# THE STUDY OF CANCER CELL HETEROGENEITY, BY IMAGE-BASED MIGRATION ANALYSIS

By

Mark Philip Harris

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

August, 2009

Nashville, Tennessee

Approved:

Vito Quaranta

Alissa Weaver

Donna Webb

Jeffery Davidson

Yu Shyr

To my parents,

Howard and Linda Harris,

for a lifetime of support and encouragement

ACKNOWLEDGEMENTS

First, I would like to thank a number of students, postdocs, and professors for their help, suggestions, and contributions to this work. My research has been a multidisciplinary affair, and without the assistance, enthusiasm, and ideas of these diverse individuals, this dissertation would not have been possible. Thanks to Shawn Garbett for the analysis of cellular movement parameters, including coding, production, and final analysis of graphs/statistics for cellular persistence and step-length distributions. Also, for the many helpful conversations on these topics, as well as chaos, higher ordered systems, and mathematics in general. Your programming wizardry was a pleasure to observe. Thanks to Walter Georgescu for discussions and tips about image segmentation and individual cell tracking in phase contrast videomicroscopy. I appreciated your assistance with my attempts at Excel and MATLAB programming. Thanks to Mohamed Hassanein for the production of CA1d clones, and for his aid in cell culture during experiments involving the clones. Our many conversations on the biological relevance of phenotypic variation always helped put my experiments in context. Our frequent coffee trips were always a welcome diversion from the rigors of science. Thanks to Junhwan Jeon, who had the idea to utilize exponential fits of cellular data to model cellular migration using the Langevian equation. This idea did not pan out, but was a very useful way to think about modeling migration nonetheless. Junhwan also jump started my thinking on analysis of cellular persistence and modeling, as well as conducting the original analysis of cellular

persistence on my dataset. Thanks to Eric Kim, who did the coding and data analysis of the Surface Area and DECCA algorithms. DECCA measurement was developed from his original analysis. Thanks to Elizabeth Koeler, who created the original statistical analysis for cell motility. Elizabeth was also very helpful in explaining many statistical principles to me. Thanks to Alka Podar, who had the idea to search for Levy flight/walk behavior on my dataset. This behavior was not found, but it was a critical step towards the finding of a Pareto distribution of step-lengths. Thanks to Brandy Weidow for statistical analysis, help on figure production, and editing of both publications, as well as many tips on scientific communication and statistics. Your hard work and dedication will never be forgotten. Thanks to John Wikswo who had the excellent idea to collaborate with Eric Kim and utilize differential imaging to measure cell motility characteristics.

I would also like to thank my thesis committee (Donna Webb, Jeff Davidson, Alissa Weaver, and Yu Shyr), for all their constructive criticism and their support along the way for a non-traditional cancer biology dissertation topic. This project remained murky and confusing for quite some time as I gradually built up the methods and techniques I needed to explain my thoughts and findings to others. The committee's patience and willingness to let me find my own way was exactly what I needed to pursue the questions that inspired me.

I am most indebted to the dedication of my mentor, Vito Quaranta. His profound ability to inspire kept me motivated through the ups and downs of this project over the years. When others doubted my findings, his insightful insistence to "listen to the data" will never be forgotten.

Next, thanks to all the climbers, cyclists, trail runners, mountain bikers, triathletes, adventure racers, and all the other crazies I call my friends. Your world of commitment to physical excellence and pushing the limits of human endurance was the balance to my life as a lab rat. Thanks for embracing me into your community, bringing me along on all your adventures, and motivating me every day.

And finally, to my wife Elizabeth Harris, who is always there with her radiant smile to keep my spirits high, and our newborn son, Gavin, who (unbeknownst to him), strongly motivated me in the writing of this dissertation. The two of you make life more rewarding than I ever thought possible. Amazingly, it keeps getting better every day.

TABLE OF CONTENTS

Page

DEDICATION ..... iii

ACKNOWLEDGEMENTS ..... iv

LIST OF TABLES ..... x

LIST OF FIGURES ..... xi

LIST OF ABBREVIATIONS ..... xiii

LIST OF DEFINITIONS ..... xiv

Chapter

I. INTRODUCTION ..... 1

1.1 Cancer – The Problem ..... 1
1.1.1 The Origins of Cancer ..... 2
1.1.2 Variability of Cancer ..... 3
1.1.2.1 Variation in Cancer ..... 3
1.1.2.2 Sources of Variation ..... 5
1.2 Summary of Dissertation ..... 8

II. MIGRATION ASSAYS ..... 11

2.1 Migration's Role in Cancer ..... 11
2.1.1 Metastasis ..... 11
2.1.2 Invasion ..... 13
2.1.3 Migration ..... 14
2.2 Migration Assays ..... 14
2.2.1 Quantitation ..... 17
2.2.2 Software ..... 18
2.3 Migration as a Model System to Study Variability ..... 18
2.4 Experimental Design ..... 19
2.5 Summary of Migration Experiments ..... 21

III. ANALYSIS OF CELL LINE SPEED VARIABILITY ..... 26

3.1 Background ..... 26
3.2 Experimental Design ..... 29
3.2.1 Materials and Methods ..... 30
3.2.1.1 Cell Culture ..... 30
3.2.1.2 Single Cell Motility Analysis ..... 32
3.2.1.3 Cell Speed Data Analysis and Statistics ..... 33
3.3 Results ..... 34
3.3.1 Single-cells Exhibit Greater Speed Fluctuation in Cancer Cell Lines ..... 34
3.3.2 Cell-to-cell Variability of Speed is Greater in Cancer Cell Lines ..... 37
3.3.3 Population-level Speed Heterogeneity is Greater in Cancer Cell Lines ..... 37
3.3.4 Day-to-day Speed Variability is Greater in Cancer Cell Lines ..... 38
3.3.5 Cancer Cell Speed Variability Increases in Serum/EGF-Depleted Media ..... 40
3.3.6 Cloned Cancer Cells Maintain Speed Variability and Response to Serum/EGF-Depleted Media ..... 41
3.4 Significance and Discussion ..... 42

IV. DESIGN OF ADDITIONAL METRICS TO QUANTITATE MOTILITY ..... 47

4.1 Background ..... 47
4.2 Experimental Design ..... 49
4.2.1 Persistence ..... 49
4.2.2 Step-Length ..... 50
4.2.3 IMF ..... 50
4.2.4 Materials and Methods ..... 50
4.2.4.1 Persistence Quantitation ..... 51
4.2.4.2 Step-Length Quantitation ..... 51
4.2.4.3 IMF Quantitation ..... 52
4.3 Results ..... 52
4.3.1 Highly Motile Cells are a Minority ..... 52
4.3.2 Persistence Increases for all cell lines in Depleted Media ..... 56
4.3.3 Step-Lengths Appear to Follow a Pareto Distribution ..... 58
4.3.4 IMF of Cancer Cells Increases in Depleted Media ..... 62
4.4 Significance and Discussion ..... 63

V. DEVELOPMENT OF DYNAMIC MULTIVARIATE ASSAYS TO STUDY VARIABILITY ..... 67

5.1 Background of System's Biology ..... 67
5.1.1 System's Approach to Studying Cellular Migration ..... 68
5.1.2 Multivariate Analysis ..... 69
5.2 Experimental Design ..... 70
5.2.1 Surface Area ..... 73
5.2.2 DECCA ..... 77
5.2.3 Materials and Methods ..... 77
5.2.3.1 Cell Culture ..... 77
5.2.3.2 Cell Preparation ..... 77
5.2.3.3 Time-lapse Microscopy ..... 78
5.2.3.4 Cell Speed Quantitation ..... 78
5.2.3.5 Computer-assisted Quantitative Analysis ..... 79
5.2.3.6 Surface Area Quantitation ..... 79
5.2.3.7 DECCA Quantitation ..... 81
5.2.3.8 Data Analysis and Statistics ..... 82
5.3 Results ..... 83
5.3.1 Cell Speed Quantitation ..... 83
5.3.2 Surface Area Quantitation ..... 85
5.3.3 DECCA Quantitation ..... 86
5.3.4 Multivariate Correlation ..... 86
5.4 Significance and Discussion ..... 87

VI. SUMMARY AND DISCUSSION ..... 95

6.1 Project Summary ..... 95
6.1.1 Demonstration of Heterogeneity in Cell Speeds ..... 95
6.1.2 Development of Metrics to Further Quantify the Motility of Epithelial Cells ..... 96
6.1.3 Exploration of Changes in Cell Motility Between Cancer and Non-tumorigenic Cell Lines ..... 98
6.2 Conclusions ..... 100
6.3 Significance ..... 101
6.3.1 Single-cell Assays ..... 101
6.3.2 Influence of Variability on Other Phenotypic Traits ..... 103
6.3.3 Phenotypic Plasticity ..... 104
6.3.4 Relevance to Cancer Progression ..... 105
6.3.5 Distinguishing Cancer from Non-cancer ..... 106
6.4 Questions Raised ..... 109
6.5 Concluding Remarks ..... 111

REFERENCES ..... 112

LIST OF TABLES

Table Page

2.1 Current migration assays ..... 16

2.2 Summary of cell migration experiments ..... 22-25

3.1 Individual cell speeds fluctuate highly and are non-constant ..... 35

4.1 Cell lines populations are non-normal and positively skewed ..... 53

4.2 Persistence time ..... 58

4.3 MLE analysis of step-lengths ..... 61

LIST OF FIGURES

Figure Page

1.1 Tumor microenvironment ..... 4

2.1 Metastasis progression ..... 12

3.1 Measured motility metrics and analyses ..... 31

3.2 Individual cells move with non-constant speed ..... 36

3.3 Cell-to-cell variability of speed is greater in cancer cell lines ..... 39

3.4 Cell tracks ..... 44

3.5 CA1d clones exhibit similar cell speed variability as the parental line ..... 45

4.1 Cell speed is non-normally distributed ..... 54

4.2 Cell speed is non-normally distributed (by experiment) ..... 55

4.3 Persistence time increases for all cell lines in serum/EGF-depleted media . 57

4.4 Step-length distribution is linear on a log-log plot ..... 59

4.5 Step-lengths are organized around a power-law distribution ..... 60

4.6 IMF of cancer cells increases upon serum/EGF-depletion ..... 64

4.7 Single-cell motion model ..... 66

5.1 Overview of multivariate profiling of single-cells ..... 72

5.2 Surface area algorithm for image analysis ..... 74

5.3 DECCA algorithm for image analysis ..... 76

5.4 Quantitation of cell speed, surface area, and DECCA ..... 84

5.5 Correlation between metrics ..... 88

5.6 3-D multivariate results ..... 94

6.1 Motility changes upon serum/EGF-depletion ..... 99

6.2 Non-genetic heterogeneity theory of cancer ..... 107

# LIST OF ABBREVIATIONS

AT1 – MCF10A-AT1 cell line

bLG4 – bacterially derived LG4 (a domain of laminin)

BM – basement membrane

CA1d – MCF10A-CA1d cell line

Col – Collagen IV

DECCA – dynamic expansion/contraction of cell area

ECM – extracellular matrix

EGF – epidermal growth factor

Fn – fibronectin

HS – horse serum

IMF – instantaneous motion fraction

L-15 – Liebovitz L-15 media

Ln – laminin

Ln-5 – also known as Ln-332

MCF – MCF10A cell line

MLE – maximum likelihood estimate

PBS – phosphate buffered saline

PDF – probability density function

PRW – persistent random walk

ROI – region of interest

RT – room temperature

# LIST OF DEFINITIONS

The field of single-cell variability contains a number of terms which are used interchangeably in the literature (e.g. diversity, heterogeneity, noise, variability, and versatility). For this dissertation, the following terms are used consistently for clarity:

- “fluctuation” is a variation in a single-cell metric over time (e.g. speed, surface area).
- “heterogeneity” is a variation in a population-level metric (e.g. average speeds, average DECCA).
- “plasticity” is the ability to maintain metric variation, in a wide variety of conditions.
- “variation” is a set of observed differences.
- “variability” is the tendency of a system to generate differences.

