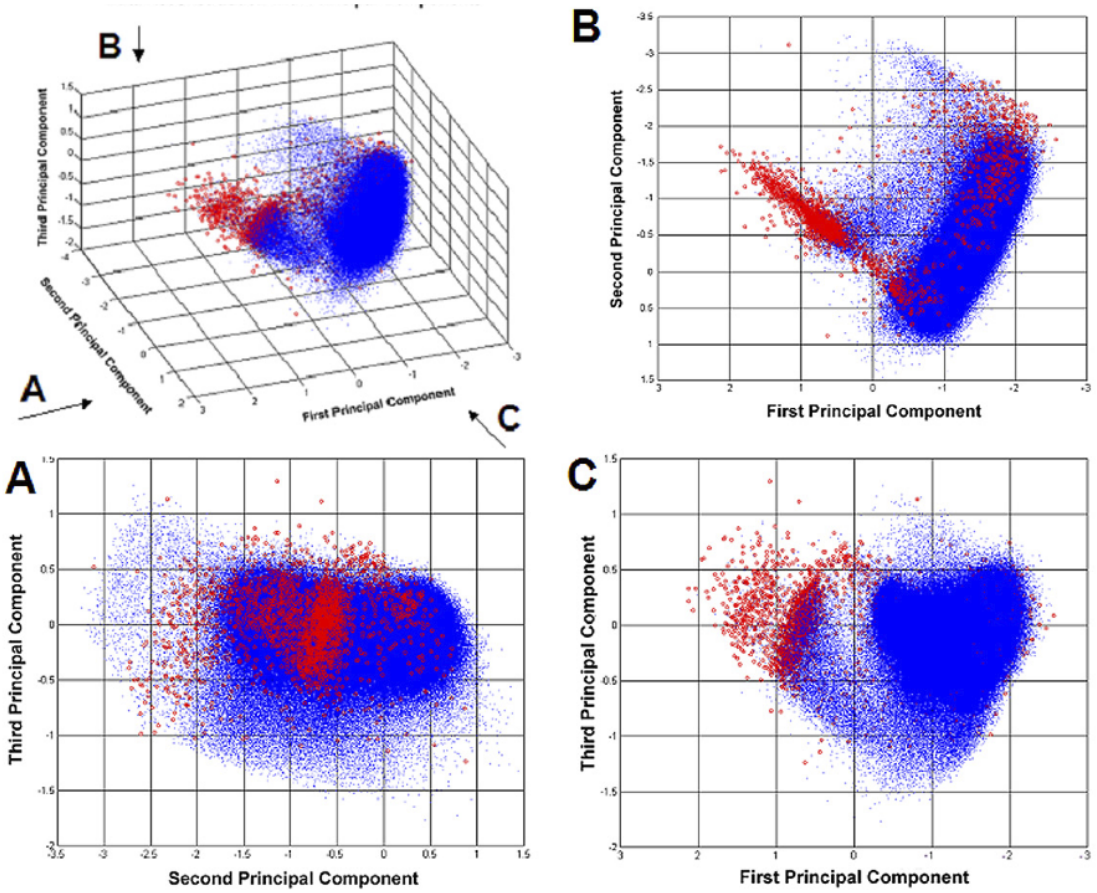


Figure 2.11: PCA components of the *in vivo* melanoma mouse model. Cancer cells are plotted in red while WBCs are plotted in blue according to the GFP ground truth. The three dimensional view of all three components is shown in the top-left panel. The rest of the panels are two-dimensional projections of PCA data along the direction illustrated in the three-dimensional plot. These plots illustrate that there are still many melanoma cells among the main cluster of WBCs.

Table 2.4: Classification performance of sequential PCA application on the *in vivo* melanoma mouse model. The cell line used was melanoma B16F10. The performance of the second PCA is calculated with respect to the original data set (i.e. after applying two PCAs sequentially).

	First PCA	**Second PCA**
Accuracy (%)	13.67	25.87
Sensitivity (%)	68.87	65.84
Specificity (%)	98.75	99.46
True Positives	1637	1565
True Negatives	819064	824921
False Positives	10342	4485
False Negatives	740	812

Chapter 3

Microscope Cell Sorting

To study rare circulating tumor cells in blood, it is not merely sufficient to detect and locate them on a microscopy slide. Ideally, these cells of interest should be physically isolated with high purity to enable further biological studies such as Polymerase Chain Reaction (PCR) for specific genes or microarray studies of the expressional profile of the cells. The current cell-separation technologies developed for diagnostic or research applications include fluorescence-activated cell sorting (FACS) and magnetic-activated cell sorting (MACS). Both methods offer throughputs (>10,000 cells/sec) that surpass the requirements of our application of interest [85, 86]. However, both of these methods lack the purity and sensitivity that is required in an ultra-rare model such as that of circulating tumor cells. At best, custom single-cell FACS machines can sort samples as rare as 1 in 10,000 cells up to %99 purity on [85]. This specification renders the current FACS- and MACS-based separation methods insensitive for sorting rare cells in blood that can be as rare as 1 in 1,000,000. Since the separation technique to be developed had to follow classification methods that are based on cytometric measurement of microscopy images, one requirement of the technique should be its ability to collect cells directly off of a microscope slide. One such technique that meets this requirement is laser microdisection that was originally designed for isolation of regions of interest from tissue sections [45]. This commercially available method, however, requires a dedicated microscope base or extensive customization of the optics of an existing microscope. Because of this limitation and given the practical

and financial requirements, using laser microdisection was not a viable option. Therefore, an inexpensive and ubiquitous technique was needed to manipulate and isolate individual cells which could be set up on any microscopy platform.

3.1 Methods

3.1.1 Mechanical Setup

To collect the cells, a glass micropipette with the inner diameter of 20-30 microns (Humagen Fertility Diagnostics, VA, USA) was attached to a micropipette collar which in turn was connected via rigid plastic tubing to a CellTram Vario Microinjector (Eppendorf, Germany). The microinjector is a syringe that has a finely controlled plunger which is turned, rather than pushed or puled, to displace fluid content. The specifications of the microinjector allows for 960 nL volume transfer per revolution and a minimum volume transfer of 2 nL. The common sizes of the cells to be collected was estimated at 30 microns or a volume of 0.1131 nL assuming a spherical shape. The microinjector's smallest volume transfer of 2nL or 20 times the volume of an average cell. This was thought as a good initial guess for the magnitude of the volume that would need to be aspirated in order to detach the cell from the glass surface of the coverslip. The microinjector was filled with oil. The microinjector was setup on an Eidaq100 High Throughput Microscope (Q3DM, San Diego, USA) to enable automated rare cell detection and retrieval on the same instrument. A metal post was used to raise and hold the micromanipulator at the hight necessary to reach the stage of the Eidaq100. The Eidaq100 system is controlled via its hardware control suite Cytoshop (Ver 2.0, Beckman Coulter, USA) that enables acquisition as well as segmentation and cytometric analysis of the acquired images.

3.1.2 Computer Control Scheme

The system was automated to lift the glass micropipette, navigate it in the horizontal plane and to aspirate to collect cells. The general schema of the control

system is illustrated in 3.1.

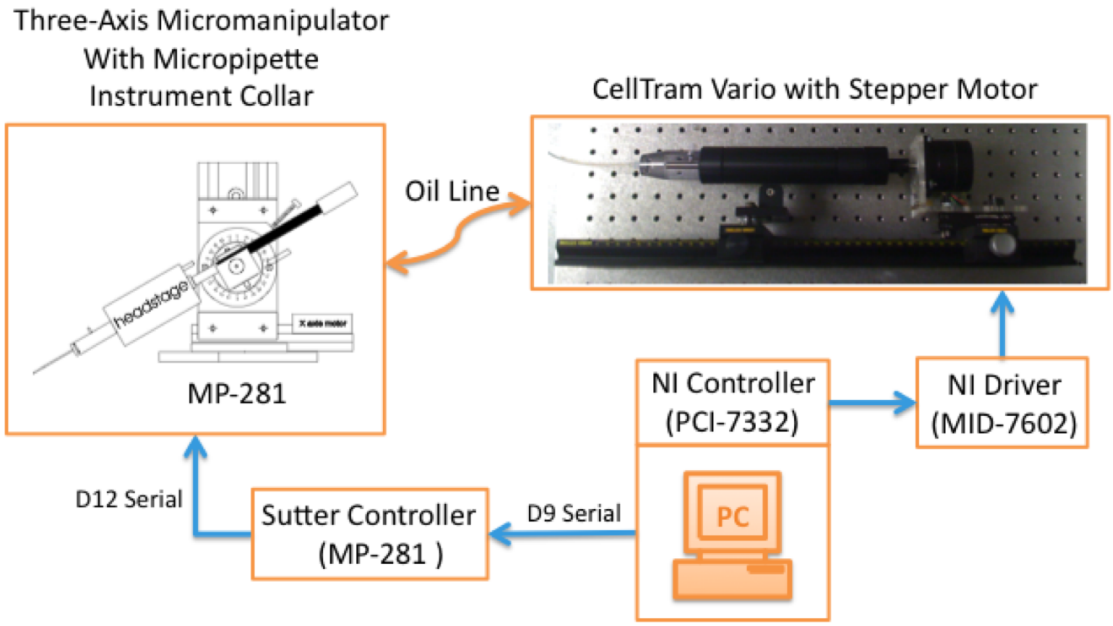


Figure 3.1: Control scheme for automation of the cell isolation system.

Automation of the Syringe

The CellTram Vario Microinjector offers a coarse control and a fine control for the aspiration of its content. The fine knob allows for aspiration resolution of up to 960 nL/revolution. This knob was connected via a joint shaft connector to a stepper motor (Vexa, USA) (Figure 3.2). LabView (Version 8.0, National Instruments, USA) was used to control the stepper motor. A LabView Virtual Instrument (VI) was made along with custom controller functions to communicate with the stepper motor (Figure 3.2). After some experimentation, the velocity of the stepper motor was set to 20,000 steps/sec. A VI, called *CellTramControl.vi* was designed to perform the aspiration at a set stepper size, velocity and number of steps every time it was invoked.

Automation of the Micromanipulator and the Stage

A Sutter MP-281 (Sutter Instruments, California, USA) micromanipulator and controller combination unit was used to navigate the glass micropipette in three axes. The MP-281 unit is comprised of the three-axis kinetic head, a three-axis rotary optical encoder for manual operation and a control box. The kinetic head has a moving plate with a dovetail interface to which an instrument collar is attached. The dovetail mounting allows for precise adjustment of the angle of attack for the instrument collar. For this study, we used the instrument collar provided with the CellTram Vario microinjector (Figure 3.4). A LabView VI, *MP281Control.vi*, was written for basic control of the MP-281 using serial interface (Figure 3.5). Functions were made to calibrate the micromanipulator upon startup and to navigate to desired positions during each call.

Stage automation was achieved by accessing the National Instrument control board inside the Eidaq100 using. A VI named *StageControl.vi* was written to navigate the stage to a given x and y position. Positions of cells were obtained from segmentation of nuclear images in CytoShop. An offset was introduced for each direction to account for differences in calibration. The offset was by moving the stage to a given cell and verifying the position of the cell by aligning it to the center of the field of view through the eye-pieces. This offset was found after each scan and was part of the procedure for switching to retrieval mode.

Collection Sequence

LabView VIs, to control the stage, syringe and the micomanipulator, are used to make collection of cells easy. The following is a sequence of events generally followed in an experiment: 1) The slide was scanned and analyzed using Cytoshop (Ver 2.0, Beckman Coulter, USA) software. A list of nuclear coordinates was generated. 2) Cytoshop is terminated and the Automated Sequence LabView VI was launched. 3) The desired cell is identified under the eyepieces and the stage is moved using *StageControl.vi* interface to bring the cell close to the center of the field of view. An offset can be introduced here if the desired cell is not positioned in the center of the field of view. 4) The micromanipulator, which was previously

centered at the center of the FOV, is lowered until the tip touches the cell. 5) *CellTramControl.vi* is used to aspirate the cell, while observing the cell carefully to ensure it is detached from the substrate and is sucked into the micropipette. 6) If the cell remains attached, step 3 should be repeated until successful. 7) The micromanipulator is lifted a few tens of microns while still keeping the tip under the surface of the liquid surface. 8) At this point, another cell can be collected by starting at step 1. Once the collection is finished, the direction of the Vexa stepper motor can be reversed to push out the collected cells into a new container for storage or culture. Also, the tip can be broken into a centrifuge tube and cell can be lysed directly for biological experiments such as microarrays.

3.2 Proof of Concept Experiment

3.2.1 Experimental Design

A proof of concept experiment was designed to show that the retrieval processes works as intended. The criteria for success in the experiment is to visually confirm that cells are collected in the micropipette. To do this, we follow the steps presented in 3.1.2 to collect fluorescently-labeled cells.

3.2.2 Cell Preparation

Multiple cell lines, most prominently 3T3 and SW1, were used for testing and validation of the system. Cells were fixed with 4% paraformaldehyde for 1 hour, then counterstained with DAPI for fluorescent imaging. Adapted from Hamada[39], the preparation of DAPI solution consisted of 75ng/ml DAPI, 10nM Tris, 10nM EDTA, 100nM NaCl, and 2% 2-mercaptoethanol. The fixing and staining were done after live cells were centrifuged onto a single-chamber Lab-Teck II^{TM}Chambered Coverglass (Cat#155360, Nunc, Thermo Scientific, USA). The DAPI solution was kept in the chamber and not washed away to compensate for the effect of photobleaching throughout the experiment. The chamber was then set in the stage insert of the Eidaq100.

3.2.3 Results

Scanning, analysis and collection of cells were done according to the sequence previously described. After timing the collection of about 50 cells, the throughput of collection was found to be 20 cells/sec which does not include scanning, analyzing and calibration time. In a separate run, after some cells were collected into the glass micropipette tip, the tip was brought into focus and the transmitted light source was turned off. The tip was viewed under fluorescence setting for DAPI and an image was captured. Figure 3.6 shows the tip of the micropipette and cells that are collected. The cells are DAPI-stained and hence their nuclei is visible in the image. Under fluorescent light, the tip is only visible due to the scattering of fluorescent signal from cells and perhaps autofluorescence of the glass. Therefore, the grayscale image in Figure 3.6 is inverted and its contrast adjusted to make the tip more visible. The few cells collected in the tip of the pipette are also visible. As more cells are collected, the cells form a line inside the lumen of the pipette.

3.3 Discussion

Since the selection of cells in this project is based on morphology quantified from microscopy images, the method desired here was one that could be implemented directly on a microscope. Most of the currently available microscopy specimen collection methods are laser-based. For example, laser microdissection uses a high powered through-the-lens laser beam to cut the area around a desired region and physically catapult the region off of the slide. Such methods require a dry sample prepared on special substrates which may interfere with the quality of collected specimen for down-stream testing. These methods also require extensive customization of the optical instrument to implement the laser light path, control the laser and physically collect samples. This makes their implementation both expensive and intrusive. The method proposed here uses very few components that are inexpensive and off-the-shelf. Other than for controlling the stage, the method proposed here does not interface in any way with the microscope. Therefore, the

system is flexible for implementation onto any microscopy system with minimal customization.

As compared with the laser-based methods, the method designed here has inherently lower throughput as it requires some user supervision and control for collection of each cell. As a trade-off, however, the system allows for visual verification of cells collected and therefore can be potentially very specific and sensitive in collecting a valuable subpopulation of cells. Overall, the system described here is an inexpensive and easily implementable system for semi-manual collection of cells that could be further automated for better throughput.

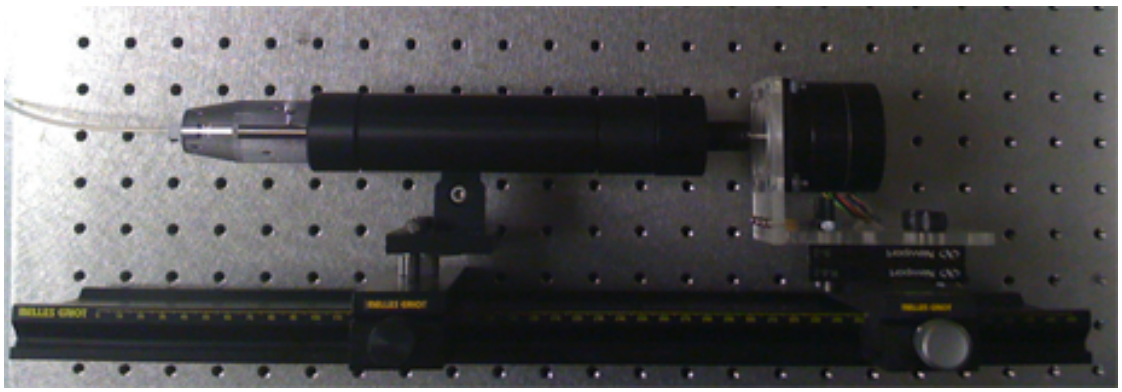


Figure 3.2: Fluidic setup of the microscope cell sorter. The CellTram Vario is shown coupled with a Vexa motor.

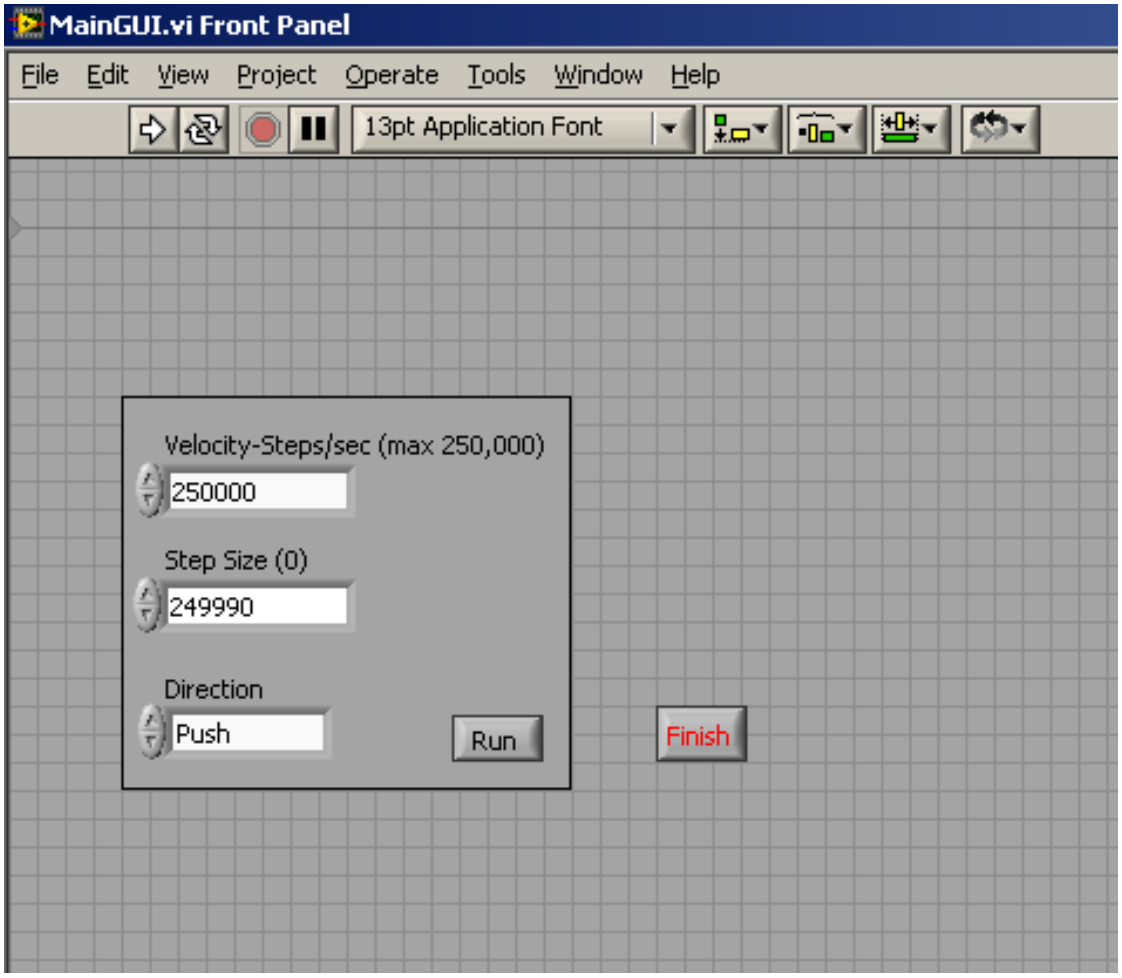


Figure 3.3: Screenshot of the VI used to control the syringe setup.