Using these definitions, speed fluctuation, speed heterogeneity, and speed variability are three unique terms.

In addition, many metrics in this analysis were analyzed at both the single-cell and population level. The terms “single-cell” and “population-level” are employed to assist the reader with the distinction, and the following definitions are consistent throughout the text:

- “metric” is a type of measurement used to gauge some quantifiable component of a cell.

- “single-cell metric” is a measurement in which any cell has multiple values over time.
- “population-level metric” is a measurement in which each cell in a population has a single value.

Many metrics can be represented as both single-cell and population-level metrics. For example, speed is a single-cell metric when evaluating the speed fluctuation for individual cells. However, when individual cell speeds are averaged, and all these values are displayed for a cell line, speed becomes a population-level metric.

# CHAPTER I

# INTRODUCTION

## 1.1 Cancer – The Problem

Cancer is a major cause of death throughout the world, accounting for 7.4 million deaths worldwide in 2004 alone. If current worldwide trends continue, 83.2 million people will die of cancer between 2008 and 2015 (World Health Organization, World Heath Statistics, 2008). In the United States, cancer is the second leading cause of death, and accounts for 22.8% of deaths in the general population (Center for Disease Control, National Vital Statistics Reports, 2008). Due to the prevalence of this disease, a large amount of research and funding has been directed at finding a cure for cancer. Great strides have been made, as evidenced by a downward trend in cancer deaths in the US since 1993 (National Cancer Institute, American Cancer Society, 2007). However, there continues to be a great need for additional research, due to the complex nature of the disease. Cancer is not simply one disease, but a cluster of several dozen related diseases. For this reason, an optimal cure for one type of cancer is not necessarily transferable to the others. Most likely, targeted strategies of treatment must be developed for all cancer types.

All subtypes of cancer are progressive diseases, and most are divided into five distinct clinicopathologic stages. Stage 0 refers to carcinoma *in situ*, a cancer that has not spread beyond its encapsulated boundary (the basement

membrane). Stage I cancers indicate penetration, and limited invasion beyond the BM. Stage II and Stage III denote more advanced local spread of the cancer. Stage IV refers to a cancer which has metastasized to a distant area of the body. Cancers of stage 0 and I generally have a good prognosis and are curable (Kumar *et al.*, 2004). Patient prognosis and life expectancy drops significantly when a patient's cancer is found to be Stage III or IV. The overwhelming cause of death for the majority of cancer sub-types is Stage IV metastasis. For this reason, early detection and treatment of low-stage cancer is currently the best chance for cure. However, there is still a great need for treatment of Stage III and IV cancers. Attempts to halt or slow metastasis could dramatically increase the quality and length of life for those living with high-stage cancers.

### 1.1.1 The Origins of Cancer

The complexity of tumor progression and metastasis has driven decades of a reductionist approach to cancer research that has largely focused on identifying the causes of oncogenic transformation, tumor growth, and on dissecting the molecular bases of the disease (Christofori, 2006; Woodhouse *et al.*, 1997). Many molecular markers have been identified as contributors to the formation of detectable metastases (Woodhouse *et al.*, 1997; Chambers *et al.*, 2002), but no single mutation can explain the pathology of cancer. Cancer was first suggested to be a multi-mutational disease in 1953 (Nording, 1953), and since then it has become clear that all human cancers display a multitude of

genetic changes, and that a number of such alterations are required for the step-wise progression of tumor development (Knudson, 1971).

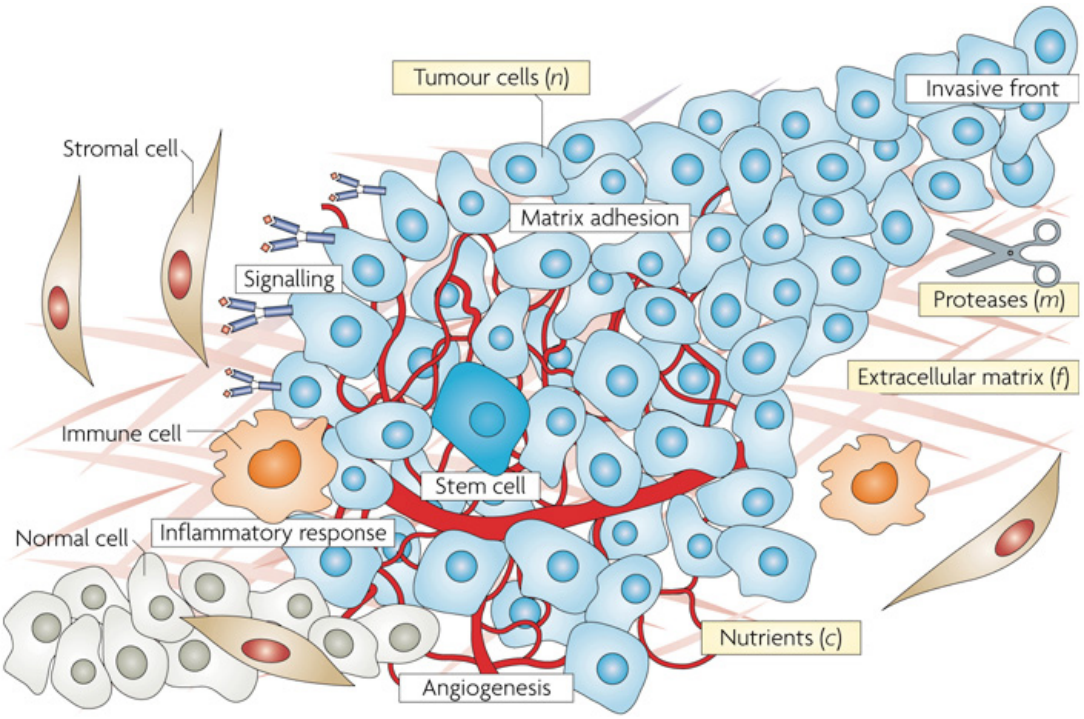
However, disease progression and variation between cancers are also affected by interaction with the surrounding tissue and soluble ligands (microenvironment). These other factors, along with genetic mutation, may help explain the complexities of the disease of cancer (Sherwood, 2006; Bissell and Labarge, 2005). It has become increasingly clear that the tumor microenvironment has a critical role in malignant tumor progression and invasion (Figure 1.1; Liotta *et al.*, 2001; Geho *et al.*, 2005; Quaranta, 2002; Quaranta and Giannelli, 2003; Quaranta *et al.*, 2005).

In total, cancer progression is thought to be a selective process determined by two key factors: the generation of heterogeneity, and the selection of variants most suited to survive (Dexter and Leith, 1986). Understanding how heterogeneity develops is a crucial step towards enhanced clinical outcome.

### 1.1.2 Variability of Cancer

#### *1.1.2.1 Variation in Cancer*

In human medicine, variation in cancer is seen at all levels, including clinical behavior, response to treatment, location of tumor, type of tumor, and variation of each cell within a tumor. Pathologically, cancers are well known to be highly heterogeneous, both grossly and microscopically (Kumar *et al.*, 2004). Morphological heterogeneity, known as pleomorphism, is commonly seen in malignant tumors (Kumar *et al.*, 2004). Cell-to-cell variability is also seen in




Anderson and Quaranta, 2008 **Nature Reviews** | Cancer

**Figure 1.1 - Tumor microenvironment.** Many extrinsic factors present in and around a tmor are involved in cancer progression.

cancer cell lines *in vitro* (Tomelleri *et al.*, 2008; Brodt *et al.*, 1985), including variation of proliferation, migration, metabolism, and morphology. Understanding the causes and significance of cancer cell heterogeneity is a critical step in cancer research, and may have a key impact on clinical cancer therepies and prediction of response to treatment.

*1.1.2.2 Sources of Variation*

There are many sources that may generate heterogeneity within tumors, including: mutations, epigenetics (chromatin remodeling, gene silencing, and X chromosome inactivation), alternative splicing, variation in mRNA expression levels, variation in protein production, microenvironmental influence, and asymmetric distribution of organelles and cellular components during cell division (Goswami *et al.*, 2008). These sources can be broken down into three groups: genetic, epigenetic, and non-genetic sources of heterogeneity.

There is ample evidence supporting the role of genetic heterogeneity in cancer progression. For example, accumulation of mutations and chromosomal abnormalities/instability is common in cancer (Dutrillaux, 1995; Anderson *et al.*, 2001; Giaretti *et al.*, 2003; Hermsen *et al.*, 2002; Rabinovitch *et al.*, 1999; Risques *et al.*, 2003; Sieber *et al.*, 2002), and has been shown to drive heterogeneity of cancer cell lines (Cifone and Fidler, 1981; Cram *et al.*, 1983; Kraemer *et al.*, 1983). This heterogeneity has been correlated with malignant potential by contributing to the creation of variant subpopulations of cells with altered abilities (Chow and Rubin, 1999; Dexter and Leith, 1986; Poste *et al.*, 1981).

Epigenetic heterogeneity has also been implicated in cancer; specifically DNA methylation and histone modification (Egger, et al., 2004; Zhang and Dent, 2005; Santos-Rosa and Caldas, 2005; Feinberg and Tycko, 2004).

Non-genetic heterogeneity's role in cancer, in contrast, has not bee studied in detail. One method that could be used to study non-genetic variability is to examine the variation in specific traits between individual cells within a population. Intriguing observations have been reported through application of these types of experiments, such as the phenotypic variation in genetically identical individuals (Raser and O'Shea, 2005; Samoilov *et al.*, 2006), genes regulating the variability of expression of other genes (Colman-Lerner *et al.*, 2005), methods to separate subpopulations (Loo *et al.*, 2007; Slack *et al.*, 2008), and variation as an evolvable trait (Fraser *et al.*, 2004; Fraser and Kaern, 2009).

Currently, there is renewed interest in investigating the cell-to-cell phenotypic heterogeneity in genetically homogenous cell populations. This is due in part to available new technology that allows quantitation of mRNA expression in single cells (Lin *et al.*, 2007; Subkhankulova *et al.*, 2008), the rise of high-content microscopy (Bullen, 2008; Glory and Murphy, 2007; Pepperkok and Ellenberg, 2006), and the availability of image feature extraction software (Nixon and Aguado, 2007; Lamprecht *et al.*, 2007), all of which make it possible to quantify heterogeneity at the single-cell level.

Through the further development of assays to quantify single-cell variation, we can better understand the non-genetic heterogeneity present within cancer cell lines (Slack *et al.*, 2008; Gordeon *et al.*, 2007). In general, biology at the

single-cell level has diverged sharply from our expectations (Levsky and Singer, 2003), but significant progress has been made in the field with the application of model systems. For example, research on *Escherichia coli* (Rosenfield *et al.*, 2005; Elowitz *et al.*, 2002; Pedraza *et al.*, 2005; Rosenfeld *et al.*, 2005) and *Saccharomyces cerevisiae* (Colman-Lerner *et al.*, 2005; Raser *et al.*, 2004; Bar-Even *et al.*, 2006) demonstrate that even genetically identical single-celled organisms have substantial variation in phenotypic traits (Samoilov *et al.*, 2006). Multiple traits demonstrating variation have been studied at the single-cell level, including: mRNA expression (Grossman *et al.*, 1995; Elowitz *et al.*, 2002; Kemkemer *et al.*, 2002), protein transcription (Rosenfeld *et al.*, 2005; Yu *et al.*, 2006), and phenotypic responses (Gascoigne and Taylor, 2008, Colman-Lerner *et al.*, 2005). These findings further demonstrate that genetically identical cells can have surprisingly large variation in these assorted traits.

Non-genetic heterogeneity can occur along different time courses. For example, variation in mRNA expression of a protein with a high turnover may lead to a subtle, transient phenotypic variation between cells. However, alterations in a more stable, upstream-acting protein, might be passed along during cell division and lead to temporary inheritance, lasting two or more generations. Thus, non-genetic heterogeneity could produce either transient or long-lasting effects, depending on the nature of the sub-cellular variation. In this dissertation, various experimental and analytical methods are developed and utilized to study cancer cell heterogeneity.

## 1.2 Summary of Dissertation

The following chapters are laid out in a logical order, and often refer to one another. The overall outline is: i) introduction of the problem, ii) identification of the model system and experimental design, iii) evidence in support of the hypotheses, iv) design of additional metrics of analysis, v) development of a multivariate method of single-cell analysis, and vi) a summary of the results. To a large extent, the following chapters can be read independently, since they lay out different areas of this research.

The introductory Chapter 1 lays out the problem of cancer as a disease, and identifies variability as a research topic that may spur novel research and advance our understanding of cancer.

Chapter 2 outlines the history, development, and use of migration assays in cancer research. The logic behind utilizing cellular motility as a system for studying variability is laid out. This chapter also introduces the experimental method used for data acquisition in this dissertation, and summarizes all data produced for the work.

Chapter 3 presents current research into variability between single cells, highlighting the phenotypic heterogeneity for a number of motility metrics. The first hypothesis of this dissertation is presented:

*Hypothesis 1: a non-tumorigenic cell line will exhibit less variability of motility than a cancer cell line.*

Data is presented in support of this hypothesis, utilizing the model system introduced in Chapter 2. As a continuation of this research, a second hypothesis is proposed to examine the effect of the microenvironment on cell variability:

*Hypothesis 2: the removal of serum and mitogenic factors (EGF) will decrease the variability of motility of a non-tumorigenic cell line, and will not affect the variability of a cancer cell line.*

Results are provided which demonstrate that serum/EGF-depletion decreases the variability of the parental MCF10A cell line, but actually *increases* the variability of the derived, CA1d cancer cell line. Evidence for these two hypotheses is provided through the quantification of the following metrics: individual cell speed fluctuation, variability of speed, population-level speed heterogeneity, day-to-day speed variability, and speed variation in response to serum/EGF-depleted media.

Chapter 4 reports results from the design of additional metrics to further examine motility for data captured for analysis in Chapter 3. Persistence, step-length distribution, motile cell fraction, and a novel metric, termed “instantaneous motion fraction” (IMF) are presented. Information garnered from these metrics is used to probe differences between non-tumorigenic and cancer cell motility.

Chapter 5 outlines the development of a multivariate dynamic assay to study cell variability. The primary goal was to quantitate multiple single-cell metrics from the same images. Two algorithms are presented here, a surface

area algorithm, and a novel metric of cellular activity, termed “dynamic expansion/contraction of cell area” (DECCA). Data is presented to demonstrate the feasibility of this technique, and its ability to produce three single-cell metrics dynamically (at every time point). This dataset should be considered a proof-of-principle, but some small conclusions can be gleamed from the example presented.

Finally, Chapter 6 offers a brief summary and discussion of the dissertation research as a whole, and the significance and implications of the work.

# CHAPTER II

# MIGRATION ASSAYS

## 2.1 Migration's Role in Cancer

The overwhelming cause of death for the majority of cancer sub-types is Stage IV metastasis (Section 1.1). Attempts to halt or slow metastasis could drastically increase the quality and length of life for those living with high-stage cancers. Migration plays a clear role in metastasis, and further knowledge of this process may aid in the development of new therapeutic interventions.

### 2.1.1 Metastasis

Metastasis occurs when cells from a primary tumor are transported through the bloodstream or lymphatic system, and begin to grow at a secondary site. Metastasis of cancer cells results in 90% of the morbidity and mortality of solid tumors, and is one of the most critical factors for determining cancer prognosis (Sporn, 1997; Gupta *et al.*, 2006). Metastasis occurs in multiple steps, and these steps allow the tumor to: invade through surrounding tissue, intravasate into the vasculature, survive in the blood stream, extravasate at a foreign site, and grow and survive at the foreign site (Figure 2.1, Chambers and Matrisian,1997; Chambers *et al.*, 2000).

A single mutation can rarely account for a metastatic phenotype. Metastasis is most often the result of multiple mutations and phenotypic selection

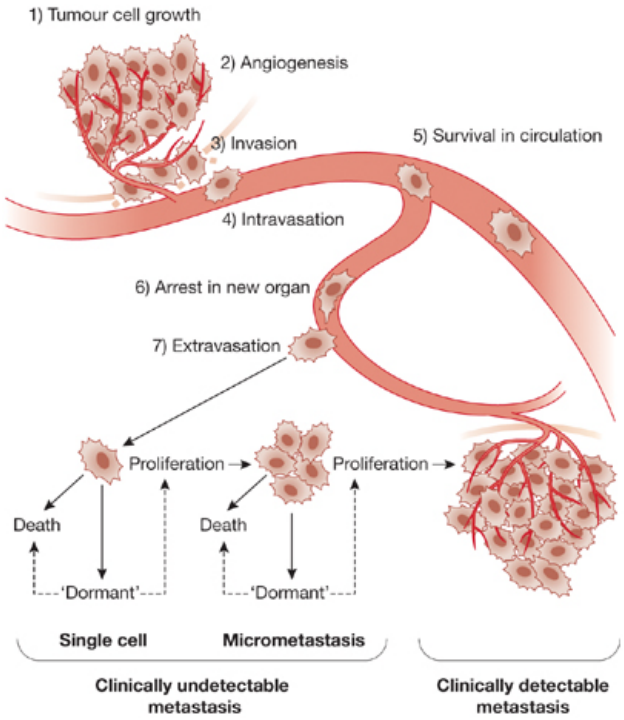



**Figure 2.1 - Metastasis progression.** Metastasis is a multistep process, including: (1) primary tumor cell proliferation, (2) recruitment of blood vessels through angiogenesis, (3) local invasion by cancer cells, (4) intravasation into the circulatory system, (5) survival in the cirulation, (6) arrest at a secondary site, (7) extravasation out of a blood vessel and into the tissue. The process is then repeated at the metastatic site.

within a tumor (Fidler and Hart, 1982; Liotta *et al.*, 1983). As an example, the Massague group used microarray technology to probe for genes that mediate breast cancer metastasis to the lung. Their analysis revealed dozens of genes involved in a clinically relevant manner (Minn *et al.*, 2005). Metastasis is a complex process, and therefore the study of metastasis has been dissected into its component steps, and these processes are studied independently all over the world.

### 2.1.2 Invasion

The ability of cells to invade into surrounding tissues is a critical first phase in metastasis (Aznavoorian *et al.*, 1993; Juhasz *et al.*, 1993; Fidler, 2002). This ability also involves multiple steps, and in the case of carcinomas, cells must cross the basement membrane. To do so, cells must be capable of degrading the extracellular matrix (ECM), and migrating into surrounding tissue. *In vivo*, cancer cells are often seen to invade through a morphological pattern called fingering (Kumar *et al.*, 2004).

Invasion is studied in a variety of ways. *In vitro* assays have been designed to mimic cell invasion, these include: Boyden chamber assay, nest invasion assay, Platypus Oris invasion assay. In this manner, researchers can examine factors that may increase or decrease the ability of cells to invade. It is important to note that while these experiments can suggest particular cell lines or conditions that can lead to invasion, all experiments must be further tested *in vivo* to truly demonstrate an effect on a cell line's invasive capacity.

### 2.1.3 Migration

Cell Migration is often studied in an attempt to further understand the mechanisms behind cancer cell invasion and metastasis, since both mechanisms rely on the motility of cancer cells (Woodhouse *et al.*, 1997, Clark *et al.*, 2000, Condeelis and Segall, 2003). Migration is essential for many physiological and pathological processes such as normal embryonic development, inflammation, wound healing, angiogenesis, and cancer metastasis (Trinkaus, 1984; Lauffenburger and Horwitz, 1996). The process of eukaryotic cell migration includes a set of underlying, interlinked sub-processes including cytoskeletal reorganization, cell protrusion, attachment at the leading edge, cell contraction for physical translocation, and detachment of adhesion at the trailing edge of the cell (Lauffenburger and Horwitz, 1996; Sheetz, 1994; Chicurel, 2002). Thus, migration is a highly complex, dynamic process that is stringently regulated not only by internal cellular signals, but also by external cues from the surrounding microenvironment (Lauffenburger, 1991; Quaranta, 2002). Due to the intricacies of migration, and advancements in technology, migration has been studied in a number of ways.