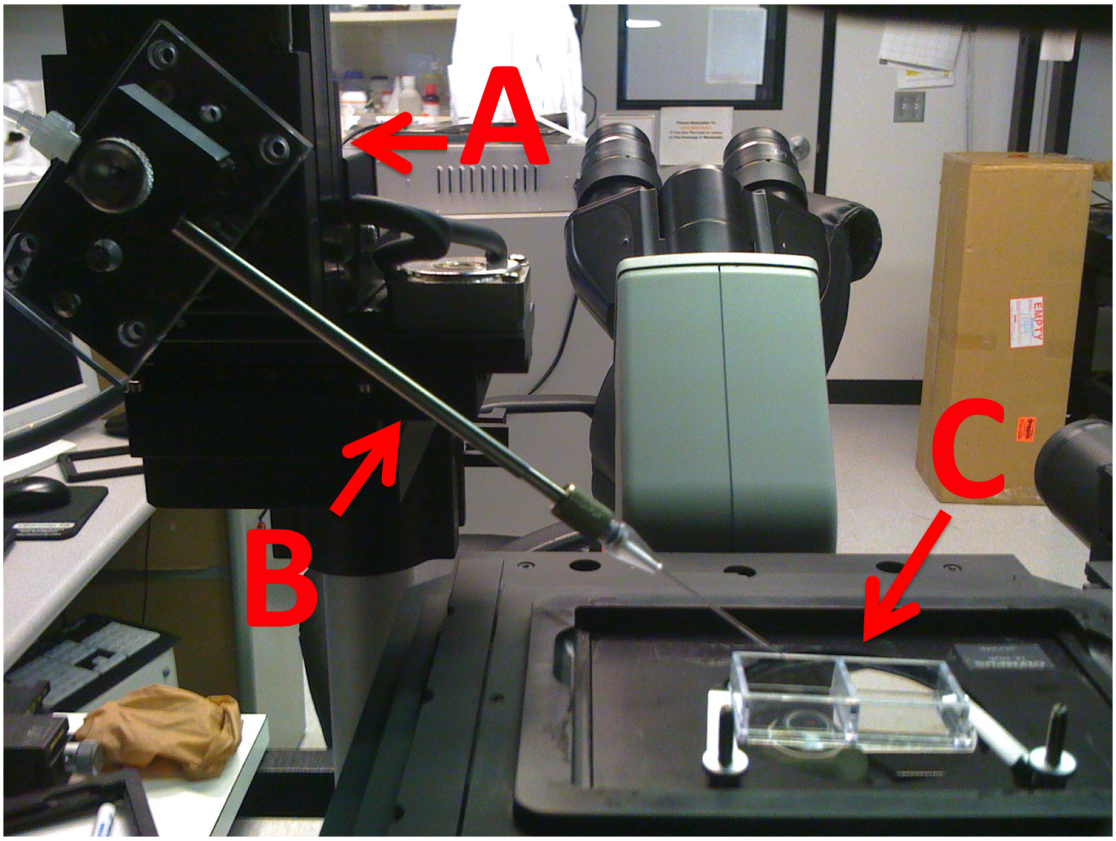


Figure 3.4: The kinetic plate setup of the microscope cell sorter. The dovetail (A) and an instrument collar attached (B) is shown. The attached glass micropipette is lowered into a chambered coverglass mounted on the stage of an Eidaq 100 (C).

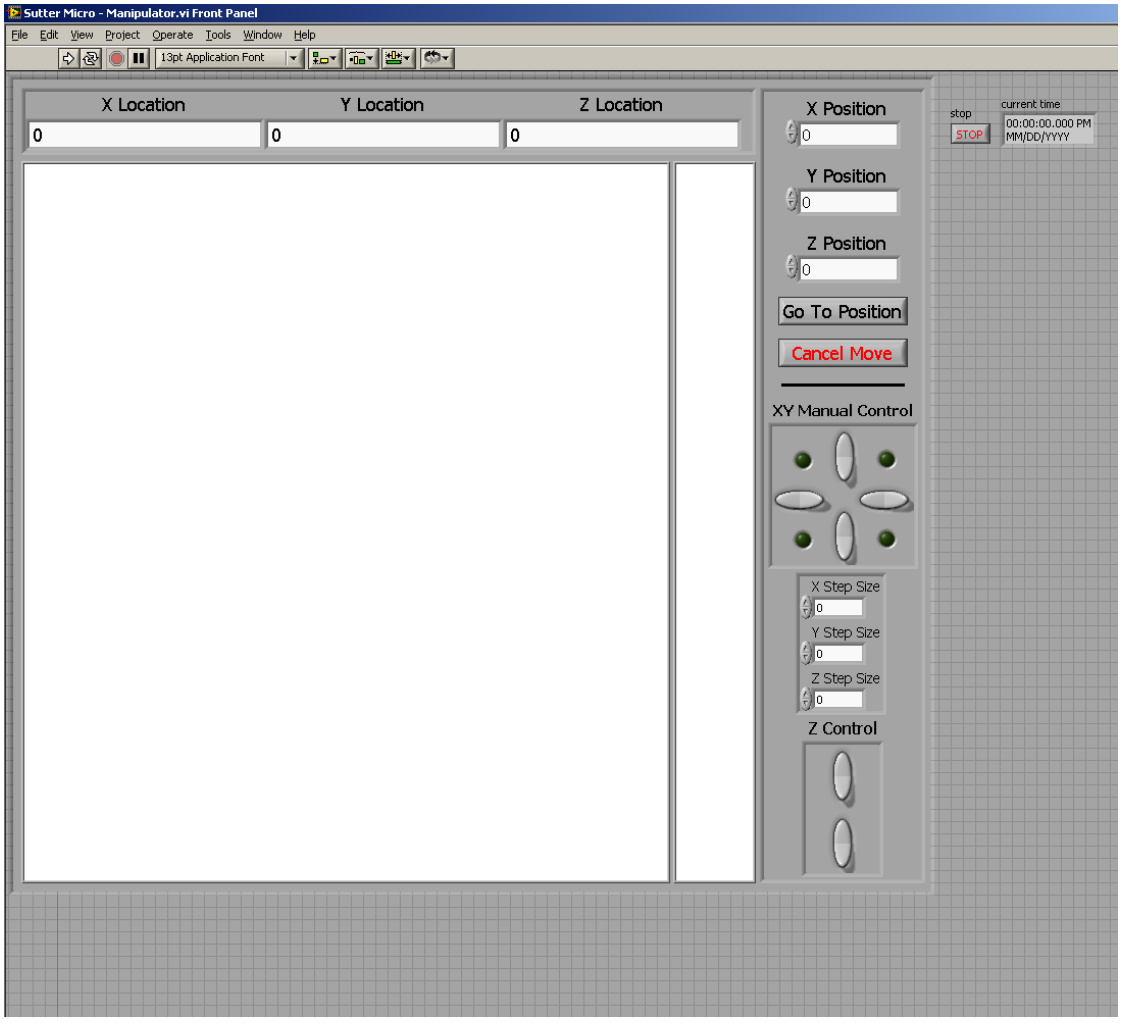


Figure 3.5: Screenshot of the *MP281Control.vi*, the VI used to control the MP-281 micromanipulator.

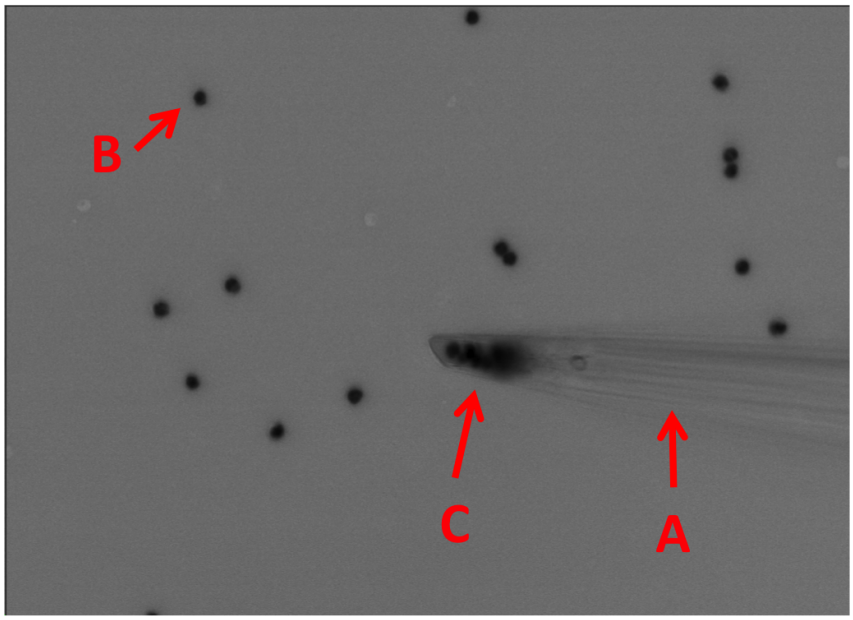


Figure 3.6: Micropipette tip containing collected DAPI-stained cells. The grayscale image has been inverted for better viewing of the glass pipette (A). DAPI-stained nuclei are visible on the coverglass (B) and also collected in the tip of the micropipette (C).

Chapter 4

Morphometric Characterization of Cytotoxicity

4.1 Introduction, Background and Significance

Detecting and quantifying toxicity of compounds have become an important challenge in small molecule screening. Early detection of toxic compounds in the drug development pipeline can significantly reduce cost and development time by helping to prevent toxic compounds from advancing into becoming drug candidates, and even approved pharmaceuticals. Toxic compounds that do advance through the discovery process often pose great cost as they fail in early trials, due to their toxicity, or cause cardiac, hepatic and renal toxicity after administration. With the advent of image-based high-content screening (HCS), assays can be implemented in parallel with drug screening to acquire toxicity data and prevent toxic compounds from advancing into the discovery pipeline. These methods often involve adding additional, commercially available, fluorescent markers for detection of toxicity and cell death. However, in high-throughput HCS (HT-HCS), where many hundreds of thousands of compounds are tested in a given screen, multiplexing a toxicity screen that involves additional biological preparation and staining can be prohibitively. Therefore, a cytotoxicity screen is desired that is sensitive, accurate and requires minimal additional preparation and cost.

It has been long known that cell death affects cytoplasmic and nuclear morphology and that morphology can be used as a method of assessing cell death and toxicity. For example, Clarke et. al. reported in 1990 in a review article that the three types of cell death known then (apoptotic, autophagic, and non-lysosomal vesiculate) were detectable by visual assessment of cell morphology from electron micrographs [23]. Specifically for apoptosis, the changes in the morphology of the nucleus, chiefly chromatin reorganization and aggregation, has been widely studied with both electron and transmitted light microscopy using immuohistochemistry (IHC) [57, 10]. Cell toxicity and death have also been assessed using fluorescent microscopy [92, 93, 112, 38, 80]. Fluorescent microscopy, as used in HCS, enables extraction of numeric measurements of the image and region of interest (ROI) that can potentially be used as metric for biological phenomena. Segmentation is used to detecting pixels that belong to background and those that belong to ROI (cells, nuclei, etc.) and morphometric and fluorometric measurements are calculated for each ROI. This affords HCS with an unprecedented sensitivity and specificity in assessing biological events. For cytotoxicity, when compared to plate-readers where only a a single fluorescent measurement is recorded for each experiment, HCS was shown to have higher sensitivity (90% versus 25%) and higher specificity (98% versus 90%) when using specific fluorescent toxicity markers [80].

To assess toxicity in HT-HCS, therefore, nuclear fluorometric and morphometric features are good candidates for use as predictors especially since a vast collection of them can be easily obtained using existing algorithms [46, 43]. Databases resulting form these numerical measurements can be vast because they are made for each cell or even sub-cellular entity. As a result, finding and analyzing biologically relevant features becomes the key challenge of performing meaningful analysis based on HT-HCS data [31, 7]. Often, measurements made for cells that are stained with more than one marker have redundancies in their information content in a way that one could replicate a certain feature using a subset of the rest of the features. For example, Loo et. al. showed that in heterogenous cell populations stained with a collection of markers, one can eliminate the data from one of the discriminating markers and still recreate the subpopulations with the

remaining marker set by classifying the cells in a feature space that includes phenotypic features [69]. Such studies suggest that cytometric feature space of cells are highly redundant and potentially valuable metrics can be found from the readily available features that are seemingly irrelevant. In all HT-HCS projects, nuclear images of often DAPI-stained cells are collected to perform segmentation are left largely unanalyzed as a byproduct. Since toxicity has great effects on nuclear morphology [43], we hypothesize that nuclear features could be used as a means of calculating a metric that is useful in evaluating cytotoxicity.

4.2 Methods

The flow chart illustrated in Figure 4.1 shows the overall process of calculating a mimicked Cytotoxicity Index (CI) that mimics a given toxicity marker and is only calculated based on nuclear features of DAPI-stained cells. This cytotoxicity index is found and verified using dose responses of known toxic compounds.

4.2.1 Mathematical Notation

Cytometric data are often obtained from multi-wellplate experiments, where each well contains many cells. Because attributes are measured at various levels (i.e., for individual cells, wells and plates) it is necessary to define a set of notations that succinctly and accurately describes and how each term is calculated. For an attribute A, therefore, we can describe a set of variables to accurately define the domain for which the attribute is calculated:

Range is the scope for which the attribute was calculated. Although other scopes are possible, this variable usually takes the form of $Cell$, $Well$ or $Plate$. For instance, the $Range$ of an attribute calculated for a single cell is $Cell$ while the $Range$ of an attribute measured for each well (e.g., number of cells) is $Well$.

Treatment is the experimental condition under which the cells were kept throughout the experiment. Only differential conditions are described here and treat-

ments common to all cells in the experiment are omitted from this description. For example, it is only necessary to define a positive control $Treatment$ in terms of how it is different from the negative control and further information is not necessary for the purposes of notation.

Instance is the mathematical index of the item for which the attribute is measured. For example, $Instance$ can be the cell index i for an attribute calculated for the ith cell or the ith item in the $Range$.

It is worthy to note that not all of these variables are necessary to describe all of the attributes presented here. In most cases, it is trivial to assume what the variable is and its explicit indication is not needed. However, if we were to illustrate the attribute A with all its variables, we would illustrate it as:

$$^{Range}_{Treatment}A_{Instance}$$

This notation can also be expanded to the functions of the attributes such as their mean and standard deviation. For example, the mean of the ith attribute of across all cells in a well which which contains compound A can be expressed as:

$$^{Cell}_{CompoundA}\mu(Attribute_i)$$

4.2.2 Preparation of $GuidePlate$ with $Target$ Cytotoxicity Measure

The method for calculation of CI employed in this study finds a subset of *Attributes* that contain cytotoxicity information. The criteria used for this selection is how well they correlate with a verified measure of cytotoxicity, or a $Target$ measure, such as obtained form staining for well known toxicity markers (e.g., TUNEL, Caspase, Cytochrome C, etc.). The method, therefore, requires a $GuidePlate$ to be prepared with the cytotoxicity staining to be mimicked ($Target$ measure) and a nuclear stain to obtain *Attributes*. The *Attributes* are then analyzed, selected and put together to create a CI which closely resembles the $Target$ measure.

To induce toxicity we prepared dose responses of MG132, Staurosporine and Taxol. Table 4.1 illustrates these toxins, their mechanism of action and their EC50 (concentration at which half the cells are dead after 24hr incubation). The dose responses were prepared to cover the full range of the toxic concentration, especially the lower concentrations where earlier responses to the toxin were recorded. This is due to the fact that the end-stage cell death appears more similar among all toxic responses as accompanied by complete fragmentation and disintegration of nuclear envelope. The earlier responses, however, could be distinct among different toxins. Note that the EC50s given in Table 4.1 are not necessarily for the cell lines used in this experiment as EC50s for all cell lines are not available for a given toxin through literature search.

Table 4.1: Toxins used to cause cytotoxicity.

Toxin	**Mechanism of Action**	**EC50 24hr incubation**
MG132	Proteosome inhibitor and autophagy inducer [41].	$5.0\mu M$ for HeLa cells [41].
Staurosporine	Protein kinase inhibitor and apoptosis inducer [115].	0.5-1.0μM for cardiomyocytes [115].
Taxol	Mitosis inhibitor [70].	2.5-7.5nM for eight various tumor cell lines [64].

A number of the commercially available toxicity kits were used to study how well the CI method can mimic their measure of toxicity. The following is a list of these kits and the corresponding protocol that was used.

TUNEL is a method of detecting DNA fragmentation by staining the open nicks of fragmented DNA using a fluorescent dye [34]. This method is widely used to assess cell toxicity. However it is incapable of distinguishing between different types of cell death [3]. A plate of AML12, a mice liver cell line[109], was prepared according to the protocol used in the PubChem screen (AID: 1656) [33] except for lipid staining. The DAPI staining was postponed until after the TUNEL protocol. Next, the plate was stained with the

In Situ Cell Death Detection Kit, Flourescein (Roche, Cat#11684795910, Mannheim, Germany). In short, the fixed cells were washed with PBS and incubated in the permeabilization solution for 2min on ice. The solution was then removed followed by another PBS wash. About $50\mu L$ of the TUNEL reaction mixture was added to each well and the plate was incubated in a cell culture incubator for 1hr. The wells were then washed three times with PBS and stained with DAPI as described in [33].

Cell Permeability Dye is a dye that fluoresces once inside the cell. A healthy cell is supposed to be impermeable to the dye, while the compromised membrane of a treated cell will let the dye in. The dye is often designed to only fluoresce when inside the cell. For the experiments described here, the Cell Permeability reagents form the Multiparameter Cytotoxicity Kit (Cellomics/Thermo Scientific, Cat#8408001, IL, USA) was used to stain AML12 cells. The fluorescent dye was designed to excite at 491 nm and emit at 509 nm. The staining was done in a 384 well plate and according to the provided protocol and in conjunction with the other staines contained in the Multiparameter Cytotoxicity Kit, namely Cytochrome C (550/568nm), LC3B (646/674nm) and DAPI (358/461nm).

Cytochrome C is one of the most common commercially available apoptosis markers. When apoptosis is initiated, Cytochrome C detaches from the inner mitochondrial membrane where it is part of the electron transport chain. It then diffuses to the cytoplasm and nuclei and its translocation is a common marker for apoptosis. In this experiment, the fluorescent conjugated antibodies from the Multiparameter Cytotoxicity Kit (Cellomics/Thermo Scientific, Cat#8408001, IL, USA) were used to stain AML12 cells for Cytochrome C. The dyes used were to fluoresce in 568 nm and excite at 550 nm. The staining was done in a 384 well plate and according to the provided protocol and in conjunction with the other staines contained in the Multiparameter Cytotoxicity Kit, namely Cell Permeability Dye (491/509nm), LC3B (646/674nm) and DAPI (358/461nm).

4.2.3 Imaging Protocols

The *GuidePlate* and any plate for which the CI is to be measured should be imaged with exactly the same instrument and imaging parameters. For this study, depending on the experiment the plates are imaged using one of the following protocols.

For the experiments involving the Lipid Droplet screen and TUNEL mimic, the plates were imaged on an Opera QEHS (Perkin Elmer) with a 20x 0.45 NA air objective. Images were binned 2-by-2 for an image size of 688 by 512 pixels. The nuclear channel was acquired using 365 nm Xenon lamp excitation and 450/50 emission filters. Two fields per well were acquired for the screening plates and nine fields were acquired for the *GuidePlate*.The integration times were set so that less than 5% of pixels are overexposed.

For other experiments that involve the Cytochrome C and Cell Permeability Dye, plates were imaged on an IC100/Eidaq100 (Beckman Coulter/Q3DM, USA) using a Nikon Plan Fluor 20X 0.50 NA air objective. Nine nuclear images per well were acquired in 8-bit mode on a ORCA-ER camera (Hamamatsu, Bridgewater, NJ) using standard DAPI filter cubes (Chroma Technology, Brattleboro, VT). The integration times were set so that less than 5% of pixels are overexposed.

4.2.4 Calculation of the Cytotoxicity Index

The cytotoxicity index (CI) is defined here as a single numerical value calculated based on easily-measurable *Attributes* of a cell and is designed to replace or mimic the *Target* cytotoxicity measure that is expensive or not feasible to implement in hight-throughput screening. The CI is measured from a subset of attributes with trends that correlate well with the direct cytotoxicity measure to be replaced. For a description of how such attributes are selected please refer to the section 4.2.7. In order to make CI a controlled metric, the attribute of each cell is first converted to z-scores with respect to the mean and standard deviation of of the same attribute of an untreated population of cells as illustrated in Equation (4.1). The weighted Euclidian distance, Equation (4.2) was then used to combine the *Attributes* into a single measure. In that equation, W_a is the weight of each

attribute determined manually or, as described later, by finding the correlation coefficient with respect to a *Target* measure. A cell with *Attributes* equal to the mean of the *Attributes* of an untreated population will have a very small or zero z-score for each *Attribute*. Therefore, its CI will be close or equal to zero and the metric itself can be viewed as distance away from the untreated population.

$$ {}^{Cell}_{Treated}Z_a = \frac{|{}^{CellPopulation}_{Untreated}\mu(Attribute_a) - {}^{Cell}_{Treated}Attribute_a|}{{}^{CellPopulation}_{Untreated}\sigma(Attribute_a)} \tag{4.1} $$

$$ {}^{Cell}_{Treated}CI = \sqrt{\sum_{a=1}^{Number of Attributes} W_a \times \left({}^{Cell}_{Treated}Z_a\right)^2} \tag{4.2} $$

4.2.5 Calculation of Attributes

Attributes are defined as metrics of a cell or a collection of cells that are readily available and are to be used to mimic and replace the expensive toxicity measure. In this application, these metrics include morphometric and fluorometric features of a cell as well as their corresponding components obtained through Principle Component Analysis (PCA) or Independent Component Analysis (ICA) or any other space-transformation operation.

Features

The degradation of the nuclear structure as a result of cell death, in particular apoptosis, and the morphology characteristics of it is well documented in literature [92, 93, 53].

In fluorescent microscopy, these late-stage cell death morphologies are characterized by reduction in nuclear size (change in nuclear area), condensation of DNA (bright spots), and nuclear fragmentation (nuclear blobs) as illustrated in Figure 4.2 for J774 cells treated with Taxol and MG132 and stained with DAPI.