## 2.2 Migration Assays

Eukaryotic cell migration in various microenvironments has been qualitatively and quantitatively measured using a variety of *in vitro* methods (Chicurel, 2002; Dormann and Weijer, 2006; Roy *et al.*, 2002; Stephens and Allan, 2003; Guan, 2004; Bahnson *et al.*, 2005). These methods generally fall

into two major categories: the first includes techniques that monitor the average behavior of a large populations of cells, such as classical wound-healing assays and Boyden chamber techniques (Guan, 2004); the second involves tracking individual cell motility, which commonly uses time-lapse video-microscopy (Dormann and Weijer, 2006; Guan, 2004). Both types of assays have their merits and limitations (Table 2.1). Studying population motility allows researchers to obtain quantitative data very quickly, often with minimal expense, but only provides information about the average behavior of cells. Single cell analysis allows the acquisition of more detailed information about cell motility (e.g. turn angle, velocity, cellular persistence), but the reaping of quantitative data can be extremely labor intensive. The second category of assays can also reveal sub-cellular dynamics of single cells, such as changes in cytoskeletal organization, lamellipodial protrusion, and focal adhesion turnover (Weaver *et al.*, 2006; Webb *et al.*, 2003).

Single cell analysis is also useful because it eliminates a number of variables that can confound the results obtained from classical wound-healing and Boyden chamber assays, such as cell-cell adhesion and proliferation (DiMilla *et al.*, 1993; Zygourakis and Marckenscoff 1996; Watanabe *et al.*, 1995). Furthermore, cells freshly removed from culture can be plated just prior to performing an experiment; whereas in wound-healing or Electric Cell-substrate Impedance Sensing (ECIS$^{TM}$) assays, cells must first be grown to confluence (Guan, 2004; Lovelady *et al.*, 2007; Keese *et al.*, 2004). Seeding cells just prior to time-lapse microscopy allows the experimenter to control the extracellular

**Table 2.1. Current migration assays.**

| Category | Assay Name | Quantification | Strengths | Weaknesses |
|---|---|---|---|---|
| **Population Level Migration Assays:** | | | | |
| Wound | Scratch | - speed of wound closure quantified by sheet movement<br>- number of cells which have migrated into the wound<br>- time to re-establish confluent monolayer | -Fast<br>-Cheap | -Population based<br>-Poor repeatability<br>-Low sensitivity<br>-No control of ECM<br>-Semi-quantitative<br>-Cell proliferation interferes |
| Wound | Electronic Cell Impedance Sensor (ECIS) | % wound closure quantified by impendence sensor | -Fast<br>-Highly repeatable<br>-High sensitivity | -Requires special equipment<br>-Expensive<br>-Population based<br>-No real-time imaging<br>-No control of ECM<br>-Cell proliferation interferes |
| Wound | Circular Wound Healing (CWA) | % wound closure quantified by image analysis software (e.g. Photoshop) | -Fast<br>-Cheap<br>-Highly repeatable<br>-Tip does not remove matrix | -Requires special equipment<br>-Low sensitivity<br>-No control of ECM<br>-Cell proliferation interferes |
| Wound | Platypus Oris System | % wound closure quantified by plate reader or image analysis software | -Fast<br>-Highly repeatable<br>-Quantifiable with plate reader | -Requires special equipment<br>-Low sensitivity<br>-No ECM present in wound<br>-Cell proliferation interferes |
| Boyden | Boyden | Number of cells which have migrated through the filter | -Fast<br>-Cheap<br>-Chemotaxis<br>-Can utilize ECM | -Cells must migrate through membrane<br>-Only looks at a subset of cells<br>-Low sensitivity<br>-False negatives |
| **Single cell Level Migration Assays:** | | | | |
| Single cell | Microfluidics | X, Y coordinantes of every cell at every time point | -Highly repeatable<br>-Real time imaging<br>-High resolution<br>-Creation of gradients: both soluble and ECM | -Requires special equipment<br>-Cells are unhealthy<br>-Data analysis is labor intensive |
| Single cell | Cell tracking | X, Y coordinantes of every cell at every time point | -Real time imaging<br>-High resolution<br>-Can utilize ECM | -Data analysis is labor intensive<br>-Difficult at high cell density |

matrix (ECM), a useful variable for migration studies. ECM molecules, including proteins such as laminin (Ln) and fibronectin (Fn), play a critical role in regulating cell shape, polarity, and growth (Ingber, 1990; Wang and Inber, 1994). In classical assays, which often involve incubating cells for extended periods of time, experimenters are commonly burdened by cells laying down their own matrix components, confounding ECM experiments using population migration studies. Although single cell analysis overcomes many problems associated with other types of assays, it continues to be a labor intensive measure of cellular migration due to its capacity for generating an abundance of quantitative data.

### 2.2.1 Quantitation of Migration

As outlined on Table 2.1, every migration assay has different strengths, weaknesses, and methods of quantification. These assays are quantified in a number of ways. Most cell migration assays can be divided up into four types of quantification: 1) counting cells (Boyden camber assays, chemotaxis assays); 2) tracking individual cells using software programs (single cell migration assays); 3) distance covered by a migrating sheet of cells (scratch assay, CIA, NEA); and 4) electrical impedance (ECIS). The time involved in quantitation, the level of automation, and the accuracy of all measurements vary between methods. Due to our goal of quantifying heterogeneity of motility within and between cell lines, a single cell migration assay (Section 2.4) was used to produce data for this dissertation. It is clear that only single-cell migration assays can provide

quantitative analysis of cell-to-cell variation, and a multitude of other metrics (e.g. turn angle, cell size, shape).

### 2.2.2 Software used to Quantify Single-cell Migration

Accurately quantifying single-cell migration assays is necessary to dissect out the higher order behavior of cell migration. A whole range of commercially available software programs have attempted to provide automated cell tracking (e.g. MetaMorph, OpenLab and Image J plugins) (Zimmer *et al.*, 2006). These programs typically rely on thresholding, edge detection filters, or template matching—functions useful for capturing cell motility when applied appropriately. Some semi- and fully-automated programs designed for analysis of amoebae, progenitor cells, and various other motile cell populations have also seen some success (Ray and Acton, 2005; Soll, 1995; Wessels *et al.*, 2006); however, such software has failed to catch on with eukaryotic cell migration researchers due to high error rates for tracking phase contrast images (Zimmer and Olivo-Marin, 2005). Manual tracking currently remains the gold standard for phase contrast movies of epithelial cell migration (Zimmer *et al.*, 2006).

## 2.3 Migration as a Model System to Study Variability

Cancer was introduced as a disease of variability in Section 1.1.2. For this dissertation, migration is used as a model system to study the variability of cancer cell lines at the single-cell level (see hypotheses, Section 1.2). There are

a number of reasons why we chose migration as a model system to study cancer variability *in vitro*:

- observation of migrating cells easily demonstrated the extreme amount of variation present, while methods to quantify and statistically represent this variation were lacking
- in cell biology, cell speeds are typically examined by population based motility assays, and thus are represented by an average number, which can lead to misleading conclusions
- our laboratory had access to automated stage microscopes, allowing the collection of up to 120 movies at a time, which made these experiments feasible
- computationally, we now have the storage capacity and image processing ability to undertake these studies
- to our knowledge, there have been no large scale studies of single-cell cancer migration in multiple microenvironments

## 2.4 Experimental Design

ECM proteins were coated on Nunc$^{TM}$ polystyrene, non-tissue culture treated, 6-well microplate dishes (Cole Parmer, Vernon Hills, IL) for 1 h at room temperature (RT). Dishes were then blocked with 5% milk (Regilait, France) in phosphate buffered saline solution (PBS, Invitrogen) for 1 h at 37ºC. Cell lines were trypsinized (TrypLE Select, Invitrogen, Sunnyvale, CA), neutralized with growth media (L-15 media supplemented as appropriate for each cell line),

washed three times in PBS, and $2x10^4$ cells per well were resuspended in growth media in 6-well microplates. Cells were allowed to adhere for 1 h within the heated (37ºC) microscope chamber prior to image capture.

Time-lapse microscopy was conducted using a Zeiss Axiovert 200M microscope (Zeiss, Thornwood, NY) equipped with a temperature-controlled chamber and an automated x-y-z stage. Microscopy was under the control of MetaMorph software (Molecular Devices, Sunnyvale, CA). At the beginning of each experiment (0 h), 4-8 regions of interest (ROI) were selected at random from within each well for imaging. Each ROI was focused manually, and the coordinates saved using MetaMorph's "Multi-dimensional Acquisition" tool for subsequent imaging. Phase-contrast images were captured automatically at each ROI every 5 min, for 4 h. Following image capture, all 49 images from a particular coordinate were combined using MetaMorph to produce image stacks. Cell speed was quantified manually by tracking the center of each cell's nucleus, using MetaMorph's "Track Points" function.

The selection of 5 min time resolution and 4 h movies was based on several criteria:

1. A time resolution of 5 min allows accurate determination of persistence time for epithelial cells (previously reported persistence time was roughly 10 min (Dunn and Brown, 1987)).
2. A movie duration of 4 h is required to determine if cells become diffusive over time, as calculated by the Furth equation for a 10 min persistence time (Furth, 1920).

3. Longer movie durations (more than 8 h) would have allowed cells to alter their matrix, and would not allow us to examine cell movement on different matrices.
4. Limitations of storage space, processing power, and time spent on manual data analysis.

Therefore, the time resolution and overall time for the experiment carefully balanced the need for a detailed dataset, and the need for a large dataset. Since we were interested in studying heterogeneity, we needed hundreds of cells in a number of microenvironmental conditions. Overall, the resolution and time course was based on the ability to answer our questions, as based on previously published literature.

## 2.5 Summary of Migration Experiments

For the analysis of cell migration, 47 individual experiments were undertaken. These include the analysis of 6 cell lines, 2 conditions, and 6 different types of ECM. A full list of every experiment can be found on Table 2.2. In total, over 7,300 cells were tracked by hand, resulting in over 454,000 individual x,y coordinates. Actual x,y coordinates and/or average cell speed values were not included here due to size constraints (300+ pages). The full dataset is located in an electronic copy in the Quaranta laboratory. Chapters 3, 4 and 5 utilize a subset of experiments, highlighted in Table 2.2, to test specific hypotheses.