For example, the commercially available cytometry software package CytoShop (Ver 2.0, Beckman Coulter, USA) produces the nuclear features illustrated in Table 4.2 by default. For the experiments involving the *a priori* feature set (see Section 4.2.7), CytoShop was used to generate features values. For all other

experiments, a custom pipeline was created in CyteSeer (Ver 2.0.4, Vala Sciences Inc., San Diego, CA, USA) to recreate the features listed in Table 4.2. Segmentation was performed using CyteSeer's Otsu method with sensitivity and minimum object size set at 80% and 5 pixels, respectively.

Multivariate Component Analysis

The cytometric features calculated for each cell describe the position of the cell in the multi-dimensional, or multivariate, feature space. Since one can theoretically calculate an unlimited number of features for each cell, it can be deduced that the features produced may not be independent of each other. Also, while some features could indicate toxicity, others may contain little or no toxicity information. As a result the desired toxicity information may be hidden in multiple features at once. Furthermore, the toxicity information itself may be a multivariate space, requiring more than one dimension of data to accurately describe. Using only the dimension of the feature space can therefore be unwise as the toxicity information may remain contained in multiple features and unrevealed. A common way to investigate the dimensionality of a multivariate space is to employ factor analysis and in particular PCA [51] and ICA [25]. These methods transform a multivariate feature space to a new space where dimensions (components) satisfy criteria such as independence or maximum data content. The components and the original dimensions are linear combinations of each other, and a transformation matrix stores the coefficients of the linear transformation. The transformation is often reversible by multiplying the components and the inverse of the transformation matrix through which the original data can be reproduced in a lossless fashion. Since the dimensionality of the dataset is limited to the number of existing dimensions, the transformed space can contain a lower number of dimensions compared to the original space.

The dimensions resulted form PCA are created with the criteria of being orthogonal and having the highest data content. The components are produced such that their variance is the maximum present in the data, where variance is assumed to be the primary indicator of data-content. PCA components are then

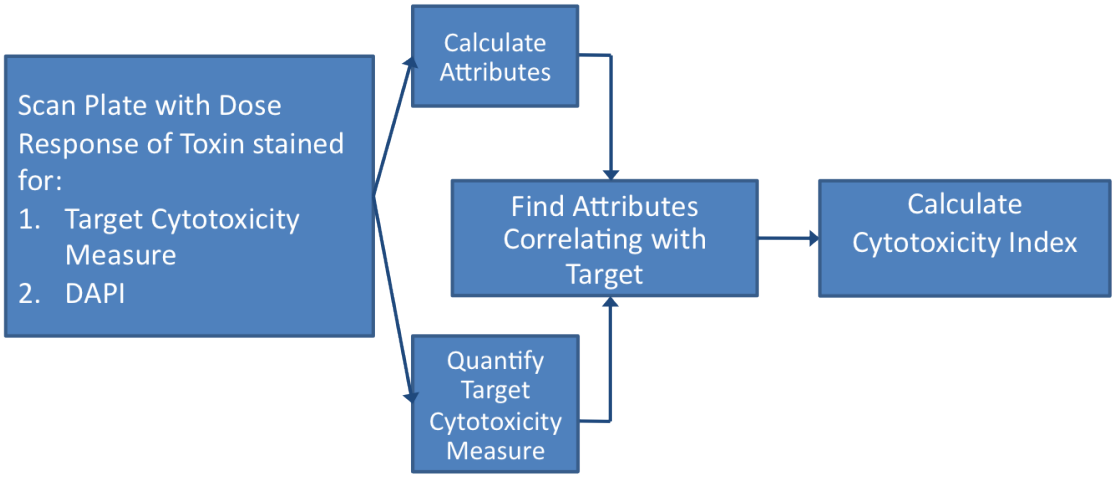


Figure 4.1: Flowchart showing the process involved in calculating CI. The process starts with the scanning and cytometric analysis of a *GuidePlate* prepared as described in Section 4.2.2 and finishes with the calculation of CI. The list of significant *Attributes* found is saved and is later used to calculate CI in HT-HCS mode.

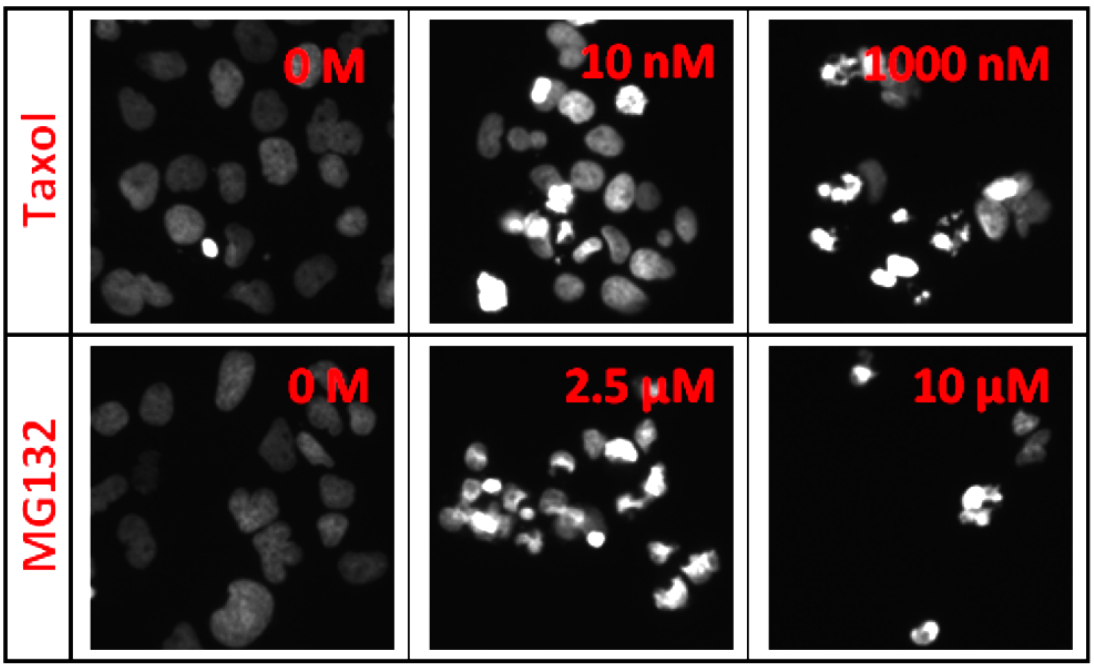


Figure 4.2: Typical end-stage toxicity effect of Taxol and MG132. The effect is shown on J774 cells stained with DAPI. Nuclear condensation and fragmentation is apparent with higher treatment concentration.

Table 4.2: List of default nuclear features calculated by CytoShop 2.1 (Beckman Coulter, USA).

Feature	**Description**
X-size	X-size of nucleus in pixels.
Y-size	Y-size of nucleus in pixels.
Width	X-size of nucleus in microns.
Height	Y-size of nucleus in microns.
Area	Area of nucleus in microns^2.
Area**1/2	Square root of area of nucleus.
Perimeter	Perimeter of the nuclear mask.
Wiggle	Perimeter divided by area of the nucleus. Measures how convoluted the boundary of the nucleus is; high for lobed nuclei.
Wiggle_Nrm	Perimeter divided by the square root of the area of the nucleus
Integ_gs	Integrated grayscale of nuclear brightness.
Aver_gs	Average grayscale of nuclear brightness.
Percents_0	Pixel intensity that is darker than 0% of nuclear pixels.
Percents_5	Pixel intensity that is darker than 5% of nuclear pixels.
Percents_25	Pixel intensity that is darker than 25% of nuclear pixels.
Percents_50	Pixel intensity that is darker than 50% of nuclear pixels.
Percents_75	Pixel intensity that is darker than 75% of nuclear pixels.
Percents_95	Pixel intensity that is darker than 95% of nuclear pixels.
Percents_100	Pixel intensity that is darker than 100% of nuclear pixels.
IQ_range	Interquartile range measures spread of dispersion. It is the difference between the Percent_75 and Percent_25.
Variance	Variance of intensity of pixels in the nucleus.
StdDev	Standard deviation of intensity of pixels in the nucleus.
AbsDev	Absolute deviation of intensity of pixels in the nucleus.
CM3	Third central moment of pixels in the nucleus; Skewness, or measure of asymmetry in probability distribution.
CM3**1/3	Third root of CM3.
CM4	Fourth central moment of pixels in the nucleus; Kurtosis, or measure of 'peakedness' in probability distribution.
CM4**1/3	Third root of CM4.

sorted in terms of their variances, with the first component containing the highest possible variance of the original feature space. As a result, if significant toxicity variation exists in the data set, PCA is useful to extract and concentrate it in a few components. PCA involves a closed form mathematical operation with negligible computation demand. In this study, functions from the Statistical Toolbox 7.3 of Matlab 2009b (Mathworks, USA) are used to preform PCA.

ICA transformation, on the other hand, aims at minimizing the dependence between the resultant components. The ICA components are not necessarily orthogonal and are not sorted based on variance. ICA is therefore a good method for separation of sources in mixed feature sets. This makes ICA useful when the toxicity data exists in the feature set along with other significant data trends. Unlike PCA, the solution to ICA is not a close form solution and requires iterations. In this study, a Matlab implementation of the algorithm known as "fastICA" [42] is used to find the ICA components.

4.2.6 Data Normalization

Cytometric *Attributes* can have very different magnitudes and ranges (Figure 4.3). Equation (4.2), uses the z-scores to transform all features to the same range and ensure *Attributes* with intrinsically large magnitudes do not dominate those will smaller magnitudes which may be more significant. However, if the internal deviation (i.e., standard deviation) of the distribution of the *Attribute* is naturally much wider than that of another *Attribute*, the z-scores of that *Attribute* can still dominate in the CI equation just as its magnitude would have it z-scores were not used. To remedy this, all data input to the process were normalized in the following manner. First, the mean of the *Attribute* from the untreated cell population was subtracted. Next, the values corresponding to the 10th and 90th percentile in rank order of the *Attribute* of the untreated cell population was found (minus the mean) and the *Attribute* to be normalized was linearly transformed so that those values from the untreated cell population were aligned to preset values of 0.1 and 0.9. This way the z-score found to form a widely spread *Attribute* are less likely to dominate the CI, while significant changes in an *Attribute* that cor-

responds to toxicity are preserved. Non-gaussian distributions might still enable one *Attribute* to dominate over another in some situations, but this effect should be at least reduced.

4.2.7 Selection of Attributes

Once attributes are calculated for each cell, those that correlate with the target toxicity measure are selected and used to calculate the CI. Multiple correlation methods are used to help ensure accounting for as many modes of correlation as possible.

A priori Feature Set

It has been well understood that end stage toxicity is followed by degradation of the nuclear structure. Effects of toxicity of Taxol and MG132 on J774 cells are visible in Figure 4.2. These changes are characterized by reduction in nucleus size, condensation of DNA (high intensity spots), and nuclear fragmentation. Based on this preliminary understanding, an *a priori* set of nuclear morphology features, including Nuclear Area, Roundness (Perimeter over Area) and Chromatin Condensation (Percentage of pixels with maximum intensity in the nucleus) were used to calculate z-factors (number of standard deviations away from mean) for each cell, as expressed in Equation (4.1), with respect to the untreated control cell population. The resultant z-score was then averaged using Equation (4.3). In addition to the *Attribute* z-scores, we also calculate the z-score of the cell count according to Equation (4.4) as the cell number is lowered at the final stage of toxicity. Equation (4.5) was then used to calculate a Euclidian distance measure as the Cytotoxic Index (CI) for the well.

$$ {}^{Well}_{Treated}Z_a = \frac{\sum_{c=1}^{CellCount} {}^{Cell}_{Treated}Z_{a,c}}{CellCount} \tag{4.3} $$

$$ {}^{Well}_{Treated}Z_{CellCount} = \frac{|{}^{WellPopulation}_{Untreated}\mu(CellCount) - {}^{Well}_{Treated}CellCount|}{{}^{WellPopulation}_{Untreated}\sigma(CellCount)} \tag{4.4} $$

$$ {}^{Well}_{Treated}CI = \sqrt{\sum_{a=1}^{Number of Attributes} \left({}^{Well}_{Treated}Z_a\right)^2 + \left({}^{Well}_{Treated}Z_{CellCount}\right)^2} \qquad (4.5) $$

This cytotoxicity index was calculated for a murine macrophage cell line J774 and Hela Cells treated with a dose response of Taxol and is shown in Figure 4.4.

Selection of Attributes Based on Correlation

The *a priori* features were selected based on a general understanding of how the cells change shape as they die. However, this feature set may or may not be the optimum for gauging toxicity as different cell death mechanisms may be associated with different changes in the nuclear morphology. Furthermore, the individual features may not represent the best correlation to the $Target$ measure, but such correlation may arise in the set of *Attributes* calculated from the feature set as described above. In order to systematically select the best attributes, Pearson (Equation (4.6)) and Spearman (Pearson calculated for rank order) coefficients are calculated for each *Attribute* and the $Target$ measure. Pearson and Spearman coefficients are correlation coefficients that span [-1 1] with their sign reflecting whether the vectors are directly (+) or inversely (-) correlated. The magnitude of the coefficient expresses how well the values correlate, with 0 signifying lack of correlation. While Pearson takes in the actual values in the vectors being studied, Spearman uses their rank order. Therefore, Pearson coefficient quantifies linear correlation best, while Spearman reflects correlation regardless of linearity. *Attributes* with Pearson or Spearman coefficients above a certain threshold, set to 0.5 in this study, are labeled to be significant. An *Attribute* that is labeled significant for both Pearson and Spearman is labeled as linearly correlated, while one that is only significant in Spearman is labeled non-linearly correlated. *Attributes* with an insignificant coefficient in both are discarded. The significant features are then used to find the CI according to Equation (4.2).

$$\rho = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{X_i - \mu_X}{\sigma_X}\Big)\Big(\frac{Y_i - \mu_Y}{\sigma_Y}\Big) \tag{4.6}$$

4.2.8 Zprime Calculation

Zprime has been used frequently as a measure of the dynamic range in a biological assay [116]. It is defined as the distance between positive and negative controls in terms of the standard deviations of the two control distributions (Equation (4.7)).

$$Zprime = 1 - 3 \times \frac{(\sigma_+ + \sigma_-)}{\mid \mu_+ - \mu_- \mid} \tag{4.7}$$

The value of Zprime ranges from $-\infty$, for controls that overlay on top of each other and offer almost no dynamic range, to $+1$ where the distance of between the control means is three time the sum of their standard deviations. Zprime is a very stringent measure of dynamic range and almost no biological assay can obtain a perfect Zprime of $+1$. A Zprime of $+0.5$, for example, occurs when the means of the two controls are three times the sum of their standard deviations apart. To ensure statistical viability of a screen, biological assays that have Zprime lager than 0 and smaller that $+0.5$ are considered only in multi-plicate format, while assays with Zprime larger than $+0.5$ are often validated to run in HT-HCS without repeat.

For this experiments, the CI and the $Target$ measurements are found for each cell. The average of those values are averaged for each well. The mean and standard deviations were found for the average values of each group of wells with the same condition (treated with the same compound and concentration). These values were used to find the Zprime according to Equation (4.7).

4.2.9 Application to Small Molecule Library Screen

In order to find CI for a small molecule library screen, the list of significant *Attributes* should be obtained by preparing a separate microtiter plate, a *GuidePlate*, that is stained with the toxicity marker (e.g., TUNEL, Cytochrome

C, Permeability Dye, etc.) that is to be mimicked. The plate must also contain a control compound that induces the desired toxic response (such as shown in Table 4.1). Optimally the plate should contain a statistically viable dose response for the compound that is also reflective of the concentration of the compounds used in the screen. The plate should then be scanned with all channels required for the cytotoxicity marker in addition to the nuclear channel.

In this study, the CI algorithm is applied to a confirmatory screen of Lipid Droplets [33] comprised of three 384 well plates containing 869 compounds found to be positive for the Lipid Droplet screen. Each plate also contains untreated cells as controls. The confirmatory screen was run in duplicates (i.e., two plates containing the same compounds and compound map). Therefore, a total of six plates containing a duplicate set of three plates were prepared as confirmatory assay for the Lipid Droplet screen. The plates were scanned and analyzed for their intended purpose in the Lipid Droplet screen.

Next, a *GuidePlate* of TUNEL was prepared with the same protocol used to make the confirmatory screen plates, but this time the plate was also stained for TUNEL. A TUNEL-mimic CI was found from this plate. Corresponding CI were also calculated retrospectively for the six screen plates. The ability of CI to find the same toxic compound twice was then assessed by surveying the CI-positive cells across the duplicate plates.

Furthermore, the PubChem compound database was queried for the compounds found to be CI-positive to find whether their toxicity is independently confirmed. PubChem contains screening results from various researchers and the data is organized both in terms of biological assays used and compounds tested. Therefore, it is possible to search for a compound and find its activity in various biological assays. Some of the assays deposited into PubChem are assays designed to investigate toxicity.

4.3 Results

4.3.1 Cell Permeability

Cells treated with a dose response of MG132, Staurosporine and Taxol were stained with the Cell Permeability dye as described in the methods section. The Cell Permeability signal was quantified as the average pixel intensity of the corresponding channel under the cells mask calculated by Cytoshop. The CI was guided to mimic this signal as described in the methods section. The values (CI and Cell Permeability) were averaged for the well and averaged again for all wells with the same condition (i.e. for each concentration). The dose response results for Cell Permeability dye and the mimicked CI are shown in Figure 4.5. A Zprime was also calculated and reported. The Zprime reported was the better Zprime between the non-treated condition and the two highest concentrations. The Zprime was improved from negative values of -0.89, -1.08 and -29.37 to +0.30, +0.64 and -2.25 for MG132, Staurosporine and Taxol, respectively. Usually, a Zprime greater than +0.5 warrants running a screen in HT-HCS without duplicates while positive values lower than +0.5 would require duplicates. With the improvements afforded by replacing the Cell Permeability signal with CI, the screen is now statistically viable and cytotoxic hits are better characterized and controlled. All this occurs while the screen can be run without the additional cost of the cell Permeability dye.

4.3.2 Cytochrome C

Once again, a dose response of MG132, Staurosporine and Taxol were used to make a mimicked CI measurement. This time the cells were stained for Cytochrome C as described in the methods section. The Cytochrome C signal was quantified as the average pixel intensity of the corresponding channel under the nuclear mask calculated by Cytoshop. The CI was guided to mimic this signal as described in the methods section. The values (CI and Cytochrome C) were averaged for the well and averaged again for all wells with the same condition (i.e. for each concentration). The dose response results for Cytochrome C and the

mimicked CI are shown in Figure 4.6. A Zprime was also calculated and reported. The Zprime reported was the better Zprime between the non-treated condition and the two highest concentrations. The Zprime was improved from negative values of -4.13, -1.30 and -3.42 to +0.30, +0.65 and -2.06 for MG132, Staurosporine and Taxol, respectively. Once again, the improvements afforded by replacing the Cytochrome C signal with CI, the screen becomes statistically viable and cytotoxic hits are better characterized and controlled. This improvement is made even though the need for additional staining, scanning and analysis of the Cytochrome C may be eliminated.

4.3.3 TUNEL

For mimicking TUNEL, like the other mimics, dose responses of MG132, Staurosporine and Taxol were used. TUNEL staining was done as described in the methods section. The average pixel intensity of the TUNEL channel under the nuclear mask was calculated by Cytoshop to quantify DNA fragmentation. The CI was guided to mimic this signal as described in the methods. The values (CI and TUNEL) were averaged for the well and averaged again for all wells with the same condition (i.e. for each concentration). The dose response results for TUNEL and the mimicked CI are shown in Figure 4.7. A Zprime was also calculated and reported. The Zprime reported was the better Zprime between the non-treated condition and the two highest concentrations. The Zprime was improved from -15.13 and -9.51 for TUNEL to -5.57 and -2.83 for CI for Staurosporine and Taxol, respectively. The Zprime for MG132 was slightly worsened from -1.41 to -1.97 when switching from TUNEL to CI. This worsening is not too significant when considering that the CI measure is calculated without the need for TUNEL staining, scanning and analysis.

4.3.4 Uniqueness of CI and *Attribute* Set Selected for Each Marker

It is possible that the CI only looks for *Attributes* that accompany a common cellular behavior among all markers chosen. In other words, the CI may be quantifying a profound and easily detectible effect that is nonspecific to the marker, but accompanies all kinds of cell death. For example, if one uses cell count as a measure of cell death, one finds that it highly correlates with end-stage toxicity. Yet, that doesn't make it a good candidate to replace all markers since it only represents one cell behavior and cannot be used to reproduce signals from various markers. To investigate this we aimed to compare the *Attributes* sets selected for each market by the algorithm. If the algorithm selects a different set of *Attributes* to mimic different markers, then CI is successful in dynamically choosing *Attributes* that best represent the marker's quantification of specific cell behavior. In particular, studying the correlation coefficients found for the *Attributes* and the $Target$ measure will be able to reveal if the correlation patterns found are unique to each marker. For example, if the correlation coefficients found for TUNEL are identical or very close to those found for Cytochrome C, it could be deduced that the CI merely quantifies a nuclear morphology that exists in both when early apoptosis is initiated (Cytochrome C) and when late-stage DNA fragmentation occurs (TUNEL). On the contrary, if the correlation coefficients found are varied and unique to each marker, then the CI is likely quantifying different patterns of morphology that are unique to the cell behavior being quantified by the marker to be mimicked.