**Table 2.2 – Summary of cell migration experiments.** This table represents all cell migration experiments performed and quantified as a part of this dissertation. A subset of these data were used for analysis in Chapter 3 and 4 (light gray shading), and Chapter 5 (dark gray shading). Every row of this table is associated with 4 to 8 individual movies. Full movies and data are stored in the Quaranta laboratory at Vanderbilt University.

| Date | Cell | Media | Matrix | Hours | Res(min) | # cells | Data Points |
|---|---|---|---|---|---|---|---|
| 070612 | HT-1080 | Full L-15 | LG3 | 12 | 5 | 53 | 7632 |
| | | L-15 -HS | LG3 | 12 | 5 | 53 | 7632 |
| 070717 | A431 | Full L-15 | Col | 4 | 5 | 56 | 2688 |
| | | Full L-15 | FN | 4 | 5 | 13 | 624 |
| 071026 | HT-1080 | Full L-15 | bLG4 | 4 | 5 | 86 | 4128 |
| | | Full L-15 | LN5 | 4 | 5 | 91 | 4368 |
| | | Full L-15 | PBS | 4 | 5 | 90 | 4320 |
| 071120 | HT-1080 | Full L-15 | bLG4 | 4 | 5 | 65 | 3120 |
| | | Full L-15 | LN5 | 4 | 5 | 86 | 4128 |
| | | Full L-15 | FN | 4 | 5 | 91 | 4368 |
| | | Full L-15 | PBS | 4 | 5 | 59 | 2832 |
| 071129 | HT-1080 | Full L-15 | PBS | 4 | 1 | 17 | 816 |
| | | Full L-15 | LN5 | 4 | 1 | 32 | 1536 |
| 071218 | MCF | Full L-15 | FN | 4 | 5 | 141 | 6909 |
| | | Full L-15 | LN5 | 4 | 5 | 140 | 6860 |
| | AT1 | Full L-15 | FN | 4 | 5 | 95 | 4655 |
| | | Full L-15 | LN5 | 4 | 5 | 103 | 5047 |
| | CA1d | Full L-15 | FN | 4 | 5 | 113 | 5537 |
| | | Full L-15 | LN5 | 4 | 5 | 184 | 9016 |
| 080114 | MCF | Depleted L-15 | FN | 4 | 5 | 40 | 1960 |
| | | Depleted L-15 | LN5 | 4 | 5 | 121 | 5929 |
| | AT1 | Depleted L-15 | FN | 4 | 5 | 31 | 1519 |
| | | Depleted L-15 | LN5 | 4 | 5 | 103 | 5047 |
| | CA1d | Depleted L-15 | FN | 4 | 5 | 35 | 1715 |
| | | Depleted L-15 | LN5 | 4 | 5 | 104 | 5096 |
| 080128 | MCF | Full L-15 | FN | 4 | 5 | 76 | 3724 |
| | AT1 | Full L-15 | FN | 4 | 5 | 79 | 3871 |
| | CA1d | Full L-15 | FN | 4 | 5 | 90 | 4410 |
| 080204 | MCF | Full L-15 | FN | 4 | 5 | 19 | 931 |
| | | Full L-15 | LN5 | 4 | 5 | 34 | 1666 |
| | AT1 | Full L-15 | FN | 4 | 5 | 10 | 490 |
| | CA1d | Full L-15 | FN | 4 | 5 | 35 | 1715 |
| | | Full L-15 | LN5 | 4 | 5 | 34 | 1666 |
| **080213** | **MCF** | **Full L-15** | **LN5** | **4** | **5** | **51** | **2499** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **68** | **3332** |
| | **AT1** | **Full L-15** | **LN5** | **4** | **5** | **59** | **2891** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **79** | **3871** |
| | **CA1d** | **Full L-15** | **LN5** | **4** | **5** | **37** | **1813** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **51** | **2499** |

Table 2.2 continued

| Date | Cell | Media | Matrix | Hours | Res(min) | # cells | Data Points |
|---|---|---|---|---|---|---|---|
| **080228** | **MCF** | **Full L-15** | **LN5** | **4** | **5** | **32** | **1568** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **49** | **2401** |
| | **AT1** | **Full L-15** | **LN5** | **4** | **5** | **34** | **1666** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **32** | **1568** |
| | **CA1d** | **Full L-15** | **LN5** | **4** | **5** | **27** | **1323** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **36** | **1764** |
| **080313** | **MCF** | **Full L-15** | **LN5** | **4** | **5** | **35** | **1715** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **37** | **1813** |
| | **AT1** | **Full L-15** | **LN5** | **4** | **5** | **42** | **2058** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **33** | **1617** |
| | **CA1d** | **Full L-15** | **LN5** | **4** | **5** | **45** | **2205** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **51** | **2499** |
| **080314** | **HT-1080** | **Full L-15** | **FN** | **4** | **5** | **32** | **1568** |
| | | **Full L-15** | **LN5** | **4** | **5** | **110** | **5390** |
| | **A431** | **Full L-15** | **FN** | **4** | **5** | **76** | **3724** |
| | | **Full L-15** | **LN5** | **4** | **5** | **171** | **8379** |
| **080318** | **HT-1080** | **Full L-15** | **FN** | **4** | **5** | **57** | **2793** |
| | | **Full L-15** | **LN5** | **4** | **5** | **160** | **7840** |
| | **A431** | **Full L-15** | **FN** | **4** | **5** | **48** | **6027** |
| | | **Full L-15** | **LN5** | **4** | **5** | **123** | **6027** |
| 080320 | A431 | Full L-15 | FN | 4 | 5 | 151 | 7399 |
| | | Full L-15 | LN5 | 4 | 5 | 154 | 7546 |
| 080417 | CAFTD | Full L-15 | FN | 4 | 5 | 16 | 784 |
| | | Depleted L-15 | FN | 4 | 5 | 19 | 931 |
| | | Full L-15 | LN5 | 4 | 5 | 40 | 1960 |
| | | Depleted L-15 | LN5 | 4 | 5 | 34 | 1666 |
| 080513 | MCF | Full L-15 | LN5 | 4 | 5 | 118 | 5782 |
| | | Depleted L-15 | LN5 | 4 | 5 | 145 | 7105 |
| | AT1 | Full L-15 | LN5 | 4 | 5 | 68 | 3332 |
| | | Depleted L-15 | LN5 | 4 | 5 | 120 | 5880 |
| | CA1d | Full L-15 | LN5 | 4 | 5 | 82 | 4018 |
| | | Depleted L-15 | LN5 | 4 | 5 | 110 | 5390 |
| 080619 | CA1d | Full Matrigel | Matrigel | 4 | 5 | 6 | 294 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 4 | 196 |
| | A431 | Full Matrigel | Matrigel | 4 | 5 | 8 | 392 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 6 | 294 |
| 080620 | CA1d | Full Matrigel | Matrigel | 4 | 5 | 5 | 245 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 7 | 343 |
| | A431 | Full Matrigel | Matrigel | 4 | 5 | 7 | 343 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 4 | 196 |

Table 2.2 continued

| Date | Cell | Media | Matrix | Hours | Res(min) | # cells | Data Points |
|---|---|---|---|---|---|---|---|
| 080707 | MCF | Full L-15 | LN5 | 4 | 5 | 43 | 2107 |
| | | Depleted L-15 | LN5 | 4 | 5 | 72 | 3528 |
| | AT1 | Full L-15 | LN5 | 4 | 5 | 46 | 2254 |
| | | Depleted L-15 | LN5 | 4 | 5 | 63 | 3087 |
| | CA1d | Full L-15 | LN5 | 4 | 5 | 65 | 3185 |
| | | Depleted L-15 | LN5 | 4 | 5 | 51 | 2499 |
| 080708 | MCF | Full Matrigel | Matrigel | 4 | 5 | 36 | 1764 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 42 | 2058 |
| | AT1 | Full Matrigel | Matrigel | 4 | 5 | 27 | 1323 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 61 | 2989 |
| | CA1d | Full Matrigel | Matrigel | 4 | 5 | 41 | 2009 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 26 | 1274 |
| **080721** | **MCF** | **Full L-15** | **LN5** | **4** | **5** | **27** | **1323** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **41** | **2009** |
| | **AT1** | **Full L-15** | **LN5** | **4** | **5** | **34** | **1666** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **60** | **2940** |
| | **CA1d** | **Full L-15** | **LN5** | **4** | **5** | **86** | **4214** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **140** | **6860** |
| | MCF | Full Matrigel | Matrigel | 4 | 5 | 69 | 3381 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 116 | 5684 |
| | AT1 | Full Matrigel | Matrigel | 4 | 5 | 201 | 9849 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 255 | 12495 |
| | CA1d | Full Matrigel | Matrigel | 4 | 5 | 110 | 5390 |
| | | Depleted Matrigel | Matrigel | 4 | 5 | 59 | 2891 |
| **081223** | **MCF** | **Full L-15** | **LN5** | **4** | **5** | **81** | **3969** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **43** | **2107** |
| | **AT1** | **Full L-15** | **LN5** | **4** | **5** | **98** | **4802** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **65** | **3185** |
| | **CA1d** | **Full L-15** | **LN5** | **4** | **5** | **83** | **4067** |
| | | **Depleted L-15** | **LN5** | **4** | **5** | **45** | **2205** |
| 090121 | Clone10 | Depleted L-15 | LN5 | 8 | 5 | 38 | 3686 |
| | | Full L-15 | LN5 | 8 | 5 | 32 | 3104 |
| | Clone11 | Depleted L-15 | LN5 | 8 | 5 | 45 | 4365 |
| | | Full L-15 | LN5 | 8 | 5 | 71 | 6887 |
| | Clone4 | Depleted L-15 | LN5 | 8 | 5 | 47 | 4559 |
| | | Full L-15 | LN5 | 8 | 5 | 65 | 6305 |
| | Clone8 | Depleted L-15 | LN5 | 8 | 5 | 47 | 4559 |
| | | Full L-15 | LN5 | 8 | 5 | 84 | 8148 |
| | Clone9 | Depleted L-15 | LN5 | 8 | 5 | 33 | 3201 |
| | | Full L-15 | LN5 | 8 | 5 | 38 | 3686 |
| | CA1d | Depleted L-15 | LN5 | 8 | 5 | 43 | 4171 |
| | | Full L-15 | LN5 | 8 | 5 | 50 | 4850 |

Table 2.2 continued

| Date | Cell | Media | Matrix | Hours | Res(min) | # cells | Data Points |
|---|---|---|---|---|---|---|---|
| 090123 | Clone10 | Depleted L-15 | LN5 | 8 | 5 | 95 | 9215 |
| | | Full L-15 | LN5 | 8 | 5 | 51 | 4947 |
| | Clone11 | Depleted L-15 | LN5 | 8 | 5 | 26 | 2522 |
| | | Full L-15 | LN5 | 8 | 5 | 50 | 4850 |
| | Clone4 | Depleted L-15 | LN5 | 8 | 5 | 122 | 11834 |
| | | Full L-15 | LN5 | 8 | 5 | 73 | 7081 |
| | Clone8 | Depleted L-15 | LN5 | 8 | 5 | 55 | 5335 |
| | | Full L-15 | LN5 | 8 | 5 | 30 | 2910 |
| | Clone9 | Depleted L-15 | LN5 | 8 | 5 | 48 | 4656 |
| | | Full L-15 | LN5 | 8 | 5 | 13 | 1261 |
| | CA1d | Depleted L-15 | LN5 | 8 | 5 | 76 | 7372 |
| | | Full L-15 | LN5 | 8 | 5 | 64 | 6208 |
| **TOTAL** | | | | **648** | | **7308** | **493,727** |