The correlation coefficients found for between each of the *Attributes* and the TUNEL, Cytochrome C and Cell Permeability markers are shown in Figure 4.8. As described before for the calculation of CI, two sets of correlations are found. Pearson correlation coefficients are found to quantify linear correlation, while Spearman correlation coefficients are used to quantify non-linear correlations. The values found are color mapped for better visual presentation. The range 0 to +1 is shown in shades of black to red while the range 0 to -1 is illustrated in the shades of black to red.

Comparing the patterns that arise for Pearson and Spearman coefficients, it is obvious that on average Spearman coefficients are higher in magnitude (brighter green or red) than Pearson coefficients found. This is evidence for the trivial fact that in comparing any random set of vectors, one is more likely to find non-linear correlations than linear correlations. Linear correlations require a much higher agreement between the rank and magnitude of numbers in the two vectors compared (i.e., first order relationship) while Spearman non-linear correlations can be found for vectors that only match in rank order (i.e., first, second or higher order relationships).

Furthermore, while there are similarities between all, the patterns found for TUNEL, Cytochrome C and Cell Permeability appear to be distinct when viewed in detail. The patterns found for TUNEL are much different from those found for Cytochrome C and Cell Permeability. This may be due to the fact that the Cytochrome C and Cell Permeability data originates from a singe physical microtiter plate while a separate plate was prepared to obtain the TUNEL data. The patterns for Cytochrome C and Cell Permeability are similar in the features and PCA regions while their ICA potion appears to be distinct. This also suggest the appropriate use of ICA in conjunction to PCA for obtaining complementary components that quantify various aspects of the data spread. While PCA is keen to calculate components with greatest overall distribution (i.e., most information content)[51], ICA aims at calculating components that are independent (i.e., most independent sources of information) [25]. Other unintended variations (e.g., temperature, drying, etc.) may affect both Cytochrome C and Cell Permeability dye signal the same way since these signal originate from the same physical well. Therefore, the dominant source of information may be due to the unintended effects. However, in that case the information that is independent to each signal, sill still be extracted by ICA. This appears to be happening for the Cytochrome C and the Cell Permeability dye as they have almost identical correlation coefficients with the features and PCA components, but very distinct coefficients for ICA components.

4.3.5 Small Molecule Library Screen Results

The TUNEL-mimicked CI was found for all six plates (three set of duplicates). An example of the CI found for a set of duplicate plates is shown in Figure 4.9. Some of the compounds that have high CI in both duplicate plates are indicated by arrows. To quantify the repeatability of CI, the z-score of CI was found with respect to the untreated wells and those compounds with z-score > 3 were labeled as CI-positive. Next, the duplicate plates were surveyed and compounds found to be CI-positive in duplicate plates were counted. Out of the 869 compounds tested, 166 were CI-positive in at least one of the duplicate plates. Among the CI-positive compounds, 146 or 88% were repeated across duplicate plates. This data indicates that the CI method can robustly detect the same hit twice with 88% probability.

Finally, to investigate whether the toxicity of the CI-positive compounds are independently confirmed, PubChem compounds database was searched for all 166 compounds. Although results were not found for all of the 166 compounds searched, 28 compounds were found to be positive in other screens that study cytotoxicity. The most common of the toxicity compounds where hits were found were high-throughput plate reader screens that quantify Cytochrome C (apoptosis marker) and ATPlite (amount of cellular ATP). The fact that no results were obtained by querying the rest of the compounds does not necessarily mean that they are not toxic. It may also mean that they have never been tested. Further, if they were tested using plate readers (as most screens in PubChem are) there is a chance that their toxicity was missed. This is due to the fact that in high-throughput screens, plate reader screens are inherently less sensitive than HCS in finding low intensity hits [18].

4.4 Discussion

Results from mimicking Cell Permeability Dye (Figure 4.5),Cytochrome C (Figure 4.6) and TUNEL (Figure 4.7) illustrate that the calculated CI closely resembles the dose response trend of the respective toxicity marker. When comparing

the magnitudes in the CI axis, it becomes apparent the CI value itself cannot be directly used as the cytotoxicity measure since it varies greatly from mimic to mimic and dose response to dose response. The CI measure is a highly relative measure and the fact that it offers viable Zprime indicates that it can resolve the biological behavior with great dynamic range. It is only when the CI value of a cell is compared to that of an untreated populations (i.e., in terms of z-scores) that the measure represents its biological meaning well. Therefore, it is recommended the z-score of CI be used as hit selection criteria instead of the CI value itself.

An added benefit of this method is that it can be retrospectively applied to images stored from previously run high content screens. Since most compounds in a large chemical library do not have toxicity information associated with them, this proves to be an inexpensive for assessing toxicity on images that are already collected and stored. The only effort needed is the preparation and analysis of a single *GuidePlate* with the desired marker to be mimicked.

In this study, some compounds found to be toxic by CI were previously known to be toxic, while no toxicity information is available for other detected compounds. Although requiring further biological experiments to confirm their toxicity, the fact that the findings of the CI potentially included previously unknown toxic compounds poses as an example of the motivation behind this project. Creating a database of toxicity information for various chemical libraries and different cell lines would not only add to our understanding of cytotoxicity in various cells, it could also save us the great cost of testing known toxic compounds over and over. The robustness and ease of use of CI, and the ability to apply it retrospectively, makes it a great candidate for such a role.

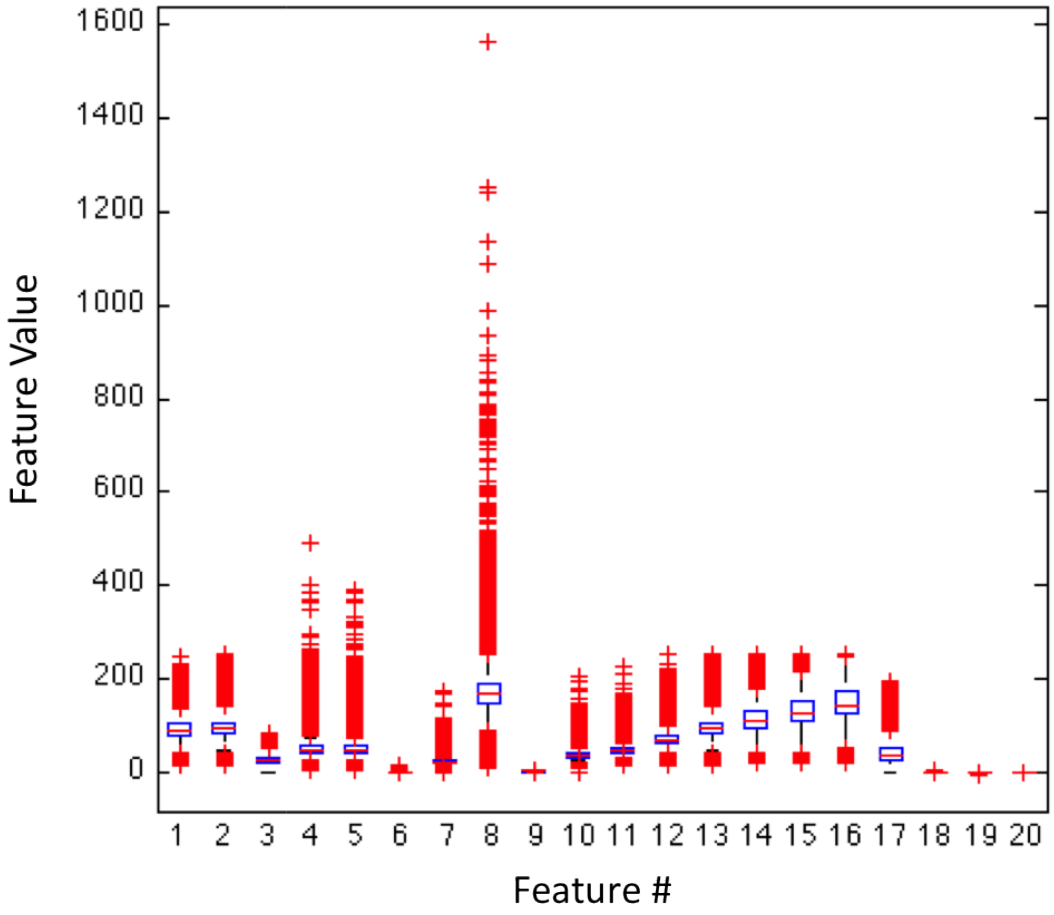


Figure 4.3: Variability of cytometric feature magnitude and range is illustrated in a box plot. A sample set of nuclear features were chosen and plotted to to illustrate the variability. In the plot, the central red mark is the median, edges of the blue box are the 25th and 75th percentiles, the black whiskers are the limits at which extreme data points not considered outliers and outliers are plotted individually as red crosses.

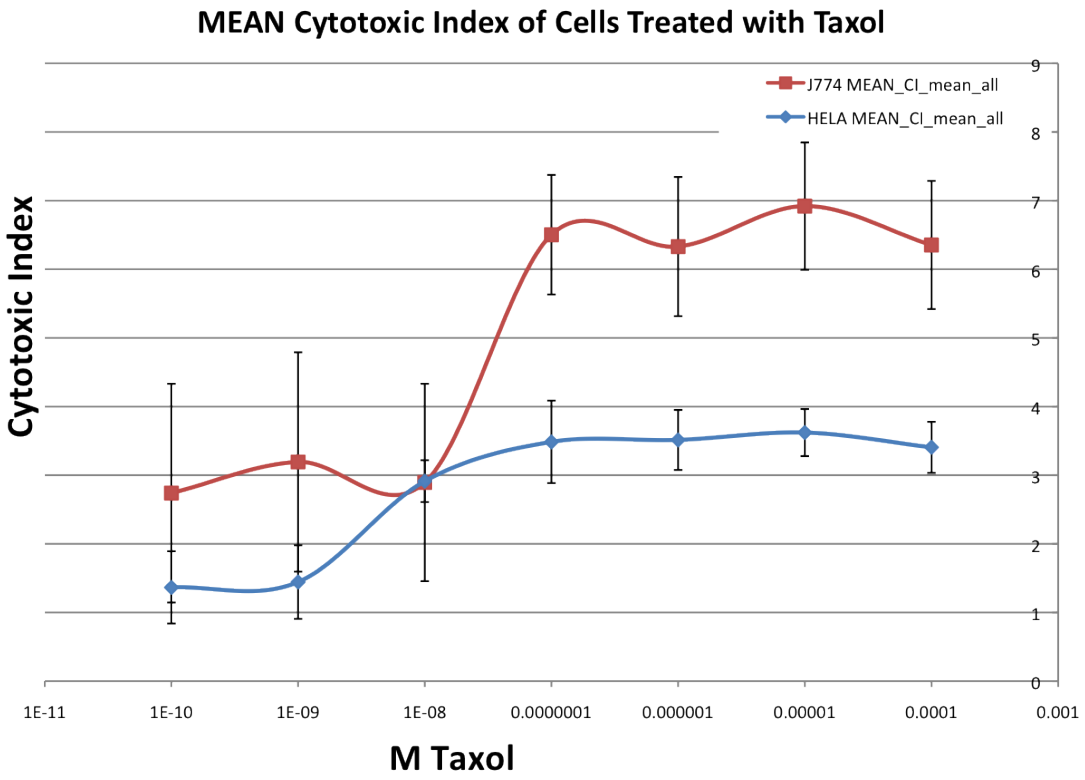


Figure 4.4: Dose response of Taxol for J774 and HeLa cells as quantified by the *a priori* feature set CI.

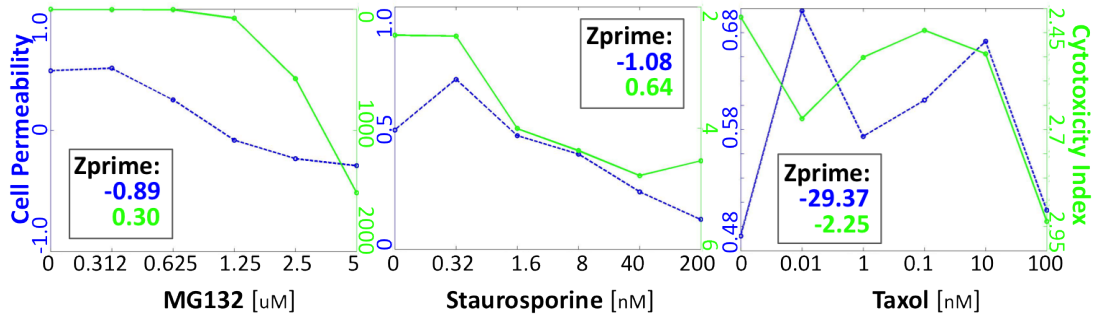


Figure 4.5: Dose responses of Cell Permeability and its mimicked CI. Dose responses of MG132, Staurosporine and Taxol are shown from left to right for the average cellular intensity of the Cell Permeability dye (blue) and CI (green). The color-coded Zprimes are also shown in the insets.

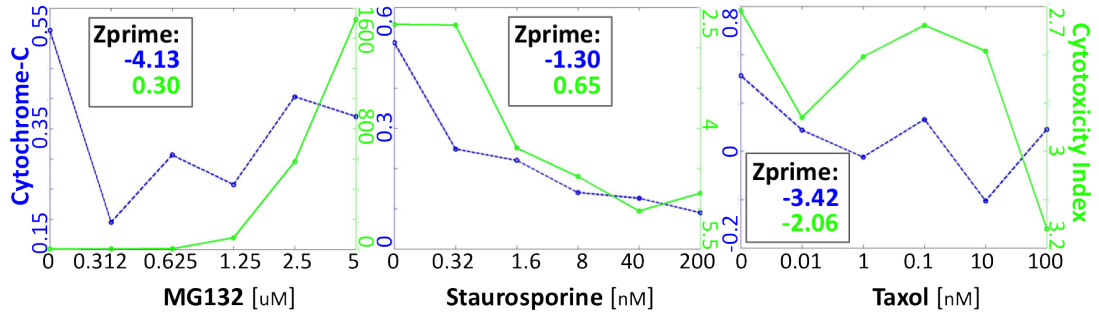


Figure 4.6: Dose responses of Cytochrome C and its mimicked CI. Dose responses of MG132, Staurosporine and Taxol are shown from left to right for the average nuclear intensity of Cytochrome C (blue) and CI (green). The color-coded Zprimes are also shown in the insets.

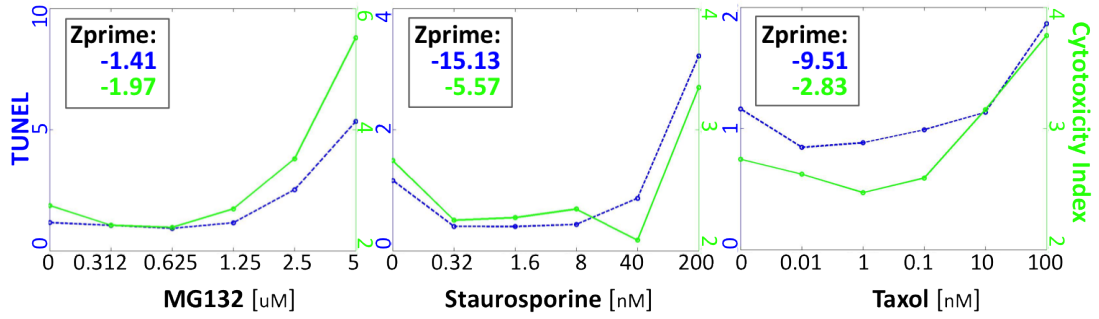


Figure 4.7: Dose responses of TUNEL and its mimicked CI. Dose responses of MG132, Staurosporine and Taxol are shown from left to right for the average nuclear intensity of TUNEL (green) and CI (blue). The color-coded Zprimes are also shown in the insets.

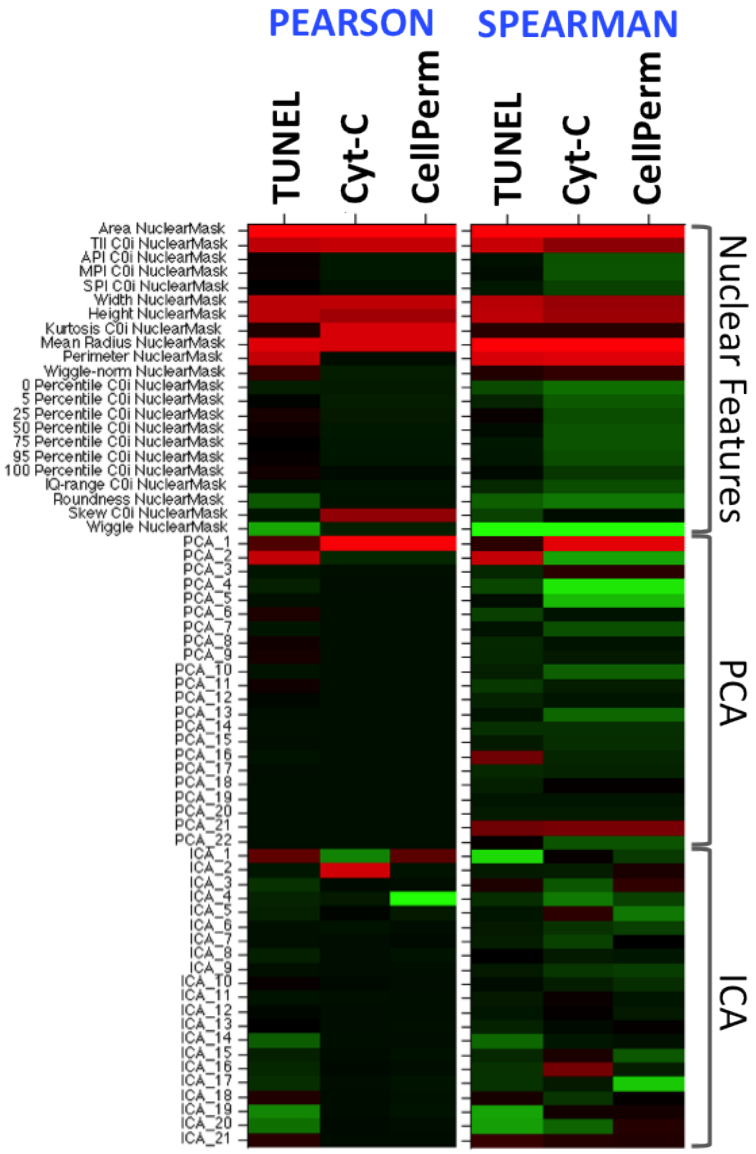


Figure 4.8: Correlation coefficients of *Attributes* found for various toxicity markers. Pearson and Spearman coefficients were found for 22 Nuclear Features and their PCA and ICA transformations with respect to TUNEL, Cytochrome C and Cell Permeability measures. Green=1, Black=0 and Red=-1.

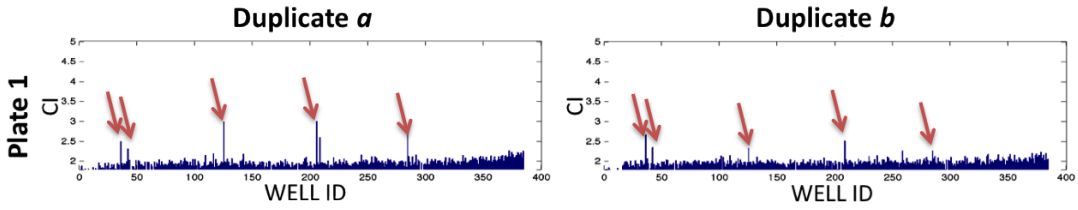


Figure 4.9: Example CI results form duplicate plates of a small molecule library screen. The CI is plotted as a bar for each well. Duplicate plates shown have the same compound map and are prepared identically. Note the repeated spikes across duplicate plates (some indicated by arrows) illustrates a high CI value for the same compound in both plates.

Chapter 5

Hybrid Median Filter

5.1 Introduction

Errors in MTP data can arise from many sources including robotic liquid handling, instrumentation and atmospheric conditions, and are frequently exacerbated by lengthy or complex assays [50, 56, 90, 114]. The detrimental localized data distortions in MTP data obtained from a chemical library screen can be random, such that sporadic errors are distributed unpredictably throughout a screen, or systematic, such that a similar pattern is repeated predictably. These distortions include edge artifacts that can be caused, e.g., by variations in temperature and humidity over the area of a MTP. Even seemingly predictable patterns such as edge artifacts usually have random components that cannot be easily modeled or corrected [90]. Random components become more evident when comparing spatially patterned data distortions from MTP to MTP, as they often defy prediction by a concise error model. Strategies for rectifying systematic errors include defining and tracking the error sources and introducing methods to compensate for their effects. A common approach is to flag and remove errors from MTP data, but this has the undesirable effect of discarding potentially useful data [50]. The primary challenge, therefore, is to design a method for correcting underlying data distortions while preserving the signal contributed by the hits, which can be modeled as sparse point noise, or outliers.

Local distortions, such as edge artifacts, can be thought of as local variations

in background. This shifts the problem to one of finding a way to correct the background values to within a narrow and uniform range (e.g., a flat surface in a 3D plot), thereby removing the distortions/artifacts and improving the confidence in resolving the sparse point noise of outliers that correspond to screening hits, while reducing the hit/background dynamic range as little as possible. We believe the simplest way to do this is to estimate the local background (L) within an appropriate sample neighborhood and scale each target data point (TargetValue) centered in that neighborhood by the ratio of the global (G) and local backgrounds (L) as shown in Equation (5.1), where the CorrectedValue replaces the TargetValue in the corrected MTP dataset.

$$CorrectedValue = \left(\frac{G}{L}\right) \times TargetValue \tag{5.1}$$

In chemical library or functional genomics (RNAi and cDNA) screens, hits are typically high- or low-magnitude values compared to the background wells. Thus, error correction consists of correcting local background distortions by, e.g., applying a median filter,[65] as illustrated in Figure 5.1 panel A, while retaining sporadic, high/low magnitude events. That is, the hits are analogous to the salt and pepper, or point noise that typically comprises errors in many other applications. For chemical library or functional genomics screening applications, non-parametric data scaling should not remove or dwarf these rare, low- or high-magnitude events, since these correspond to assay hits and are the samples of interest. MTPs with local patterns designed into the assay, such as a serial dilutions of compounds for dose-responses, should be addressed with other methods such as LOcal regrESSion (LOESS) to fit and normalize data[98, 114] via locally adaptive processes that use low-order polynomials to approximate a fit to neighborhood data [24].

For primary screens, both the median filter and the hybrid median filter (HMF) can be used to estimate the local (empirically defined neighborhood) background to scale the center element by the global background (G/L, see Equation (5.1)). G is the true or expected background, which can be estimated by the median or average of the whole dataset or a representative subset of values (e.g.,

multiple plates in a batch, or one plate) in the screen; the global background estimate is a constant for a whole dataset, a batch of plates, or a single plate respectively. Importantly, the median-based correction is outlier resistant in that a single hit/outlier will not alter the local estimate L of the background. By taking the statistical median step-wise over pre-determined sub-regions of the neighborhood, the HMF provides additional freedom to ignore multiple hits/outliers within the neighborhood to create the background estimate. Because they completely ignore the hits/outliers in estimating the background, we hypothesized that median filters would be a better choice than typical smoothing operators/filters and those based on the DFT, which invariably blunt the hits. The linear DFT views everything as spatial frequencies, which cannot always properly represent the discrete properties inherent of the assay data. The nonlinear and inherently digital properties of the HMF enable more complete insensitivity to hits/outliers and it can be applied to raw data arrays systematically and automatically without plate-by-plate operator input, which may improve consistency.

In preliminary experiments, we tested various local background estimators on MTP data for their respective capacity to diminish spatial distortion while preserving hit amplitudes, and found the bidirectional hybrid median filter (HMF) promising [77, 91]. We then adapted the HMF for better operation at MTP edges and compared it to a recently described Small Laboratory Information Management System (SLIMS) DFT technique [56, 90]. The strengths and weaknesses of the algorithms were evaluated by the reduction of background distortion and overall variations in MTP values (with hits removed), preservation of hit amplitudes, and/or hit:background dynamic range.

5.2 Materials and Methods

5.2.1 Median-Based Array Correction

MTP data can be written in the matrix notation Equation (5.2) where R and C represent the maximum number of rows and columns, respectively , in the MTP (e.g., R=16, C=24 for a 384-well MTP). Here, each element of the matrix

represents the measurement originating from a well. In index notation, this can be simply expressed as $MTP_{i,j}$, where $i = [1, 2, 3, ..., R]$ and $j = [1, 2, 3, ..., C]$ are the row and column indices.

$$MTP = \begin{pmatrix} MTP_{1,1} & MTP_{1,2} & \dots & MTP_{1,C} \\ MTP_{2,1} & MTP_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ MTP_{R,1} & \dots & \dots & MTP_{R,C} \end{pmatrix} \tag{5.2}$$

A square local neighborhood of the size $d \times d$ (e.g., 3×3 and 5×5 in Figure 5.1 panel A and B, respectively) is explicitly described with odd values of d for symmetry. The half-size of the filter, using integer math, is defined as h=d/2 (e.g., h=5/2=2 for d=5). A local background $L_{i,j}$ within each $d \times d$ neighborhood around $MTP_{i,j}$ is estimated by the HMF such that the center element of the neighborhood falls on $MTP_{i,j}$. The medians of the diagonal and axial elements of each neighborhood are defined as shown in Equation (5.3).

$$\begin{aligned} &AxialElementsMedian = \\ &MEDIAN\langle H \cup V \mid H = MTP_{i,j\neq k}, V = MTP_{i\neq k,j}; k = [1, 2, ..., h]\rangle \end{aligned}$$

$$\begin{aligned} &DiagonalElementsMedian = \\ &MEDIAN\langle D \mid D = MTP_{i\neq k,j\neq k}; k = [1, 2, ..., h]\rangle \end{aligned} \tag{5.3}$$

Note that since k is a non-zero index, the central element is excluded from these median calculations. Moreover, when the neighborhood is close to the periphery of the MTP, the filter elements that fall outside of the defined range of MTPi,j are simply ignored by dynamically shrinking the neighborhood size at the edges. Finally, the estimate of the local background is as shown in Equation (5.4).

$$L_{i,j} = MEDIAN[AxialElementsMedian, DiagonalElementsMedian, MTP_{i,j}] \tag{5.4}$$

Using the global background G of the entire MTP and this local background estimate, each data point MTPi,j is then scaled to a corrected value Ci,j using a more specific version of Equation (5.1), given in Equation (5.5).

$$C_{i,j} = \left(\frac{G}{L_{i,j}}\right) \times MTP_{i,j} \tag{5.5}$$

For the more common median and average filters, each element was also scaled in the manner described in Equations (5.1) and (5.4), with L defined as the median and average of the neighborhoods (Figure 5.1 panel A), respectively, and G defined as the MTP median (Figure 5.1).

The kernel sizes of 3×3 and 5×5 for the HMF were chosen empirically. In preliminary experiments (data not shown), the 5×5 size appeared to be optimum for 384-well MTP arrays; however, application to higher density formats may require additional testing to confirm an optimal kernel size and it is possible that larger kernel sizes (e.g. 7×7, 9×9, etc) might perform better in higher density formats where, e.g., edge artifacts might also extend inward over a larger number of wells. Since the local background estimation depends heavily on obtaining a representative sample population from the MTP, kernel size is important particularly at the MTP edges. An alternative to ignoring the kernel elements outside the array when operating at the edges is to keep the kernel size constant and move the target pixel within the kernel and appropriately alter the sub-regions for calculations of the intermediate medians. Alternatives from image processing are also available [58, 59, 97].

Application of the common median and average filters (see Figure 5.1 panel A), and the bidirectional HMF (see Figure 5.1 panel B), are described as follows. For the example 5×5 neighborhood in Figure 5.1 panel A, the median and average filters used all 25 of the elements to obtain local backgrounds of 12 and 12.96,

respectively. Application of the bidirectional HMF in Figure 5.1 panel B to the same neighborhood gave a median of the diagonal elements of 11 and a median of the axial elements of 13. With a central data element of 19, the final rank ordered list yields a hybrid median of 13.

The operation of the HMF is illustrated in more detail in the diagram of an example MTP (Figure 5.2) created using Microsoft Excel by weighting array values appropriately to create a uniform contour with a pronounced edge artifact. The HMF with kernel (or neighborhood) size d=5, was applied to the MTP and the medians of the diagonal (gray) and axial (black) elements are noted in the table (Figure 5.2, bottom), along with the original and scaled value of the center MTPi,j position (red). Note that elements of the HMF outside of the MTP array are ignored when they overhang the plate edge at the peripheral wells. The total number of contributing elements (N) is listed for each of the masks, which are located at center coordinates IDi,j.

For the example targeting well A-01 for correction in Figure 5.2, there is no hit, but the modeled edge effect decreases the values toward the edge and Eq. 5 scales MTP1,1 = 25 to the corrected value C1,1 = 50. At well D-16, MTP4,16 is a down hit, there are two up hits in the 5×5 neighborhood (95 in the diagonal and 75 on the vertical axis), and the well value is unchanged at C4,16 = 10. At H-24, there is an up hit distorted by the edge effect and MTP8,24 = 95 is scaled up to C8,24 = 113. At well K-10, MTP11,10 = 50 is the same as the global estimate, resulting in no change for C11,10 = 50. Well O-03 is a down hit of MTP15,3 = 4 that is distorted by the edge effect and is corrected up to C15,3 = 7. A program for implementing the bidirectional hybrid median filter can be downloaded at http://sdccg.burnham.org/metadot/index.pl?iid=2363 (password pbushway).

5.2.2 Discrete Fourier Transform-Based Data Correction

DFT calculations of biological array experiments,[56, 90] can be used to create a periodogram of the array data [5, 105]. The periodogram describes the degree of distortions that can be modeled by patterning in the original array data. In a MTP, plate-wide, well-by-well random variations including point-noise (out-

liers; hits) represents low periodogram amplitudes and composes the bulk of the array data in a screen. Localized spatial distortion and patterning appears as high magnitude signal on a periodogram due to spatial agreement in array values.

Our MTP data were processed with the Small Laboratory Information Management System (SLIMS) DFT software (downloaded from
http://slims.sourceforge.net) for comparison with the hybrid median filter. The VisTa standalone was used to confirm SLIMS-implemented DFT and DFT corrections
(see http://www.columbia.edu/cu/biology/faculty/stockwell/StockwellLab/index to download). [90]

5.2.3 MIN6 eGFP/DsRED cells

The MIN6 mouse insulinoma cell line [74] was stably modified following lentiviral infection to express eGFP under the control of the human insulin promoter [81] and DsRED under control of the phosphoglycerate-kinase (PGK) minimal promoter. [27, 40] The dual promoter system allows a direct comparison of insulin and PGK promoter activities.

5.2.4 Experimental Array Data

Our comparative analysis of correction methods used a combination of three computationally-derived (synthetic, created in Microsoft Excel) and four experimentally-derived, cell-based (experimental, MIN6 eGFP/DsRED cells) MTP array data, which are summarized in Table 5.1 . In the three synthetic arrays, a model 384-well MTP was created to mimic the edge artifact commonly observed in real-world assay data. Synthetic1 (Figure 5.3) edge distortion was created in the outer wells (a distance of three or less from the edge), spanning the range 0.25-0.5 (out of 0.0-1.0). The Synthetic1 background values were generated randomly around a mean of 0.5 with a 5% deviation (using a uniform distribution) to mimic noise. Synthetic1 hits were inserted randomly with magnitudes 0.6 and 1.0 (Figure 5.3 panel A). Synthetic2 was created with edge effects within three wells

of the edge but without background noise or hits (Figure 5.4 panel A). Sythethic3 is similar to Synthetic2 but with hits created by a random multiplier or divisor in the range 1 to 10 (Figure 5.6 panel A).

The four experimental datasets were created from two different MIN6 MTPs that were prepared to mimic minor edge distortion and hits (Experimental1 in Figure 5.4 panel D, Experimental2 in Figure 5.7 and Figure 5.5) MIN6 cells were seeded into a black-wall, clear-bottom 384 well MTP to achieve 60% confluence 24 hrs after seeding. Four image fields per well were acquired for both MTPs using a Nikon 10x, NA 0.45 objective at 12-bits/pixel binned 4x4 using the following filters: 360/40, 460/40 (DAPI), 475/20, 535/50 (eGFP), 535/50, 620/60 (DsRED) excitation/bandpass, emission/bandpass, respectively. TIFF images were analyzed using the IN Cell Analyzer 1000 Developer Toolbox. First, the eGFP image channelwas flat-field corrected from previously acquired blank reference images and then an isotropic diffusion filter was passed over the image for 5 iterations to smooth image noise in image areas devoid of significant signal variation (terms and algorithms from INCell 1000 Developer Toolbox). Images were then segmented using threshold $T = 1.04 * Mean_Image_Intensity$ for the eGFP channel. The well readouts were the density multiplied by the area under the segmented masks produced as a standard metric (DxA measurement) in the GE Developer Toolbox. The first plate was incubated for a total of 3 days at 37 oC and 10% CO2 until cells achieved 90-100% confluence, then fixed in 5% paraformaldehyde (FLUKA) and imaged on the INCell 1000 to generate the Experimental1 dataset. Then randomly positioned wells were incubated with 0.04% Phalloidin-Alexa488 (Invitrogen/Molecular Probes) in PBS to simulate up hits and the MTP was re-imaged to generate the Experimental2 dataset.

The second MIN6 MTP for datasets Experimental3 and Experimental4 was created to mimic more severe edge distortion and hits with decreased fluorescence intensity. Cells were seeded to achieve 40% confluence the day after seeding and maintained in culture for an extended incubation time of 5 days to maximize edge distortions, fixed and read using the EnVision plate-reader (PerkinElmer, MA, USA). Methanol was then added to randomly selected wells and groups of wells,

Table 5.1: List of datasets used in evaluation of HMF. The datasets listed are created and used for gauging the performance of the correction methods.

Dataset Name	Description
Synthetic1	Synthetic data created in Microsoft Excel with a background SD of 5%, containing hits less than or equal to 5x SD. (Figure 5.3 panel A)
Synthetic2	Synthetic data created in Microsoft Excel to represent edge effects in the three most peripheral wells. (Figure 5.4 panel A)
Synthetic3	Synthetic data created in Microsoft Excel to represent edge effects in the most peripheral wells (within three from the edge), with hits created by a random multiplier or divisor in the range 1 to 10. (Figure 5.6 panel A)
Experimental1	eGFP Readout of MIN6 cells incubated for 2 days and imaged on an INCell1000 imager (Figure 5.4 panel D) with no hits implemented.
Experimental2	eGFP Readout of MIN6 cells incubated for 2 days and imaged on an INCell1000 imager. Up hits were simulated using phalloidin-FITC for illustration in (Figure 5.7 panel) and analysis presented in (Figure 5.5 panel).
Experimental3	Data are eGFP or DsRED (two colors, same MTP) readout of MIN6 cells on an Envision plate-reader after incubation for 5 days.
Experimental4	eGFP or DsRED readout of MIN6 cells on an Envision plate-reader (Figure 5.4 panels D and G respectively) after incubation for 5 days. Down hits were simulated by addition of methanol as described in Methods.

to decrease fluorescence and thereby mimic down hits or toxic compounds for the Experimental4 dataset, which was also read on the Envision instrument. Both eGFP and DsRed spectra data were collected. This diverse collection of datasets helped delineate the differences in performance.

5.2.5 Calculations

The measurements used to gauge the performance of arrays are described as follows. The coefficient of variation of the background (CV = 100xSD/mean) was used to describe the variation in the MTP data array background values. In arrays with hits, the hits were removed prior to calculating the CV, unless otherwise noted.

Over each MTP dataset, we define the average hit amplitude and Dynamic Range as shown in Equations (5.6) and (5.7), respectively, where $SDBackground$ is the SD of the plate with the hits excluded. Dynamic Range and CV both distinguish array corrections that improve assay performance by decreasing background variation and/or increasing hit amplitudes, rather than by a simple multiplication that increases the amplitudes of both the hits and the background. The data is objectively improved hit magnitudes increased and/or background variations decreased with smaller background CV and larger Dynamic Range.

$$AverageHitAmplitude = \frac{\sum \mid Hit - Mean_{Background} \mid}{NumberofHits} \tag{5.6}$$

$$DynamicRange = \frac{AverageHitAmplitude}{SD_{Background}} \tag{5.7}$$

5.2.6 HMF Performance vs. Hit Density

The HMF assumes that hits are relatively rare in the neighborhood. As hit density increases, the probability increases that enough hits reside within a given neighborhood to cause errors in the local background estimate. To investigate the sensitivity of the HMF to the number of hits in an MTP, we mathematically introduced increasing numbers of hits into the synthetic and experimental datasets

and measured the corresponding changes in the CV and Minimum Dynamic Range (which is defined by a variation of Equation (5.7) wherein the Average Hit Amplitude is replaced by the Minimum Hit Amplitude). For this experiment, variations of the Synthetic2 (Figure 5.4 panel A) and Experimental1 (Figure 5.4 panel D) MTPs were used. Originally, neither of these datasets contained any hits. To mimic a relatively weak hit, we chose a hit signal magnitude of mean plus 3 x SD, where SD is the Standard Deviation of the MTP without hits. Hits were randomly positioned in the MTP and Dynamic Range and CV were measured with and without correction by the 5×5 HMF. This experiment was performed with number of hits varying from 1 to 381 in increments of 5.

5.3 Results

5.3.1 Comparisons of Average, Median and Bidirectional HMF Corrections on a Synthetic MTP

In the first experiment, the background estimates provided by HMF were directly compared to those from the median and average filters (see Figure 5.1) for their ability to correct localized distortions in the Synthetic1 dataset which contains simulated hits (Figure 5.3). All filters improved the Dynamic Range in Synthetic1 to some degree, but the HMF achieved the greatest reduction in CVs (Figure 5.3 panel E). The 5×5 HMF (Figure 5.3 panel B) and the 3×3 median filters (Figure 5.3 panel C) corrected the edge distortion and retained the amplitude of simulated hits. In contrast, the smoothing effects of the 3×3 average filter (Figure 5.3 panel D) reduced simulated hit magnitudes and demonstrated a diminished capacity to correct edge distortions (Figure 5.3 panels B-C).

Background smoothing was compared via the CVs as shown in Figure 5.3 panel E; a lower MTP background CV indicates a more uniform background. The 3×3 median filter (Figure 5.3 panel E) reduced the whole background CV to 8.2% from 24.9% for raw MTP and increased the Dynamic Range (see Eq. 7) by 7.1-fold, while the 5×5 median filter decreased the whole background CV to 13.8%

and increased Dynamic Range by 3.1-fold. The 3×3 average filter resulted in less improvement, with a whole background CV of 11.2% and Dynamic Range 3.3-fold better. The 5×5 average filter decreased the whole background CV to 16.7% and increased the Dynamic Range by 2.1-fold. The 3×3 median filter increased the dynamic range the most (7.1-fold), with the HMF next best (6.5-fold), whereas the HMF reduced the whole background CV the most (to 7.9%) and the 3×3 median next best (to 8.2%). At the edges, the HMF also reduced the CV the most (from 14.2% to 8.8%), with the 3×3 median filter next with a reduction to 9.7%. Inspection of the corrections along the 2-well-wide edges of the MTP in Figure 5.3 panel C vs. A shows the differences. The 3×3 median filter, results in Figure 5.3 panel C, demonstrate more corruption in at the edges (e.g., regions A-15, A-09, D-02 and O-23 in Figure 5.3 panel C vs. panel 3). Thus, the HMF enhanced Dynamic Range less than the 3×3 median filter, but created a lower background CV and left hit amplitudes closer to RAW data values. Also notable is how differently the HMF correction effected the CV of the peripheral wells compared to the central wells. HMF reduced the CV of the peripheral wells the most (14.2% in raw versus 8.8% in corrected) while changing the CV of the center region the least (6.6% in raw versus 6.8% in corrected). This shows the effectiveness of the HMF correction, especially for edge effects.

5.3.2 Performance comparison of the 5×5 HMF and DFT Corrections on Spatially Distorted MTP Array Data

We compared the performances of the 5×5 HMF and DFT (as applied in the SLIMS software package) on synthetic and experimental datasets, both with and without simulated hits as described in the Methods. Example images from the Experimental2 dataset are shown in Figure 5.7 and include: 1) A negative control central well (Figure 5.7 panel A); 2) a dim negative control edge well (Figure 5.7 panel B) that resulted from long-term cell culture; and 3) a very bright simulated hit positioned at the MTP edge (Figure 5.7 panel C, note higher contrast vs. panels A and B) that were artificially brightened to visualize the cells.

Figure 5.4 panels A and D are surface plots of the Synthetic2 and Exper-

imental1 datasets, respectively. These datasets did not contain simulated hits. Figure 5.4 panel B illustrates the Synthetic2 MTP treated with the SLIMS DFT method. The waves in the surface plots that are especially prominent at the edges in Figure 5.4 panel B are typical ringing artifacts often created by transforming the image data to the discrete frequency domain and back to the spatial domain,[108] as done by DFT techniques. By inspection, the HMF produced a flatter 3D plot overall (Figure 5.4 panel F) vs. the DFT correction (Figure 5.4 panel E, compare, e.g., areas M-J/13-16). Neither method completely corrected areas near corner wells A-1 and A-22 (Figure 5.4 panels E and F), where more prominent noise might be mistaken for hits.

The Synthetic2, Synthetic3, Experimental1, Experimental2 and both eGFP and DsRed datasets of Experimental3 and Experimental4 were corrected using the DFT, 3×3 median, 5×5 median and 5×5 HMF correction methods. The CVs were calculated and plotted in Figure 5.5 panel A for datasets with no simulated hits and 6B for data arrays with simulated hits. In data arrays without hits (Figure 5.5 panel A), the 5×5 median background estimator performed the worst (decreased the CVs the least) on the eGFP and DsRed Experimental3 and Synthetic2 datasets, and the DFT performed the worst on the eGFP and DsRed Experimental3 dataset. The 5×5 HMF lowered the CVs the most on all datasets; with DFT second best on the Experimental3 eGFP dataset (4.2% vs. 4.5%), and tied with HMF in the Experimental3 DsRed dataset (both 4.3%). The 3×3 median was second to HMF on both the Experimental1 (2.7% vs. 2.8%) and the Synthetic2 (2.6% vs. 4.1%) datasets. For analogous corrections of the arrays with simulated hits included (Figure 5.5 panel B), the DFT performed the worst on all datasets, except for Experimental4 DsRed, and actually increased the CVs compared to the raw data on the Experimental2 and Synthetic3 datasets (to 36.0% from 14.4% and to 33.1% vs. 23.8%, respectively). The 5×5 HMF again performed the best on all datasets, with the 3×3 median second best on the Experimental4 eGFP dataset (5.1% vs. 5.9%), the DFT second best on the Experimental4 DsRed dataset (6.5% vs. 8.8%), and the 3×3 median second best on both the Experimental2 (4.5% vs. 4.9%) and Synthetic3 datasets (3.9% vs. 5.7%).