# CHAPTER III

# ANALYSIS OF CELL LINE SPEED VARIABILITY

## 3.1 Background

Currently, there is renewed interest in investigating cell phenotypic heterogeneity within genetically homogenous cell populations. This is due, in part, to newly available technologies, such as the ability to accurately quantify mRNA expression in single cells (Lin et al., 2007; Subkhankulova *et al.*, 2008), the rise of high-content microscopy (Bullen 2008; Glory and Murphy, 2007; Pepperkok *et al.*, 2006), and the development of image feature extraction software (Nixon and Aguado, 2007; Lamprecht *et al.*, 2007)--all of which make it possible to quantify heterogeneity at the single-cell level. Data collected at the single-cell level can then be used to derive phenotypic distributions within a cell population, rather than be limited to average measurement values. Interest has also been piqued by intriguing observations, such as phenotypic variation in genetically identical individuals (Raser *et al.*, 2005; Samoilov, 2006), genes regulating the variability of expression of other genes (Colman-Lerner *et al.*, 2005), introduction of methods to represent population heterogeneity as subpopulations (Loo *et al.*, 2007; Slack *et al.*, 2008), and the concept of variation itself as an evolvable trait (Fraser *et al.*, 2004). However, investigations into dynamic eukaryotic cell heterogeneity have not yet received the same level of attention. This lag is not surprising, in part because the technology necessary to

quantitate the behavior of thousands of cells over time has just recently been introduced (affordable high-content, time-lapse microscopy). Only one report, to our knowledge, has described dynamic heterogeneity at the single-cell level to date (Gascoigne and Taylor, 2008).

Microscopy has been long used in the biological sciences as a tool for extraction of cellular data. As computer processing power and storage capacity continue to advance, there is a push towards automated quantitation of cellular features from microscopic images. The information contained in biological images is vast, and there are currently groups working both on feature extraction (Lamprecht *et al.*, 2007) and data analysis techniques in order to make sense of overwhelmingly abundant datasets (Loo *et al.*, 2007; Slack *et al.*, 2008; Jones *et al.*, 2009). Both feature extraction and data analysis become more complex when attempting to draw quantitative data from cells dynamically over time.

The motility characteristics, in particular persistence, of many cell types have been studied, but the study of dynamic heterogeneity has been mainly reserved for model organisms such as *S. cerevisiae* and *E. coli*. Therefore, it is not surprising that investigations into cancer cell heterogeneity have not received the same level of attention as that of model organisms. However, the concept of cancer cell heterogeneity is central to several hypotheses of cancer progression and thus research in this area may lead to novel cancer insights. For example, tumor progression is commonly described as a selective process determined by two key factors: the generation of heterogeneity and the selection of variants most suited to survive (Dexter and Leith, 1986). Heterogeneity is most often

viewed as the accumulation of genetic mutations within a tumor, which gives rise to increased variation of cellular phenotypes (Cifone and Fidler, 1981; Cram *et al.*, 1983; Kraemer *et al.*, 1983). Meanwhile, selective pressures within a tumor are also varied, but are known to include: anoxia, malnutrition, fluctuating hormone levels, and interaction with immune cells (Cahill *et al.*, 1999; Casanovas *et al.*, 2005; Mitsumoto *et al.*, 1998; Pennacchietti *et al.*, 2003; Smalley *et al.*, 2005). The interplay between phenotypic selection and tumor progression is a central feature of the clonal evolution theory of cancer (Nowell, 1976), and is thought to be a driver of cancer progression and invasion (Anderson *et al.*, 2006; Anderson and Quaranta, 2008). Further, in the clonal evolution theory, genetic mutations arise, and tumor cells compete against one another for space and resources. Those clones that grow the fastest and are more suited for survival will, over time, constitute the majority of the tumor volume (Hanahan and Weinberg, 2000). There are other sources that also generate heterogeneity within tumors, such as epigenetics, alternative splicing, and variation in mRNA expression levels (Goswami *et al.*, 2009). These types of heterogeneity generation could increase the inherent flexibility within a cancer cell population, and allow the tumor to adapt to and survive in many types of microenvironments.

Here, heterogeneity was examined at the single-cell level through the lens of cell motility. We studied three genetically-related cell lines (one non-tumorigenic, the others cancer) using high-content microscopy. We examined over 1,500 cells and characterized their motility, both at the individual and population levels, with respect to several motility-based metrics, including:

individual cell speed fluctuation, variability of speed, population-level speed heterogeneity, day-to-day speed variability, and speed variation in response to serum/EGF-depleted media. Our experimental workflow is depicted in Figure 3.1. Our primary goal was to pursue a proof-of-principle that heterogeneity of cell motility exists within cell lines and that it can be estimated quantitatively.

Two hypotheses are proposed to address heterogeneity of motility in non-tumorigenic and cancer cell lines:

- *Hypothesis 1: a non-tumorigenic cell line will exhibit less variability of motility than a cancer cell line.*
- *Hypothesis 2: the removal of serum and mitogenic factors (EGF) will decrease the variability of motility of a non-tumorigenic cell line, and will not affect the variability of a cancer cell line.*

### 3.2 Experimental Design

To address these hypotheses, the MCF10A family of cell lines was utilized. MCF10A, a human cell line derived from spontaneous immortalization of normal breast epithelial cells that is non-tumorigenic in nude mice (Miller *et al.*, 1993), MCF10A-AT1 (AT1), a tumorigenic *ras* oncogene transformed version of the parental cell line (Dawson *et al.*, 1996), and MCF10A-CA1d (CA1d), a cell line derived from xenograft-passaging the parental line in nude mice creating a highly invasive cell line (Santner *et al.*, 2001). This model system was selected for a variety of reasons:

- all three cell lines stem from the same genetic background

- the system represents a progression of increasing invasive potential, from non-tumorigenic, to tumorigenic, to invasive
- their *in vivo* ability to form tumors and invade are well characterized
- the system is used regularly in our laboratory for other research projects, so cell behavior *in vitro* is well characterized, and reagents are readily available

Single-cell migration assays (Section 2.4) were used to analyze the speed of the three cell lines in two types of media (full media and serum/EGF- depleted) to examine the role of the microenvironment on cell speed. The logic behind analyzing motility to study heterogeneity can be found in Section 2.3.

The majority of the data analyzed within this chapter was compiled from 5 experiments (see Table 2.2, dates: 080213, 080228, 080313, 080721, 081223). These represent 5 of 7 experiments conducted where MCF, AT1, CA1d cell lines were tested, in both full and serum/EGF-depleted media, on Ln-332. Two dates (080513, 080707) were not included to eliminate an additional microenvironmental variable (different batches of Ln-332).

### 3.2.1 Materials and Methods

*3.2.1.1 Cell Culture*

MCF10A, a human cell line derived from spontaneous immortalization of normal breast epithelial cells that is non-tumorigenic in nude mice (Miller *et al.*, 1993), MCF10A-AT1 (AT1), a tumorigenic *ras* oncogene transformed version of

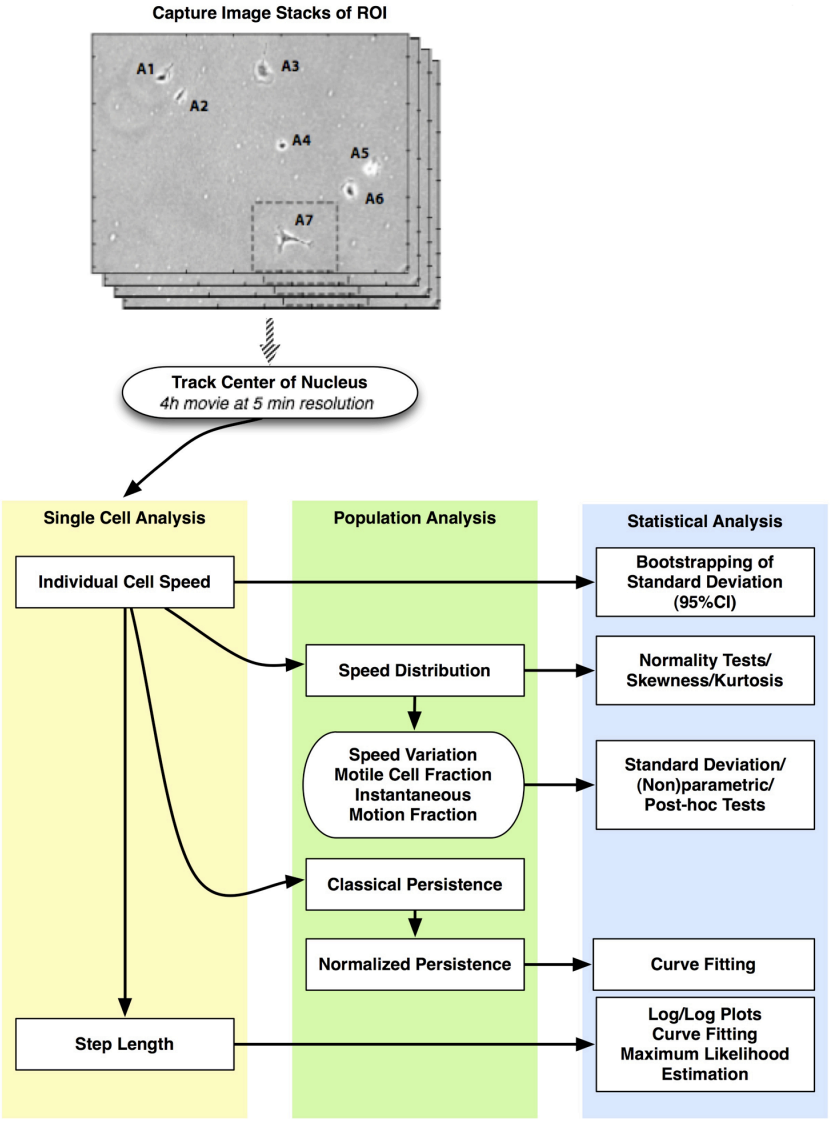



**Figure 3.1 - Measured motility metrics and analyses.** Phase-contrast images were captured automatically at each ROI every 5 min, for 4 h. Image stacks were produced and cell speed was quantified manually by tracking the center of each cell's nucleus. Individual cell speeds, and x, y coordinates were utilized for futher analysis. Data analysis was performed at both the single-cell and population-level. All single-cell speeds were analyzed statistically by bootstrapping of their standard deviations (to achieve 95% confidence intervals). Step-length was calculated as the distance a single cell travels between pauses (two consecutive frames at the same coordinate). The distribution of step-lengths was analyzed by log/log plotting of data and curve fitting. Fits were demonstrated significant using a maximum likelihood estimation method. At the population-level, cell line distributions were assessed using frequency histograms and tests for normality, skewness, and kurtosis. Further, speed variation between media conditions was calculated by standard deviation, (non)parametric tests, and post-hoc tests. The motile cell fraction and instantaneous motion fractions were also calculated from speed data. in addition, non-normalized (Dunn) persistence times were calculated, and persistence values were normalized using the Kipper method. Finally, persistence times were analyzed by curve fitting.