In Table 5.2 , the data of Figure 5.5 panels A and B were further averaged across all datasets into two groups with and without simulated hits for comparison. Without simulated hits, the improvements in CV to an average of 0.2-0.3x of raw are similar except for the 5×5 median, which improved the CVs by an average of only 0.6x. The differences are more dramatic with simulated hits, where the 5×5 HMF and 3×3 median average corrections were again similar at 0.3x and 0.4x, respectively, but the 5×5 median corrected by an average of 0.6x and the DFT actually made the average CV worse than the raw data by 1.3x. Thus, for the results in Figure 5.5 panels A-B and averages shown in Table 5.2 , the CV was decreased most consistently by the HMF method, and the DFT correction performed especially poorly on the Experimental2 and Synethetic3 datasets with simulated hits included.

Table 5.2: Performance of correction methods without hits vs. with hits averaged for all datasets. The average CVs for the datasets are tabulated for experiments without and with simulated hits in the arrays, to further compare the performances of the DFT, 3×3 median, 5×5 median, and 5×5 HMF filters in the experiments of Figure 5.5 panels A and B. CVs with hits were calculated only on non-hit wells.

	No Hits		**With Hits**	
Correction Method	**CV(%)**	**CV/Raw**	**CV(%)**	**CV/Raw**
Raw Data	16.2	1.0	17.0	1.0
5×5 HMF	3.5	0.2	5.0	0.3
3×3 median	5.4	0.3	7.3	0.4
SLIMS DFT	5.5	0.3	22.0	1.3
5×5 median	9.2	0.6	10.4	0.6

We next evaluated the Dynamic Ranges (see Equation (5.7)) and average hit amplitudes after corrections by the median, HMF and DFT (both normalized to the raw data) methods. As shown in Figure 5.5 panel C, the Dynamic Ranges increased most after correction by the HMF in all datasets, with the 3×3 median performing second best in the Experimental4 eGFP dataset (1.4 vs. 1.6), the DFT performing second best in the Experimental4 DsRed dataset (1.7 vs. 3.2), and the

3×3 median performing second best in both the Experimental2 (3.0 vs. 3.2) and Synthetic3 (3.6 vs. 5.2) datasets. The DFT method decreased the Dynamic Range in all datasets except in Experimental4 DsRed dataset.

Correction methods affect the range of simulated hit amplitudes, as shown in Figure 5.5 panel D where a consistent retention of hit amplitudes was observed in HMF corrections, while the DFT had a tendency to blunt the hits. Interestingly, while the DFT blunted the hit amplitudes the most in the Experimental4 DsRed dataset (0.69x), it apparently also decreased the CV even more to still improve the Dynamic Range, second only to the HMF (Figure 5.5 panel C). Overall the median-based background estimation method (Equation (5.1)), retained hit amplitudes representative of the raw (0.99x to 1.06x).The preservation of raw data hit amplitudes was independent of median filter type and contrasted sharply with DFT corrections which consistently reduced hit amplitudes over all of the datasets. In addition, all of the background estimation methods (Equation (5.1)) improved the Dynamic Ranges (Figure 5.5 panel C), whereas the DFT decreased them in three out of the four datasets.

The DFT and HMF corrective effects on simulated hit amplitudes are further illustrated in 3D surface plots of the Synthetic3 and Experimental4 eGFP and DsRed datasets in Figure 5.6 panes A-I. Note that artifacts in different channels in the same wells (e.g., eGFP and DsRED) are independent from each other (see Figure 5.6 panels D-F vs. G-I), which means that the background estimate of one reporter protein cannot be used to correct the distortion of another. The performance differences between these two methods can be viewed by comparing Figure 5.6 panels B, E and H to C, F and I, respectively. In each case, the DFT largely blunts the hits more and decreases the background variations less than the HMF background estimator.

5.3.3 Effect of Hit Density on HMF Correction Performance

The HMF background estimation method was found to better retain hits than the other background estimators. This ability to ignore hits in the background

estimate was expected to decrease as the hit densities increased. High hit density experiments, such as secondary screens, should be excluded from HMF correction, as should the regions with contiguous control wells. To determine at what hit densities HMF should no longer be used, simulated hits were randomly positioned 100 times each for hit densities of 1 to 382 in 384 wells with the magnitude of 3 x SD as described in the methods. The resulting CV (Figure 5.8 panel A) and SNR (Figure 5.8 panel B) plots were created from the averages of each set of 100 trials at each density and the SD envelopes are also shown. The HMF reduced the CVs at all hit densities and as shown by the SNR, began blunting the hits as the hit densities increased. The SNRs improved more than 50% at hit densities lower than 20% for the Synthetic1 data set and lower than 25% for the Experimental1 data set. Thus, the HMF decreased the CV and improved the SNR at hit densities an order of magnitude or greater than are typical for primary screens.

5.4 Discussion

One variable in the HMF that might be further optimized, especially for different (larger or smaller) MTP plate sizes is the kernel size. The 5×5 size was empirically determined to be optimum for 384-well MTP arrays; however, for application to denser formats (e.g., 1,536) it be would useful to test different kernel sizes such as 7×7 and 9×9. Since the local background estimation depends heavily on obtaining a representative sample population from the MTP, kernel size becomes a significant concern at the MTP edges. Although the total number of elements sampled by the kernel filter is reduced at the MTP periphery because a portion of the mask overhangs the MTP edge, the 5×5 HMF was robust enough to maintain a tight correlation between the obtained median value and the local sampling area in the 384 well MTP. Alternative ways to address array edges are well documented for image processing [58, 59, 97].

HMFs, like all automated methods for correcting systematic errors, require spatially random MTP data with no artificial arrangement of wells that are expected to have high or low magnitudes. As shown in Figure 5.8 panel , the effec-

tiveness of the filter diminishes as the hit density increases, with CV significantly diminished at hit rates above 20%. High densities or clustering of positive wells, or organizing MTPs with dose series or controls positioned in rows or columns would result in misrepresentation of the control values as background in the kernel and compromises filter function. One way to circumvent this problem is by nesting controls or dilution series within the MTP in such a way as to avoid having a low- or high-value cluster disturb the function of the filter (i.e. by randomizing well positions). Also, in the case of controls positioned in rows or columns, the filter can be programmed to exclude the control wells from analysis. The SLIMS interface used in the DFT corrections offers such an option [56, 90]. Alternatively, the HMF kernel could be customized by excluding axial elements (column or row) which might contain control wells and we are exploring this further.

In comparison of MTPs before and after correction with DFT, the correction restored values to the mean value range (compare Figure 5.6 panel G to H). With rare exceptions, however, the DFT corrections also reduced the amplitude of simulated hits (compare hits in Figure 5.6 panel E vs. D and panel H vs. G). For example, in the corners of the Experimental4 DsRed array, the DFT method failed to preserve simulated hits altogether and instead reduced the hit to background levels. This tendency to blunt hits is also common to the averaging correction methods (compare Figure 5.3 panels D, B and C). In contrast, the 5×5 HMF retained all simulated hits in our data sets, scaling them in agreement with local wells (e.g., compare region A1 in Figure 5.6 panels G-I).

The case of Experimental4 DsRed illustrates an interesting point in the comparison of the DFT and median-based correction methods. The background CV measurements for Experimental4 DsRed (Figure 5.5 panel B) are higher in arrays corrected with 3×3 and 5×5 median as compared to the DFT. However, viewing Figure 5.5 panel D reveals that while the DFT reduced the background CV more, it did so at the cost of blunting hits. The DFT correction is also accompanied with a reduction in Dynamic Range which is apparent in Figure 5.5 panel C. The lower CVs, blunted hits and reduced Dynamic Range after DFT treatment can be explained by a gross flattening of the array contour without regard to discrete

hits. For example, the MTP edge correction appears extensive after treatment with the DFT-based method in the Experimental DsRed (compare data in the range 37.5%-50% in Figure 5.6 panel G vs. H), but the corrections also reduced the magnitude of many of the hits (e.g., well A13).

Further, correction of corner wells by the DFT was more aggressive than the correction of the other edge-proximal wells. The DFT also blunted the hits in MTP corners more aggressively than other edge-proximal wells (e.g., compare corrections to A1 and P1 to M1 and A13 in Figure 5.6 panels G-H). This correction is unusual and suggests a DFT correction artifact based on highly conserved array symmetry (4 corners). The correction made by the HMF to MTP corner wells, however, appears to be consistent with that made to other edge-proximal wells (e.g., compare corrections to A1 and A13 in Figure 5.6 panels H-I).

Although the DFT reduced the background CVs statistically over the entire MTP for both Synthetic2 and experimental arrays (Figure 5.5 panel B and Table 5.2), it introduced waves or ringing artifacts in Figure 5.4 panels B and E (e.g. area J-17) in the corrected Synthetic2 dataset. The DFT is based on continuous functions that require special treatment to deal with finite data arrays. i.e., how does one model the region at the edge where the data ends? In order for continuous functions to work at the edges, assumptions have to be made to create data outside the original array. e.g., if one produces an artificially larger plate where the outside values are zero, a step function is produced that generates distinctive ripple patterns [108]. Various windowing functions have been designed to reduce this ripple. Primary screening data has hits that are fundamentally discrete on a background that can be modeled as continuous (a single mean with noise). Because of this it may be possible to remove the ripple by iteratively refining the estimate of the data outside the MTP array to match the mean and noise of the background.

Median filters (Figure 5.4 panel C), which are nonlinear and natively discrete, do not generate ringing artifacts. However, the HMF corrected array Figure 5.4 panel C exhibits symmetric artifacts in the corners of the plate due to the small sample size of the HMF kernel in the corners. This resulted in insufficient

sampling of the background and in turn reduced the efficacy of the correction. The corner correction failures in the HMF correction had a much smaller effect than the DFT ringing on the CVs (Figure 5.5 panel A). One possible solution to this problem is to adaptively increase the kernel area at the corners. The kernel size could also be held constant by moving the target pixel. Alternatively, we have adjusted the kernel pattern to make it more or less sensitive to outliers as the sample size decreases. In addition, we have found that the serial application of multiple discrete filters tuned to common MTP array patterns minimizes the introduction of artifacts at the MTP corner regions, while also having an additive beneficial effect on error correction.

The DFT method transforms MTP data into the Fourier space by fitting sinusoidal functions to the data [90]. Since sinusoidal functions are by definition continuous, this transformation assumes that the data is also continuous. For MTPs, this means that the DFT correction expects a hit, which by definition has an extreme magnitude, to resemble its surrounding background wells. However, since hits in MTP data are inherent spatial discontinuitiespresence of a hit is not related to the magnitude of the wells around it and is an abrupt departure from the backgroundthe piecewise continuity assumption of the DFT method is not appropriate. Fitting of a continuous function to discontinuous data reduces the hits toward background levels, i.e., blunts them. The HMF method, on the other hand, is based on nonlinear rank order calculations for each neighborhood (i.e., finding the median) and does not assume spatial continuity in the dataset. This inherently discrete method essentially ignores rare extreme values (hits) in its estimation of the background.

In summary, the 5×5 HMF performed best overall with regard to statistical improvement of the various datasets tested and this conclusion is supported by the surface plots describing the arrays before and after correction. The DFT method may benefit from case by case fine-tuning in the frequency domain. The bidirectional HMF might also be further tuned by optimizing the neighborhood size and subregions, but the easily implemented and computationally cheap 5×5 HMF performed well on all of the datasets tested here. We conclude that median

based array correction methods best reduced localized data distortion and assay noise while preserving hit amplitudes, and that discrete background smoothing approaches are superior to ones based on continuous functions for this data type rare hits in data arrays.

5.5 Acknowledgments

This chapter, in full, is a reprint of the material as it appears in Hybrid median filter background estimator for correcting distortions in microtiter plate data; Paul J. Bushway, Behrad Azimi, Susanne Heynen-Genel, Jeffrey H. Price, and Mark Mercola; ASSAY and Drug Development Technologies, 8(2), 238–250, April 2010 [17]. The dissertation author was one of the primary co-investigators and co-authors of this paper.

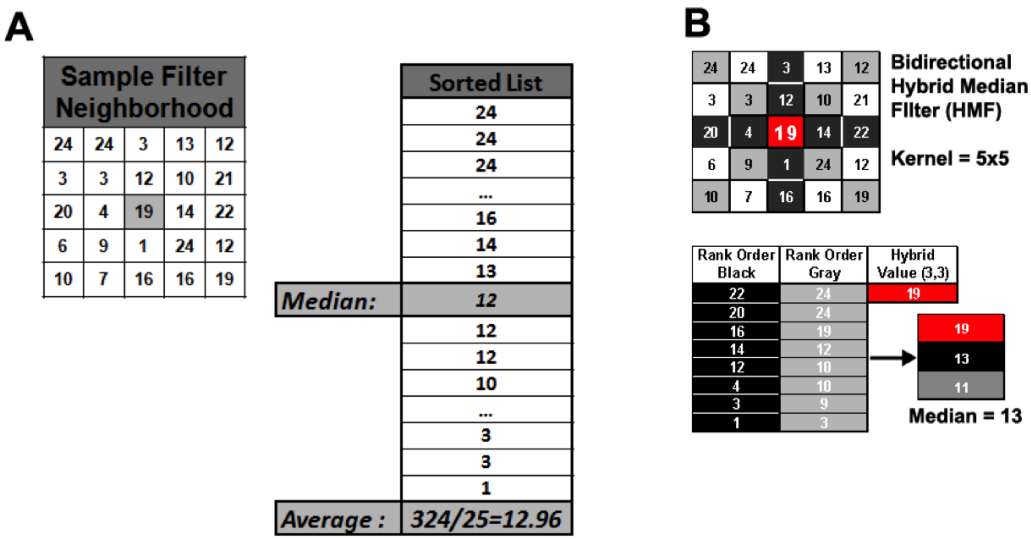


Figure 5.1: Illustration of example 3×3 median and 5×5 HMFs. (A) Elements representing both common 3×3 median and average filters, where the median is obtained from the rank order of all 9 elements and the average is the arithmetic mean of the same elements, are shown. (B) For the bidirectional 5×5 HMF, the median values are extracted from neighborhood diagonals (light gray), cross (black), and center value (hashed) and are used to generate a rank-ordered 3-number list from which a final hybrid median value is determined for each data element in an MTP array.

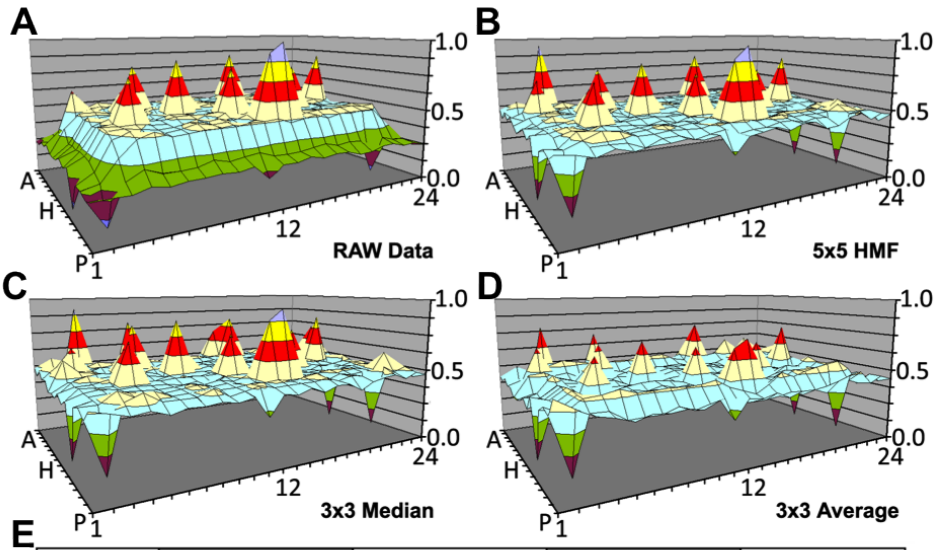
384-Well Microtiter Plate Bioarray Data Format: Plate Median = 50

50	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	25	25	25	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	25	25	25
B	25	28	28	33	33	33	33	33	33	33	33	33	33	33	33	75	33	33	33	33	33	28	28	25
C	25	28	40	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	40	28	25
D	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	10	50	50	50	50	50	50	33	28
E	28	33	50	50	50	50	50	50	50	50	50	50	50	50	95	50	50	10	50	50	50	50	33	28
F	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
G	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
H	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	95
I	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
J	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
K	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
L	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
M	28	33	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	33	28
N	25	28	40	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	40	28	25
O	25	28	4	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	28	28	25
P	25	25	25	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	25	25	25

Date Table - Properties of Highlighted Filter Mask and Raw Data Scale

		HMF Values			Background		$C_{i,j} = (G \div L_{i,j}) \times MTP_{i,j}$		
$ID_{i,j}$	N	Gray	Black	$MTP_{i,j}$	$L_{i,j}$	G	Transform	$G \div L_{i,j}$	$C_{i,j}$
A-01	7	34	25	25	25	50	=(50÷25)x25	2.0	50
D-16	17	50	50	10	50	50	=(50÷50)x10	1.0	10
H-24	11	42	28	95	42	50	=(50÷42)x95	1.2	113
K-10	17	50	50	50	50	50	=(50÷50)x50	1.0	50
O-03	14	28	33	4	28	50	=(50÷28)x4	1.8	7

Figure 5.2: Application of the 5×5 bidirectional HMF to a MTP array. Various placements of the HMF on an example array illustrate the function of the filter and the effect of peripheral wells on filter size and the total number (N) of sampled elements. As the 5×5 HMF passes over the array, it returns a unique value estimating the local neighborhood ($L_{i,j}$) corresponding to the raw data element at its center ($MTP_{i,j}$). A linear transform is then applied to the raw data center element. The linear transform divides the dataset global median (G) (50 as indicated in top left corner) by the filter hybrid median ($L_{i,j}$) to produce a simple scalar that is multiplied by $MTP_{i,j}$ to yield the corrected value ($C_{i,j}$). The table (bottom) tracks $MTP_{i,j}$ and $C_{i,j}$ for the illustrated operations.



KERNEL FILTERS	Normalized Mean*			Standard Deviation*			CV*			Dynamic Range	
	Central	Edges	Whole	Central	Edges	Whole	Central	Edges	Whole	Dynamic Range	x RAW
RAW Data	0.997	0.594	0.842	0.066	0.084	0.210	6.6%	14.2%	24.9%	0.65	1.00
3x3 Median	0.981	0.948	0.968	0.068	0.092	0.079	6.9%	9.7%	8.2%	4.61	7.09
5x5 Median	0.988	0.852	0.936	0.100	0.128	0.130	10.1%	15.0%	13.8%	2.03	3.12
3x3 Average	0.918	0.837	0.887	0.093	0.090	0.100	10.1%	10.8%	11.2%	2.12	3.26
5x5 Average	0.916	0.715	0.839	0.106	0.089	0.140	11.6%	12.5%	16.7%	1.37	2.11
5x5 HMF	0.983	0.937	0.966	0.067	0.082	0.077	6.8%	8.8%	7.9%	4.23	6.51

* Hits removed for calculation

Figure 5.3: Comparison of local background estimators. The performances of the average, median, and bidirectional hybrid median filters (HMFs) are compared using the Synthetic1 dataset. The raw data are shown in (A) and the array corrections are shown in (B) for the 5×5 HMF, (C) for the 3×3 median filter, and (D) for the average filter. A data table (E) statistically summarizes the capacity of each filter to smooth localized background distortion while preserving hit amplitudes in the Excel array. Mean, standard deviation, and coefficient of variation (CV) were calculated with the simulated hits removed. Dynamic range was calculated according to Equation (5.7). The table breaks down corrective performance at the edges and central region of the array. Edge regions are defined by a 2-element wide border plus 1 element nested at each corner. The 5×5 median and average filters are not shown in surface plots because of reduced correction efficiency and method redundancy, but are summarized in (E). The scale and color codes are identical, with each color corresponding to 12.5% of the total range.