the parental cell line (Dawson *et al.*, 1996), and MCF10A-CA1d (CA1d), a cell line derived from xenograft-passaging the parental line multiple times in nude mice creating a consistently tumorigenic and highly invasive cancer cell line (Santner *et al.*, 2001), were maintained in GIBCO® DMEM/F-12 media (Invitrogen, Carlsbad, CA) supplemented with 5% horse serum (Invitrogen), 0.1 µg/ml cholera toxin (Calbiochem/EMD Biosciences, Gibbstown, NJ), 10 µg/ml insulin (Invitrogen), 0.5 µg/ml hydrocortisone (Sigma, St. Louis, MO), and 20 ng/ml epidermal growth factor (Invitrogen). All cell lines were kept in constant culture in a humidified atmosphere of 5% $CO_2$ at 37°C. The MCF10A cell line was kindly provided by Dr. Joan Brugge (Harvard Medical School, MA) and the AT-1 and CA1d lines were provided by Dr. Fred Miller (Karmanos Institute, MI); all cell lines are readily available through the Vanderbilt Integrative Cancer Biology Center's (VICBC) Tissue Culture Core Unit (http://www.vanderbilt.edu/VICBC/general.html). For all assays performed, cells must be of consistently low passage number for repeatable results.

*3.2.1.2 Single Cell Motility Analysis*

Experiments were performed as previously described in Section 2.4 (see also Harris *et al*., 2008). Briefly, the laminin (Ln) isoform Ln-332 (1 µg/mL; purified in-house) was coated on Nunc™ polystyrene, non-tissue culture treated, 6-well microplate dishes (Cole Parmer, Vernon Hills, IL) for 1 h at room temperature (RT). Dishes were then blocked with 5% milk (Regilait, France) in phosphate buffered saline solution (PBS, Invitrogen) for 1 h at 37ºC. Cell lines were trypsinized (TrypLE Select, Invitrogen, Sunnyvale, CA), neutralized with

growth media (L-15 media supplemented with horse serum, cholera toxin, hydrocortisone, insulin, and EGF), washed three times in PBS, and $2x10^4$ cells per well were resuspended in growth media in 6-well microplates. Cells were allowed to adhere for 1 h within the heated (37ºC) microscope chamber prior to image capture.

Time-lapse microscopy was conducted using a Zeiss Axiovert 200M microscope (Zeiss, Thornwood, NY) equipped with a temperature-controlled chamber and an automated x-y-z stage. Microscopy was under the control of MetaMorph software (Molecular Devices, Sunnyvale, CA). At the beginning of each experiment (0 h), six regions of interest (ROI) were selected at random from within each well for imaging. Each ROI was focused manually, and the coordinates saved using MetaMorph's "Multi-dimensional Acquisition" tool for subsequent imaging. Phase-contrast images were captured automatically at each ROI every 5 min, for 4 h. Following image capture, all 49 images from a particular coordinate were combined using MetaMorph to produce image stacks. Cell speed was quantified manually by tracking the center of each cell's nucleus, using MetaMorph's "Track Points" function.

*3.2.1.3 Cell Speed Data Analysis and Statistics:*

Data analysis was performed using SPSS, version 17 (SPSS, Inc., Chicago, IL). For each cell line (both in the presence and absence of serum/EGF), 95% confidence intervals of individual cell standard deviations were computed to test if cells maintained constant speed. The Shapiro-Wilks W test was applied to all population-based data sets (by individual experiment, and for

pooled datasets) to test each distribution for normality (Gaussian behavior). Skewness coefficients were also calculated for each distribution. Kruskal-Wallis H tests were applied to pooled datasets to detect differences across experiments for each cell line (and medium condition) and Tamhane T2 post-hoc tests were applied to detect pair-wise individual experimental comparisons. The Kolmogorov-Smirnov two-sample non-parametric test was subsequently applied to data to check for significant differences ($P < 0.05$) across various groups (by cell line and microenvironmental conditions) for all measurements. Parameters for each cell were grouped together in analysis. To analyze bin relationships between individual cells, paired Wilcoxon tests were used for all pairs of variables. Cell speed data (unless indicated otherwise) are presented in terms of mean ± standard deviation (with 95% confidence intervals where indicated).

## 3.3 Results

Information garnered from this dataset was processed and analyzed in a number of ways. Figure 3.1 is a flowchart depicting the metrics and analyses used in this study.

### 3.3.1 Single-cells Exhibit Greater Speed Fluctuation in Cancer Cell Lines

Three genetically related breast cell lines, MCF10A, MCF10A-AT1 (AT1), and MCF10A-CA1d (CA1d) were seeded in microplates coated with laminin-332, incubated either in full-serum or serum/EGF-depleted media, and time-lapse high-content microscopy performed as described in Section 3.2.1.2. As shown in Figure 3.2A, individual cells from all three cell lines, and from both medium

conditions, displayed step-pause-step motion, which is entirely consistent with previous reports on epithelial cell motility (Potdar *et al.*, 2009). However, individual cells carried out steps at highly fluctuating, non-constant speed over time (4 h, at a 5 min sampling interval) (Figure 3.2A). Four cells from each population are presented, although approximately 500/line were tracked cumulatively. With the obvious exception of non-motile cells (cells that moved less than a cell diameter over the course of the experiment), nearly all cells exhibited speed that fluctuated widely across the 49 frames (Figure 3.2A). Due to the great difference in fluctuations between individual cell speed profiles, we applied stringent bootstrapping test to estimate 95% confidence intervals for individual cell standard deviations for each dataset (Table 3.1). This test is robust to irregularly distributed data. These calculated confidence intervals strongly support the fluctuation of speed within single cells and, of note, they never include zero, further indicating that cells consistently adopt variable speeds in consecutive frames. Both step-pause-step motion and fluctuating speed were observed in serum/EGF-depleted culture media as well (Figure 3.2A), suggesting they are intrinsic cell properties.

**Table 3.1 - Individual cell speeds fluctuate highly and are non-constant.** 95% confidence intervals for individual cell standard deviations were obtained from bootstrapping. Note that confidence intervals do not include zero, demonstrating that cells are not maintaining constant speed during our observations.

| Cell Line | Medium | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 | Experiment 5 | Pooled |
|---|---|---|---|---|---|---|---|
| MCF10A | Full | 0.115 to 0.548 | 0.085 to 0.553 | 0.077 to 0.781 | 0.030 to 0.423 | 0.075 to 0.893 | 0.035 to 0.893 |
| | Depleted | 0.239 to 0.699 | 0.043 to 0.399 | 0.059 to 0.636 | 0.040 to 0.458 | 0.063 to 0.568 | 0.036 to 0.699 |
| AT1 | Full | 0.096 to 0.607 | 0.060 to 0.814 | 0.055 to 0.565 | 0.039 to 0.232 | 0.130 to 0.669 | 0.046 to 0.814 |
| | Depleted | 0.212 to .700 | 0.042 to 0.389 | 0.043 to 0.440 | 0.047 to 0.150 | 0.179 to 0.666 | 0.036 to 0.700 |
| CA1d | Full | 0.063 to .883 | 0.078 to 0.854 | 0.056 to 1.395 | 0.064 to 0.511 | 0.052 to 1.026 | 0.042 to 1.395 |
| | Depleted | 0.181 to 1.162 | 0.062 to 0.900 | 0.044 to 0.616 | 0.032 to 0.395 | 0.269 to 1.265 | 0.035 to 1.265 |