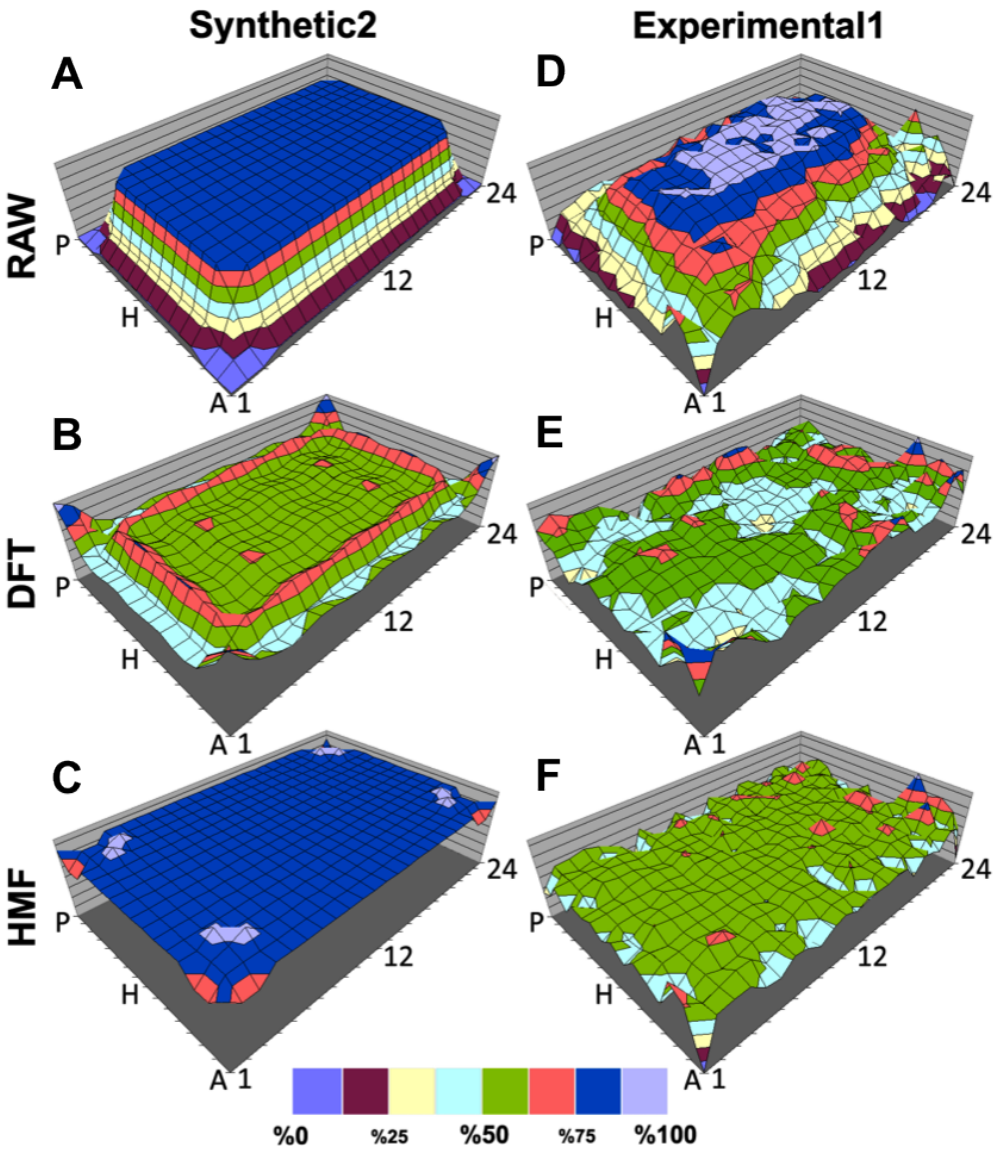


Figure 5.4: Panel of surface maps comparing DFT and HMF correction performances. 3D plots of Synthetic2 dataset (A) and Experimental1 data acquired on the INCell1000 high content microscope (D), both without simulated hits, are compared with 3D plots after correction by the DFT (B, E) and 5×5 bidirectional hybrid median (C, F) filters. Color intervals represent 12% of the RAW array range. All MTPs shown are normalized with respect to the maximum and minimum of their respective RAW dataset. The RAW datasets shown, therefore, span 0%–100%.

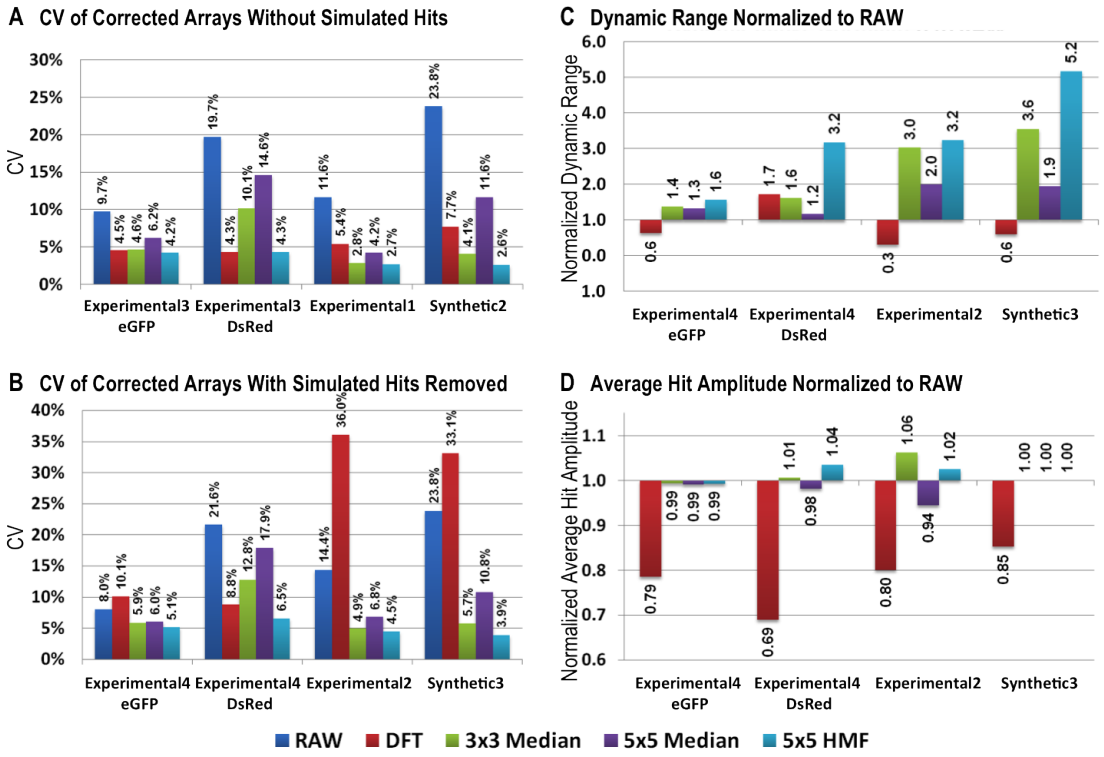


Figure 5.5: Summary of background estimator performance. (A) The coefficient of variations (CVs) for the corrected Synthetic2, Experimental1, and the eGFP and DsRed datasets of Experimental3 that contained no simulated hits are shown. (B) The CVs of corrected Synthetic3, Experimental2, and the eGFP and DsRed datasets of Experimental4, which contained simulated hits that were removed prior to CV calculations, are shown. (C) Shows the effect of array corrections on dynamic range. The average absolute deviation of hits from the array mean is divided by the background standard deviation (values shown above dynamic range bars). The corrected arrays were then normalized to the RAW data dynamic range by division. (D) The average absolute deviations of hits from each array were normalized to the RAW data and shown as a percentage of the RAW data amplitude. Importantly, (C) shows the consolidated effect of the correction method (noise, localized distortion, and preservation of dynamic range), whereas (D) tests if dynamic range improvements were made by erosion or exaggeration of array hit amplitudes. See Methods for details on calculations.

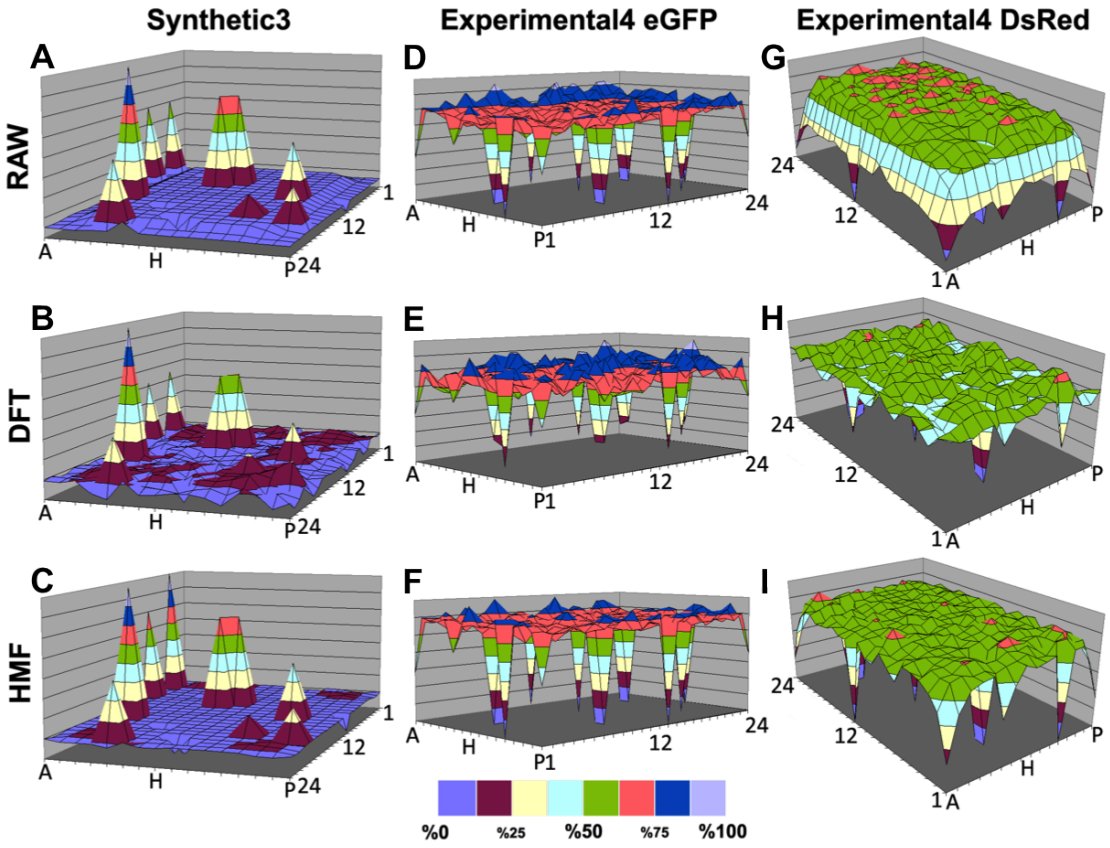


Figure 5.6: Panel of surface maps comparing corrections to arrays with nested outliers. Surface maps of the Synthetic3 dataset (AC) and Experimental4 eGFP (DF) and Experimental4 DsRed (GI) are shown, as described in Methods and Table 5.1. The RAW data are shown in (A, D, and G) and the data as corrected by the 5×5 bidirectional hybrid median filter (HMF) (C, F, and I) and the Discrete Fourier Transform (DFT) (B, E, and H) methods. The Experimental4 eGFP and DsRED data are shown separately but each well contains both fluorescent proteins. The eGFP and DsRED differ in localized edge distortion prior to correction (see Discussion). Color intervals represent 12% of the RAW array range. All MTPs shown are normalized with respect to the maximum and minimum of their respective RAW dataset. The RAW datasets shown, therefore, span 0%-100%.

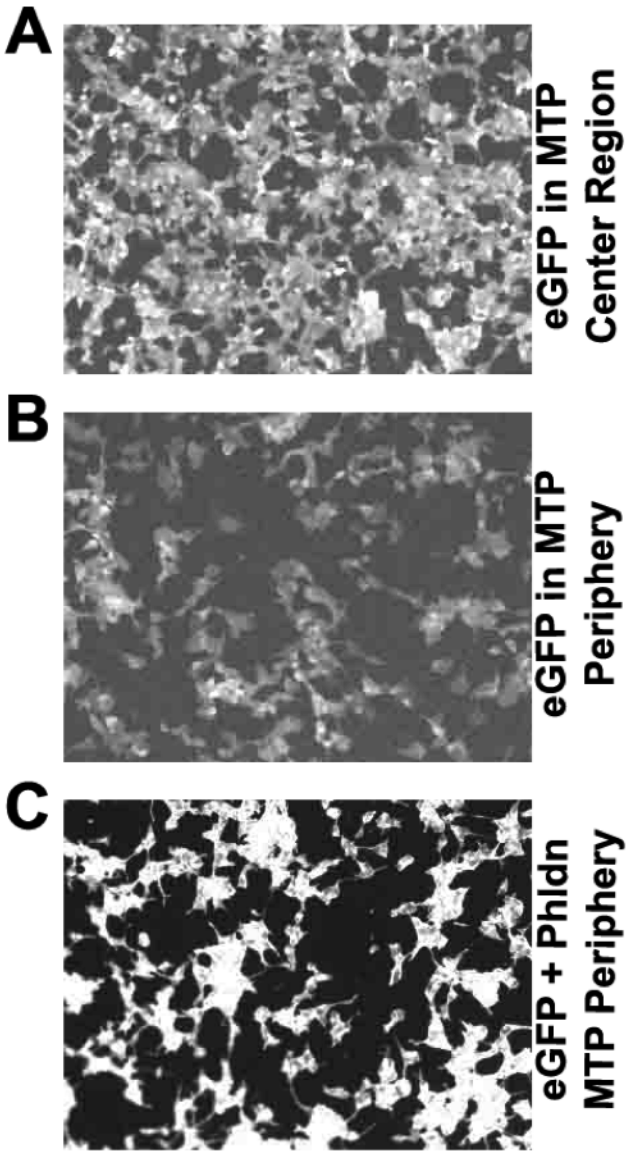


Figure 5.7: Example MTP edge distortion in the cell- based fluorescent assay. The example eGFP raw images from MTP used in the Experimental2 dataset are shown. Panel (A) from a central well is brighter than the image in (B) from a peripheral well, demonstrating an edge effect. (C) Shows an example simulated hit. Brightness was increased artificially in (A) and (B), relative to (C) for display, resulting in higher backgroundsnote the resulting higher backgrounds and lower contrast.

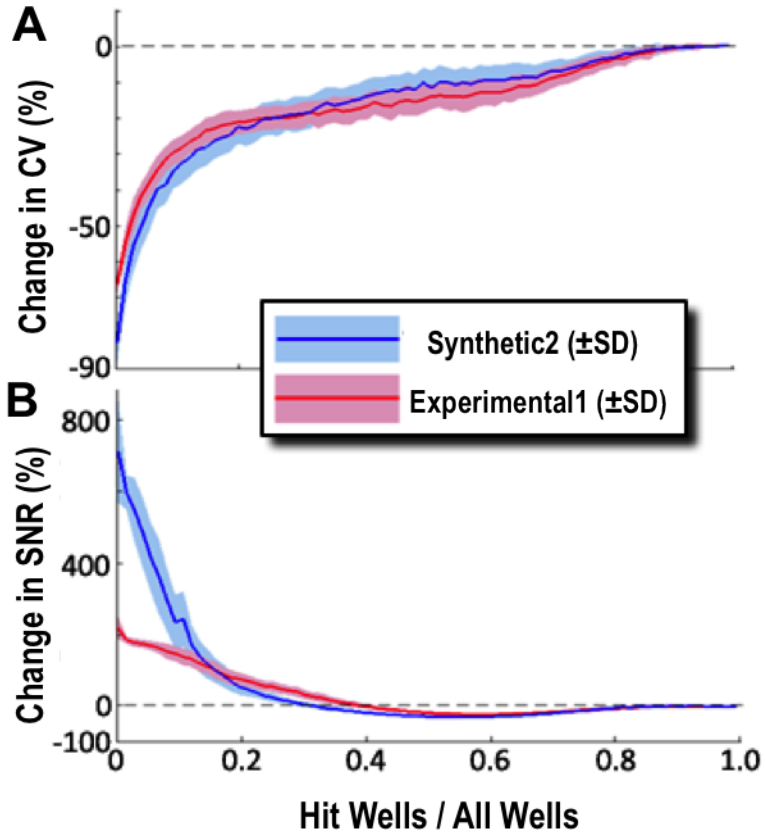


Figure 5.8: Effect of hit density on 5×5 hybrid median filter (HMF) function. Simulated hits were positioned randomly into the Synthetic2 and Experimental1 (Figure 5.4 panels A and D, respectively) and the effect on coefficient of variation (CV) and signal-to-noise ratio (SNR) were determined for increasing numbers of hits (from 1 to 382 in increments of 5) along with standard deviation obtained over 100 trials. (A) Effect on CV: a negative value in the vertical axis signifies an improvement (reduction) in CV. (B) Effect on SNR: a positive value in the vertical axis signifies an improvement in SNR. Note that the 5×5 HMF improves CV and SNR at all hit densities; however, the magnitude of the effect diminishes as the hit density increases.

Bibliography

[1] W.J. Allard, J. Matera, M.C. Miller, M. Repollet, M.C. Connelly, C. Rao, A.G. Tibbe, J.W. Uhr, and L.W.W.M. Terstappen. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clinical Cancer Research*, 10(20):6897–904, 2004.

[2] T. Anagnostou, M. Remzi, M. Lykourinas, and B. Djavan. Artificial neural networks for decision-making in urologic oncology. *Eur Urol*, 43(6):596–603, 2003.

[3] Grasl-Kraupp B, Ruttkay-Nedecky B, Koudelka H, Bukowska K, Bursch W, and Schulte-Hermann R. In situ detection of fragmented dna (tunel assay) fails to discriminate among apoptosis, necrosis, and autolytic cell death: a cautionary note. *Hepatology*, 21(5):1465–8, May 1995.

[4] J. Bartek, J. Bartkova, J. Taylor-Papadimitriou, A. Rejthar, J. Kovarik, Z. Lukas, and B. Vojtesek. Differential expression of keratin 19 in normal human epithelial tissues revealed by monospecific monoclonal antibodies. *Histochem J*, 18(10):565–75., 1986.

[5] M.S. Bartlett. Properties of sufficiency and statistical tests. *Proc Royal Soc London, Mathematical and Physical Sciences*, 160(901):268–282, 1937.

[6] P.D. Beitsch and E. Clifford. Detection of carcinoma cells in the blood of breast cancer patients. *Am J Surgery*, 180:446–9, 2000.

[7] Thomas Berlage. Analyzing and mining image databases. *Drug Discovery Today*, 10(11):795 – 802, 2005.

[8] Anindita Bhoumik, Lisa Gangi, and Ze'ev Ronai. Inhibition of Melanoma Growth and Metastasis by ATF2-Derived Peptides. *Cancer Res*, 64(22):8222–8230, 2004.

[9] Anindita Bhoumik, Tian-Gui Huang, Vladimir Ivanov, Lisa Gangi, Rui F. Qiao, Savio L.C. Woo, Shu-Hsia Chen, and Ze'ev Ronai. An atf2-derived

peptide sensitizes melanomas to apoptosis and inhibits their growth and metastasis. *The Journal of Clinical Investigation*, 110(5):643–650, 9 2002.

[10] Marco Biggiogera, Maria Grazia Bottone, Anna Ivana Scovassi, Cristiana Soldani, Lorella Vecchio, and Carlo Pellicciari. Rearrangement of nuclear ribonucleoprotein (rnp)-containing structures during apoptosis and transcriptional arrest. *Biol. Cell*, 96(8):603–615, 2004.

[11] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.

[12] N. Bjurstam. The gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. *Cancer*, 80(11):2091–9, 1997.

[13] A.J. Bosma, B. Weigelt, A.C. Lambrechts, O.J.H.M. Verhagen, R. Pruntel, A.A.M. Hart, S. Rodenhuis, and L.J. van 't Veer. Detection of circulating breast tumor cells by differential expression of marker genes. *Clin Cancer Res*, 8:1871–7, 2002.

[14] M. Bravo-Zanoguera, B.V. Massenbach, A.R. Kellner, and J.H. Price. High-performance autofocus circuit for biological microscopy‘. *Review of Scientific Instruments*, 69:3955–77, 1998.

[15] H.B. Burke. Artificial neural networks for cancer research: outcome prediction. *Semin Surg Oncol*, 10(1):73–9, 1994.

[16] H.B. Burke, P.H. Goodman, D.B. Rosen, D.E. Henson, J.N. Weinstein, F.E. Jr. Harrell, J.R. Marks, D.P. Winchester, and D.G. Bostwick. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer*, 79(4):857–62, 1997.

[17] Paul J. Bushway, Behrad Azimi, Susanne Heynen-Genel, Jeffrey H. Price, and Mark Mercola. Hybrid median filter background estimator for correcting distortions in microtiter plate data. *ASSAY and Drug Development Technologies*, 8(2):238–250, April 2010.

[18] Paul J. Bushway, Mark Mercola, and Jeffrey H. Price. A comparative analysis of standard microtiter plate reading versus imaging in cellular assays. *ASSAY and Drug Development Technologies*, 6(4):557–567, August 2008.

[19] L.F. Campos, A.C. Silva, and A.K. Barros. Independent components analysis and neural networks applied for classification of malignant, benign and normal tissue in digital mammography. *Methods Inf Med*, 46(2):212–5, 2007.

[20] H. Ceelie, R.B. Dinkelaar, and W. van Gelder. Examination of peripheral blood films using automated microscopy; evaluation of diffmaster octavia and cellavision dm96. *J Clin Pathol*, 60(1):72, 2007.

[21] R.F. Chang, S.F. Huang, W.K. Moon, Y.H. Lee, and D.R. Chen. Solid breast masses: neural network analysis of vascular features at three-dimensional power doppler us for benign or malignant classification. *Radiology*, 243(1):56–92, 2007.

[22] K. L. Cheung, C. R. Graves, and J. F. Robertson. Tumour marker measurements in the diagnosis and monitoring of breast cancer. *Cancer Treat Rev*, 26(2):91–102., 2000.

[23] Peter G. H. Clarke. Developmental cell death: morphological diversity and multiple mechanisms. *Anatomy and Embryology*, 181(3):195–213, 03 1990.

[24] W.S. Cleveland and S.J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statistical Association*, 83:596–610, 1988.

[25] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287 – 314, 1994. Higher Order Statistics.

[26] P. Corradini, C. Voena, M. Astolfi, S. Dell'Oro, S. Pilotti, G. Arrigoni, M. Bregni, A. Pileri, and A.M. Gianni. Maspin and mammaglobin genes are specific markers for rt-pcr detection of minimal residual disease in patients with breast cancer. *Annals of Oncology*, 12:1693–98, 2001.

[27] N. Deglon, J. L. Tseng, J. C. Bensadoun, A. D. Zurn, Y. Arsenijevic, L. Pereira de Almeida, R. Zufferey, D. Trono, and P. Aebischer. Self-inactivating lentiviral vectors with enhanced transgene expression as potential gene transfer system in parkinson's disease. *Hum Gene Ther*, 11(1):179–90, 2000.

[28] G. Deng and Y. S. Kim. Quantitation of erbb-2 gene copy number in breast cancer by an improved polymerase chain reaction (pcr) technique, competitively differential pcr. *Breast Cancer Res Treat*, 58(3):213–7., 1999.

[29] M. J. Duffy. Ca 15-3 and related mucins as circulating markers in breast cancer. *Ann Clin Biochem*, 36(Pt 5):579–86., 1999.

[30] L.V. Fausett. *Fundamentals of Neural Networks*. Prentice Hall, Upper Saddie River, NJ, 1994.

[31] Yan Feng, Timothy J. Mitchison, Andreas Bender, Daniel W. Young, and John A. Tallarico. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nature reviews. Drug discovery*, 8(7):567–578, July 2009.

[32] M. Fiche, H. Avet-Loiseau, M. F. Heymann, F. Moussaly, C. Digabel, M. Joubert, J. M. Classe, F. Dravet, P. Fumoleau, J. Ross, and C. M. Maugard. Genetic alterations in early-onset invasive breast carcinomas: correlation of c-erbb-2 amplification detected by fluorescence in situ hybridization with p53 accumulation and tumor phenotype. *Int J Cancer*, 84(5):511–5, 1999.

[33] Burnham Center for Chemical Genomics. High throughput imaging assay for hepatic lipid droplet formation (pubchem aid: 1656), April 2009.

[34] Y Gavrieli, Y Sherman, and S A Ben-Sasson. Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *The Journal of Cell Biology*, 119(3):493–501, 1992.

[35] A. F. Gazdar, V. Kurvari, A. Virmani, L. Gollahon, M. Sakaguchi, M. Westerfield, D. Kodagoda, V. Stasny, H. T. Cunningham, II Wistuba, G. Tomlinson, V. Tonk, R. Ashfaq, A. M. Leitch, J. D. Minna, and J. W. Shay. Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. *Int J Cancer*, 78(6):766–74., 1998.

[36] B.T. Greer and J. Khan. Diagnostic classification of cancer using dna microarrays and artificial intelligence. *Ann NY Acad Sci*, 1020:49–66, 2004.

[37] B.T. Greer and J. Khan. Online analysis of microarray data using artificial neural networks. *Methods Mol Biol*, 377(37):61–74, 2007.

[38] Georg Häcker. The morphology of apoptosis. *Cell and Tissue Research*, 301(1):5–17, 06 2000.

[39] S. Hamada and S. Fujita. Dapi staining improved for quantitative cytofluorometry. *Histochemistry*, 79(2):219–26, 1983.

[40] I. Hamaguchi, N. B. Woods, I. Panagopoulos, E. Andersson, H. Mikkola, C. Fahlman, R. Zufferey, L. Carlsson, D. Trono, and S. Karlsson. Lentivirus vector gene expression during es cell-derived hematopoietic development in vitro. *J Virol*, 74(22):10778–84, 2000.

[41] Yong Hwan Han, Hwa Jin Moon, Bo Ra You, and Woo Hyun Park. The effect of mg132, a proteasome inhibitor on hela cells in relation to cell growth, reactive oxygen species and gsh. *Oncology Reports*, 22(1):215–221, July 2009.

[42] Laboratory of Computer Helsinki University of Technology and Information Science. Fastica.

[43] Susanne Heynen. *Fluorescent feature fidelity and explora- tion of automated cell-by-cell classification in model populations for high-throughput image cytometry.* PhD thesis, University of California San Diego, 2002.

[44] J. Hoff. Methods of blood collection in the mouse. *Lab Animal*, 29(10):47–53, 2000.

[45] Verena Horneffer, Norbert Linz, and Alfred Vogel. Principles of laser-induced separation and transport of living cells. *Journal of Biomedical Optics*, 12(5):054016, 2007.

[46] K Huang and RF Murphy. From quantitative microscopy to automated image understanding. *Journal of Biomedical Optics*, 9(5):893–912, 2004.

[47] Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, M.L. Giger, R.A. Schmidt, C.J. Vyborny, and K. Doi. Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology*, 198(3):671–8, 1996.

[48] T. Jirapech-Umpai and S. Aitken. Feature selection and classification for microarray data analysis: evolutonary methods for identifying predictive genes. *BMC Bioinformatics*, 6(1):148–58, 2005.

[49] M. Jirari. A computer aided detection system for digital mammograms based on radial basis functions and feature extraction techniques. *Conf Proc IEEE Eng Med Biol Soc*, 4:4457–60, 2005.

[50] J. D. Johnson, R. A. Dennull, L. Gerena, M. Lopez-Sanchez, N. E. Roncal, and N. C. Waters. Assessment and continued validation of the malaria sybr green i-based fluorescence assay for use in malaria drug screening. *Antimicrob Agents Chemother*, 51(6):1926–33, 2007.

[51] I Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.

[52] S. Joo, W.K. Moon, and H.C. Kim. Computer-aided diagnosis of solid breast nodules on ultrasound with digital image processing and artificial neural network. *Conf Proc IEEE Eng Med Biol Soc*, 2:1397–400, 2004.

[53] S G Kalinichenko and N Y Matveeva. Morphological characteristics of apoptosis and its significance in neurogenesis. *Neurosci Behav Physiol*, 38(4):333–344, May 2008.

[54] B.L. Kalman, W.R. Reinus, S.C. Kwasny, A. Laine, and L. Kotner. Prescreening entire mammograms for masses with artificial neural networks: preliminary results. *Acad Radiol*, 4(6):405–14, 1997.

[55] S. Kasimir-Bauer, C. Oberhoff, K. Sliwinska, R. Neumann, A.E. Schindler, and S. Seeber. Evaluation of different methods for the detection of minimal residual disease in blood and bone marrow of patients with primary breast cancer: importance for clinical use? *Breast Cancer Res Treat*, 69:123–32, 2001.

[56] B. P. Kelley, M. R. Lunn, D. E. Root, S. P. Flaherty, A. M. Martino, and B. R. Stockwell. A flexible data analysis tool for chemical genetic screens. *Chem Biol*, 11(11):1495–503, 2004.

[57] J. F. Kerr, A. H. Wyllie, and A. R. Currie. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer*, 26(4):239–257, August 1972.

[58] L. Khriji and M. Gabbouj. Median-rational hybrid filters. In *International Conference on Image Processing ICIP'98*, Chicago, Illinois, USA, 1998.

[59] L. Khriji and M. Gabbouj. Median-rational hybrid filters for image restoration. *IEE Electronics Letters*, 34:977–979, 1998.

[60] S.J. Kim, N. Ikeda, E. Shiba, Y. Takamura, and S. Noguchi. Detection of breast cancer micrometastases in peripheral blood using immunomagnetic separation and immunocytochemistry. *Breast Cancer*, 8(1):63–9, 2001.

[61] D.N. Krag, T. Ashikaga, T.J. Moss, R.E. Kusminsky, S. Feldman, N.Z. Carp, F.L. Moffat, P.D. Beitsch, T.G. Frazier, T.A. Gaskin, J.W. Shook, S.P. Harlow, and D.L. Weaver. Breast cancer cells in blood: a pilot study. *Breast Journal*, 5(6):354–8, 1999.

[62] A. Kratz, H.I. Bengtsson, J.E. Casey, J.M. Keefe, G.H. Beatrice, D.Y. Grzybek, K.B. Lewandrowski, and E.M. Van Cott. Performance evaluation of the cellavision dm96 system: Wbc differentials by automated digital image analysis supported by an artificial neural network. *Am J Clin Pathol*, 124(5):770–81, 2005.

[63] I. Levner. Feature selection and nearest centroid classification for protein mass spectrometry. *BMC Bioinformatics*, 6(1):68–81, 2005.

[64] JE Liebmann, Cook J.A., Lipschultz C, Teague D, Fisher J, and Mitchell JB. Cytotoxic studies of paclitaxel (taxol) in human tumour cell lines. *Br. J. Cancer*, 68(6):1104–9, Dec 1993.

[65] J.S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice Hall, Upper Saddle River, NJ, 1989.

[66] Y. C. Lin, Y. H. Wu Chou, I. C. Liao, and A. J. Cheng. The expression of mammaglobin mrna in peripheral blood of metastatic breast cancer patients as an adjunct to serum tumor markers. *Cancer Lett*, 191(1):93–9., 2003.

[67] J. Linssen, V. Jennissen, J. Hildmann, E. Reisinger, J. Schindler, G. Malchau, A. Nierhaus, and K. Wielckens. Identification and quantification of high

fluorescence-stained lymphocytes as antibody synthesizing/secreting cells using the automated routine hematology analyzer xe-2100. *Cytometry B Clin Cytom*, 72(3):157–66, 2007.

[68] B. Liu, Q. Cui, T. Jiang, and S. Ma. A combinational feature selection and ensemble neural network method for classification of gene expression data. *BMC Bioinformatics*, 5:136–47, 2004.

[69] Lit-Hsin H. Loo, Hai-Jui J. Lin, Robert J. Steininger, Yanqin Wang, Lani F. Wu, and Steven J. Altschuler. An approach for extensibly profiling the molecular states of cellular subpopulations. *Nature methods*, 6(10):759–765, October 2009.

[70] James J. Manfredi and Susan Band Horwitz. Taxol: an antimitotic agent with a new mechanism of action. *Pharmacology and Therapeutics*, 25(1):83 – 125, 1984.

[71] A. R. Manhani, R. Manhani, H. P. Soares, I. Bendit, F. Lopes, A. G. Nicoletti, F. L. Fonseca, M. Novaes, S. M. Zatta, V. Arias, S. Giralt, and A. del Giglio. Ck-19 expression by rt-pcr in the peripheral blood of breast cancer patients correlates with response to chemotherapy. *Breast Cancer Res Treat*, 66(3):249–54., 2001.

[72] G. Mehes, A. Witt, E. Kubista, and P. F. Ambros. Circulating breast cancer cells are frequently apoptotic. *Am J Pathol*, 159(1):17–20., 2001.

[73] L.A. Meinel, A.H. Stolpen, K.S. Berbaum, L.L. Fajardo, and J.M. Reinhardt. Breast mri lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (cad) system. *J Magn Reson Imaging*, 25(1):89–65, 2007.

[74] J. Miyazaki, K. Araki, E. Yamato, H. Ikegami, T. Asano, Y. Shibasaki, Y. Oka, and K. Yamamura. Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology*, 127(1):126–32, 1990.

[75] J.E. Montie and Wei. J.T. Artificial neural networks for prostate carcinoma risk assessment: an overview. *Cancer*, 88(12):2655–60, 2000.

[76] R.N. Naguib and G.V. Sherbet. Artificial neural networks in cancer research. *Pathobiology*, 63(3):129–39, 1997.

[77] A. Nieminen, P. Heinonen, and Y. Neuvo. A new class of detail-preserving filters for image processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1):74–90, 1987.

[78] M. Novaes, I. Bendit, B. Garicochea, and A. del Giglio. Reverse transcriptase-polymerase chain reaction analysis of cytokeratin 19 expression in the peripheral blood mononuclear cells of normal female blood donors. *Mol Pathol*, 50(4):209–11., 1997.

[79] N. N. Nupponen, J. Isola, and T. Visakorpi. Mapping the amplification of eif3s3 in breast and prostate cancer. *Genes Chromosomes Cancer*, 28(2):203–10., 2000.

[80] Obrien, P., Irwin, W., Diaz, D., Howard-Cofield, E., Krejsa, C., Slaughter, M., Gao, B., Kaludercic, N., Angeline, A., Bernardi, P., Brain, P., Hougham, and C. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Archives of Toxicology*, 80(9):580–604, September 2006.

[81] H. Odagiri, J. Wang, and M. S. German. Function of the human insulin promoter in primary cultured islet cells. *J Biol Chem*, 271(4):1909–15, 1996.

[82] W.A. Osta, Y. Chen, K. Mikhitarian, M. Mitas, M. Salem, Y.A. Hannun, D.J. Cole, and W.E. Gillanders. Epcam is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. *Cancer Res*, 65:5818–24, 2004.

[83] K. Pachmann, P. Heiss, U. Demel, and G. Tilz. Detection and quantification of small numbers of circulating tumour cells in peripheral blood using laser scanning cytometer (lsc). *Clin Chem Lab Med*, 39(9):811–7, 2001.

[84] J. H. Price, E. A. Hunter, and D. A. Gough. Accuracy of least squares designed spatial fir filters for segmentation of images of fluorescence stained cell nuclei. *Cytometry*, 25(4):303–16., 1996.

[85] Nunez R. *Flow cytometry for research scientists : principles and applications.* Wymondham, Norfolk, U.K.: Horizon Press, 2001.

[86] M Radisic, RK Iyer, and SK Murthy. Micro- and nanotechnology in cell separation. *International journal of nanomedicine*, 1(1):3–14, 2006.

[87] G. V. Raj, J. G. Moreno, and L. G. Gomella. Utilization of polymerase chain reaction technology in the detection of solid tumors. *Cancer*, 82(8):1419–42., 1998.

[88] Chianese D Rao CG, Doyle GV, Miller MC, Russell T, Sanders RA Jr, and Terstappen LW. Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors. *Int J Oncol*, 27(1):49–57, July 2005.

[89] Sabine Riethdorf, Herbert Fritsche, Volkmar Müller, Thomas Rau, Christian Schindlbeck, Brigitte Rack, Wolfgang Janni, Cornelia Coith, Katrin Beck, Fritz Jänicke, Summer Jackson, Terrie Gornet, Massimo Cristofanilli, and Klaus Pantel. Detection of Circulating Tumor Cells in Peripheral Blood of Patients with Metastatic Breast Cancer: A Validation Study of the CellSearch System. *Clinical Cancer Research*, 13(3):920–928, 2007.

[90] D. E. Root, B. P. Kelley, and B. R. Stockwell. Detecting spatial patterns in biological array experiments. *J Biomol Screen*, 8(4):393–8, 2003.

[91] John C. Russ. *The image processing handbook*. CRC Press, Boca Raton, FL :, 1998.

[92] Antti Saraste. Morphologic criteria and detection of apoptosis. *Herz*, 24(3):189–195, 05 1999.

[93] Antti Saraste and Kari Pulkki. Morphologic and biochemical hallmarks of apoptosis. *Cardiovascular Research*, 45(3):528–537, 2000.

[94] F. V. Shammas, J. A. Van Eekelen, L. Wee, R. Heikkila, and A. Osland. Sensitive and quantitative one-step polymerase chain reaction using capillary electrophoresis and fluorescence detection for measuring cytokeratin 19 expression. *Scand J Clin Lab Invest*, 59(8):635–42., 1999.

[95] M.L. Slovak, L. Tcheurekdjian, F.F. Zhang, and J.L. Murata-Collins. Simultaneous detection of multiple genetic aberrations in single cells by spectral fluorescence in situ hybridization. *Cancer Research*, 61:831–6, 2001.

[96] B. M. Smith, M. J. Slade, J. English, H. Graham, M. Luchtenborg, H. D. Sinnett, N. C. Cross, and R. C. Coombes. Response of circulating tumor cells to systemic therapy in patients with metastatic breast cancer: comparison of quantitative polymerase chain reaction and immunocytochemical techniques. *J Clin Oncol*, 18(7):1432–9., 2000.

[97] Steven W. Smith. *Digital signal processing : a practical guide for engineers and scientists*. Demystifying technology series. Newnes, Amsterdam ; Boston, 2003.

[98] G. K. Smyth and T. Speed. Normalization of cdna microarray data. *Methods*, 31(4):265–73, 2003.

[99] V. Stearns, H. Yamauchi, and D. F. Hayes. Circulating tumor markers in breast cancer: accepted utilities and novel prospects. *Breast Cancer Res Treat*, 52(1-3):239–59, 1998.

[100] E.W. Thompson, S. Paik, N. Brunner, C.L. Sommers, G. Zugmaier, R. Clarke, T.B. Shima, J. Torri, S. Donahue, and M.E. Lippman. Associaton of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines. *J Cell Physiol*, 150(3):535–44, 1992.

[101] H. Thurm, S. Ebel, C. Kentenich, A. Hemsen, S. Riethdorf, C. Coith, D. Wallwiener, S. Braun, C. Oberhoff, F. Janicke, and K. Pantel. Rare expression of epithelial cell adhesion molecule on residual micrometastatic breast cancer cells after adjuvant chemotherapy. *Clinical Cancer Research*, 9:2598–605, 2003.

[102] K. Truong, M. N. Guilly, M. Gerbault-Seureau, B. Malfoy, P. Vielh, C. A. Bourgeois, and B. Dutrillaux. Quantitative fish by image cytometry for the detection of chromosome 1 imbalances in breast cancer: a novel approach analyzing chromosome rearrangements within interphase nuclei. *Lab Invest*, 78(12):1607–13., 1998.

[103] Christopher Walker Vladislava O Melnikova, Svetlana V Bolshakov and Honnavara N Ananthaswamy. Genomic alterations in spontaneous and carcinogen-induced murine melanoma cell lines. *Oncogene*, 23(13):2347–2356, 2004.

[104] L. Wasserman, A. Dreilinger, D. Easter, and A. Wallace. A seminested rt-pcr assay for her2/neu: initial validation of a new method for the detection of disseminated breast cancer cells. *Mol Diagn*, 4(1):21–8., 1999.

[105] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.

[106] Philip T. H. Went, Alessandro Lugli, Sandra Meier, Marcel Bundi, Martina Mirlacher, Guido Sauter, and Stephan Dirnhofer. Frequent epcam protein expression in human carcinomas. *Human Pathology*, 35(1):122 – 128, 2004.

[107] T.E. Witzig, B. Bossy, T. Kimlinger, P.C. Roche, J.N. Ingle, C. Grant, J. Donohue, V.J. Suman, D. Harrington, J. Torre-Bueno, and K.D. Bauer. Detection of circulating cytokeratin-positive cells in the blood of breast cancer patients using immunomagnetic enrichment and digital microscopy. *Clin Cancer Res*, 8:1085–91, 2002.

[108] J. Woods, J. Biemond, and A. Tekalp. Boundary value problem in image restoration. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85.*, volume 10, pages 692–695, 1985.

[109] J C Wu, G Merlino, and N Fausto. Establishment and characterization of differentiated, nontransformed hepatocyte cell lines derived from mice transgenic for transforming growth factor alpha. *PNAS*, 91(2):674–678, 1994.

[110] Y. Wu, Doi. K., M.L. Giger, and R.M. Nishikawa. Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks. *Med Phys*, 19(3):555–60, 1992.

[111] Y.C. Wu, M.T. Freedman, A. Hasegawa, R.A. Zuurbier, S.C. Lo, and S.K. Mun. Classification of microcalcifications in radiographs of pathologic specimens for the diagnosis of breast cancer. *Acad Radiol*, 2(3):199–204, 1995.

[112] Norio Yamamoto, Ping Jiang, Meng Yang, Mingxu Xu, Kensuke Yamauchi, Hiroyuki Tsuchiya, Katsuro Tomita, Geoffrey M. Wahl, Abdool R. Moossa, and Robert M. Hoffman. Cellular Dynamics Visualized in Live Cells in Vitro and in Vivo by Differential Dual-Color Nuclear-Cytoplasmic Fluorescent-Protein Expression. *Cancer Res*, 64(12):4251–4256, 2004.

[113] X. Yang, Q. Tahin, Y. F. Hu, I. H. Russo, B. R. Balsara, D. Mihaila, C. Slater, J. C. Barrett, and J. Russo. Functional roles of chromosomes 11 and 17 in the transformation of human breast epithelial cells in vitro. *Int J Oncol*, 15(4):629–38., 1999.

[114] Y. H. Yang, S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed. Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, 30(4):e15, 2002.

[115] Tian-Li Yue, Chuanlin Wang, Anne M Romanic, Kristine Kikly, Paul Keller, Watler E DeWolf Jr, Timothy K Hart, Heath C Thomas, Barbara Storer, Juan-Li Gu, Xinkang Wang, and Giora Z Feuerstein. Staurosporine-induced apoptosis in cardiomyocytes: A potential role of caspase-3. *Journal of Molecular and Cellular Cardiology*, 30(3):495 – 507, 1998.

[116] Ji-Hu Zhang, Thomas D. Y. Chung, and Kevin R. Oldenburg. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen*, 4(2):67–73, 1999.

[117] X. Y. Zhong, S. Kaul, Y. S. Lin, A. Eichler, and G. Bastert. Sensitive detection of micrometastases in bone marrow from patients with breast cancer using immunomagnetic isolation of tumor cells in combination with reverse transcriptase/polymerase chain reaction for cytokeratin-19. *J Cancer Res Clin Oncol*, 126(4):212–8., 2000.

[118] V. Zieglschmid, C. Hollmann, B. Gutierrez, W. Albert, D. Strothoff, E. Gross, and O. Bocher. Combination of immunomagnetic enrichment with

multiplex rt-pcr analysis for the detection of disseminated tumor cells. *Anticancer Res*, 25(3A):1803–10, 2005.

[119] R. L. Zimmerman, F. Fogt, and S. Goonewardene. Diagnostic value of a second generation ca 15-3 antibody to detect adenocarcinoma in body cavity effusions. *Cancer*, 90(4):230–4, 2000.