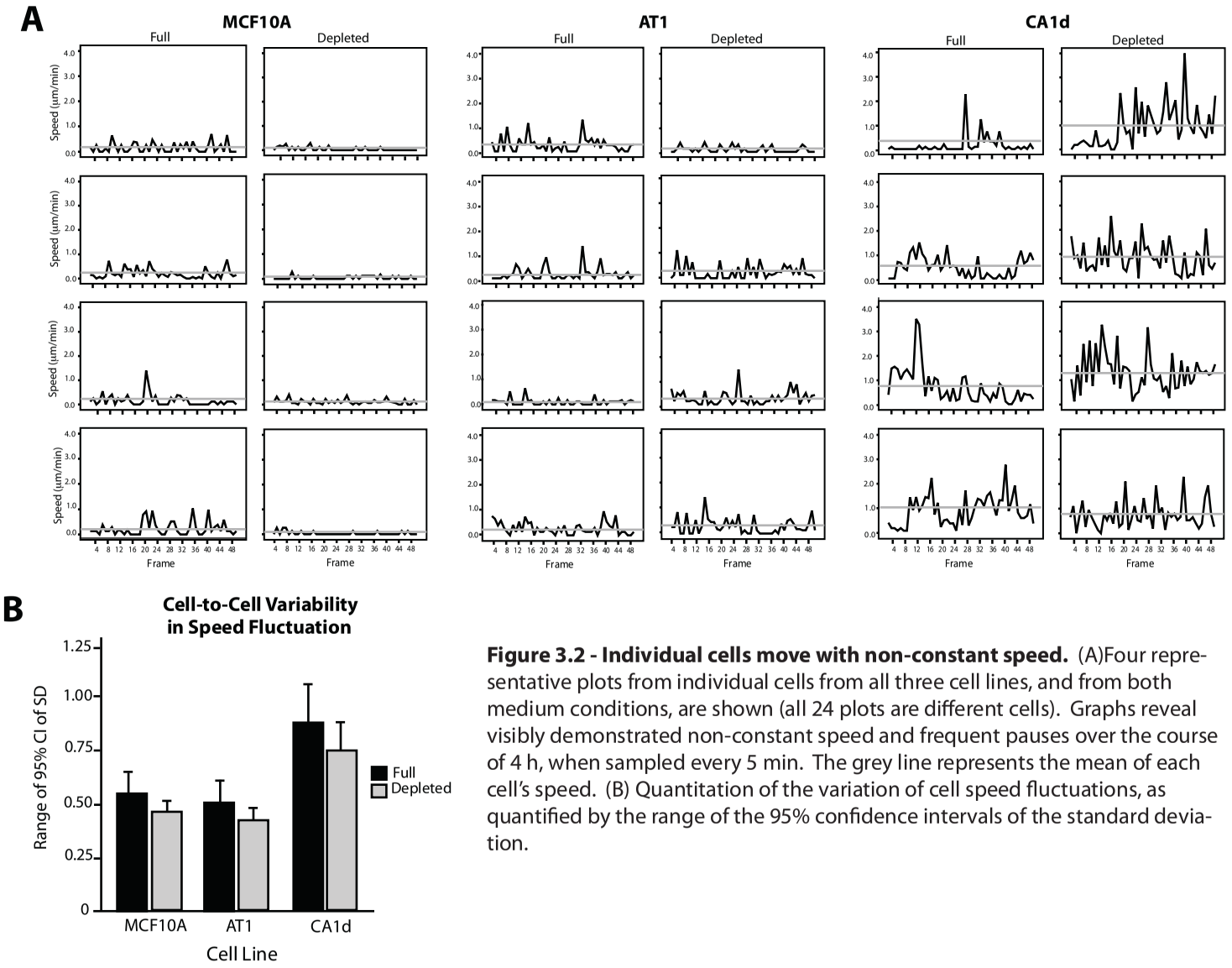



**Figure 3.2 - Individual cells move with non-constant speed.** (A)Four representative plots from individual cells from all three cell lines, and from both medium conditions, are shown (all 24 plots are different cells). Graphs reveal visibly demonstrated non-constant speed and frequent pauses over the course of 4 h, when sampled every 5 min. The grey line represents the mean of each cell's speed. (B) Quantitation of the variation of cell speed fluctuations, as quantified by the range of the 95% confidence intervals of the standard deviation.

### 3.3.2 Cell-to-cell Variability of Speed is Greater in Cancer Cell Lines

As demonstrated in Figure 3.2A, speed is variable within the cell lines we examined from single-cell to single-cell. Furthermore, the spread of speed variability is itself variable: it is broader in the aggressive cell line, and becomes even broader in serum/EGF-depleted media. In contrast, the speed variability spread appears to be dampened in the non-tumorigenic cell line.

Since these findings legitimate a comparison between cell lines, we calculated the ranges of the 95% confidence intervals. As shown in Figure 3.2B, cells within the most invasive cancer cell line, CA1d, display greater variability of speed fluctuations than MCF10A (non-tumorigenic) and AT1. Further analyses with this metric, single-cell speed fluctuation within a cell line (Table 3.1), should be interesting to pursue in a larger panel of cancer cell lines.

### 3.3.3 Population-level Speed Heterogeneity is Greater in Cancer Cell Lines

Having demonstrated that individual cell speeds fluctuate and are non-constant, we examined the distribution of cell speeds for cell lines at the population level. To simplify analyses and avoid confounding factors, we represented each cell by a single value. Individual cell speeds were therefore calculated by averaging each cell's speed across a time-lapse movie (N=49 frames) (grey horizontal lines in Figure 3.2A).

Figure 3.3B shows box-and-whisker plots for five pooled experiments, for all cell lines in both medium conditions; points, color-coded by experiment, represent individual average cell speeds. The spread of these points visualized

the cell-to-cell variability of speed within cell lines. However, because of high day-to-day variation (as is often observed with cell line motility assays), it was not possible to draw statistical conclusions from these pooled data. Therefore we analyzed cell-to-cell variability within single experiments. Box and whisker plots of individual experiments are shown in Figure 3.3A. The dark grey horizontal line within each box represents the mean speed for all pooled cells within a cell line.

Comparison of the mean values masks underlying cell line heterogeneity. Instead, examining the scattering of the data points, including “outliers”, uncovers some interesting trends. For instance, the speed scatter is broader for the most aggressive cancer cell line, CA1d. This cell line also appears to contain the larger amount of outliers, which may explain some of its aggressive traits, at least in *in vitro* assays.

### 3.3.4 Day-to-day Speed Variability is Greater in Cancer Cell Lines

As displayed through scatter plots in Figure 3.3B, we saw a large amount of cell speed variability across individual experiments (CV > 0.3 for all cell lines and conditions). However, cells were handled in a uniform fashion during preparation for individual experiments (e.g., passage number, confluence-state, micro-centrifuge settings, trypsinization time). Interestingly, levels of variability seemed to be linked to cell lines. Kruskal-Wallis tests confirmed that data from different experiments are significantly different ($P < 0.01$, in all cases). More interestingly, post-hoc Tamhame T2 pair-wise tests further revealed that only a single experiment (4) for MCF10A in full media varied

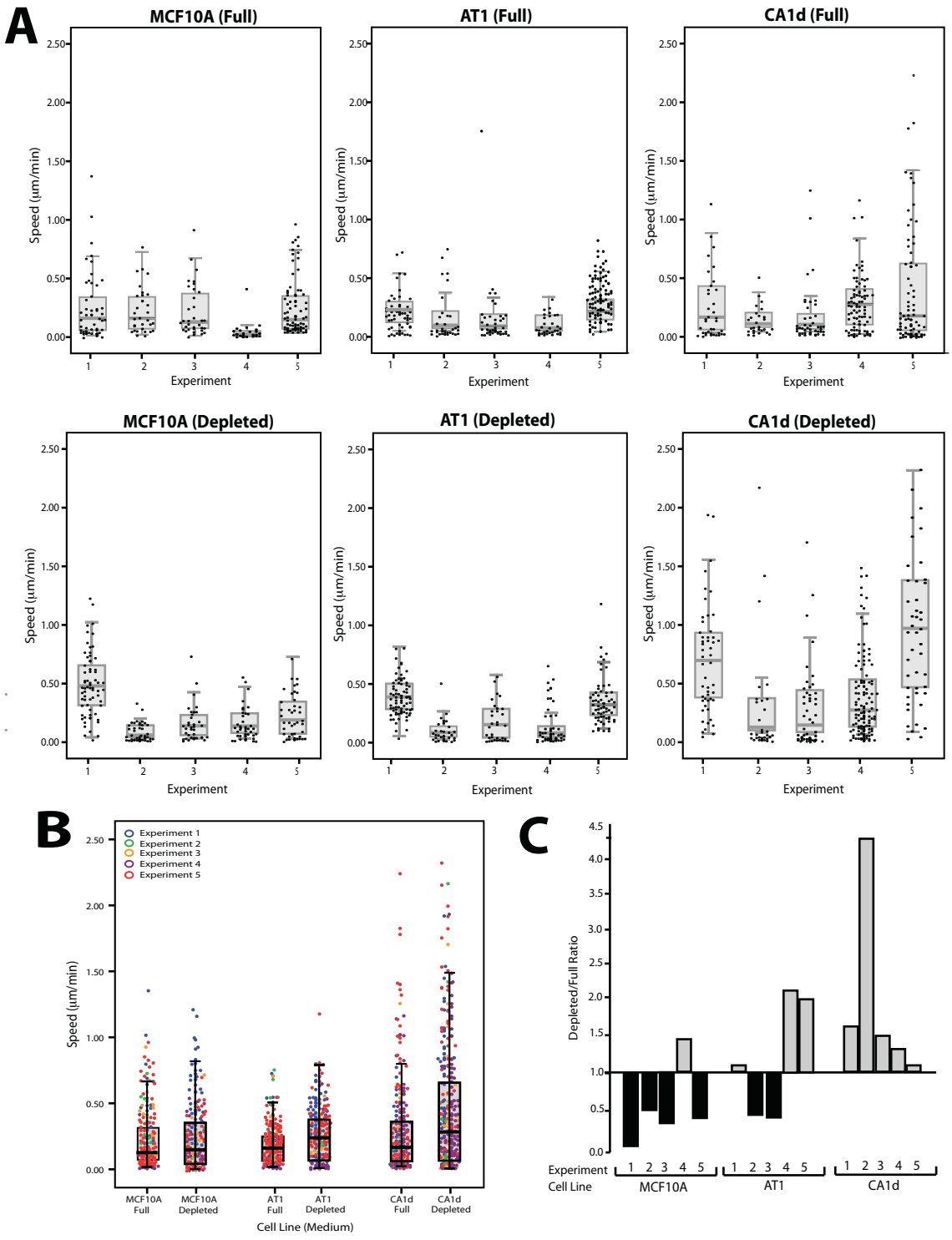



**Figure 3.3 - Cell-to-cell variability of speed is greater in cancer cell lines.** (A) Box-and-whisker plots for individual experiments. Individual cells are represented by scattered points, the box outlines the 25th and 75th quartiles, the horizontal line repreesents the mean, and the whiskers represent the 95% confidence intervals. (B) Box-and-whisker plots for pooled experiments. Points are color coded by experiment. (C)A ratio was calculated to compare the change in population-level speed heterogeneity between full and serum/EGF-depleted media conditions for all three cell lines within experiments. A ratio > 1 indicates that cells demonstrated greater speed heterogeneity in serum/EGF-depleted media.

from all other experiments performed for that cell line and condition (data not shown). In contrast, both AT1 and CA1d in full medium revealed much more day-to-day variability. Of note, all cell types also displayed more day-to-day variability in serum/EGF-depleted medium than they did in their corresponding experiments in full medium.

### 3.3.5 Cancer Cell Speed Variability Increases in Serum/EGF-Depleted Media

To examine possible effects of serum/EGF-depletion on cell speed, a ratio was created between the cell speed range (*i.e.*, variation measured for cell speed) in serum/EGF-depleted and full media, for individual experiments. The resulting ratio reflects the effect of the microenvironment on cells within the same experiment (Figure 3.3C). In all experiments performed, CA1d cells demonstrated a ratio > 1, indicating that this cell line has increased cell-to-cell variability in serum/EGF-depleted media. In contrast, MCF10A cells displayed a ratio < 1 in 4 out of 5 experiments performed, indicating that in this non-tumorigenic cell line cell-to-cell variability of motility decreased in serum/EGF-depleted media. AT1 cells showed an intermediate effect. Similar results (with identical trends) were seen when 95% confidence intervals were used to create the ratio, rather than the range. All cell tracks are displayed as windrose plots in Figure 3.4 to demonstrate this effect.

### 3.3.6 Cloned Cancer Cells Maintain Speed Variability and Response to Serum/EGF-Depleted Media

To eliminate the possibility of genetic heterogeneity influencing variability of migration speeds, we produced clones of CA1d cells. We have analyzed five CA1d clones derived from individual cells *via* serial dilution. If cell speed is genetically pre-determined, one would expect specific observations from analyzing cell speeds of cloned cells: 1) Cell speed variation would be substantially reduced, 2) Various clones would produce different mean cell speeds, and 3) The cell speed ratio (from full to serum/EGF-depleted media) would not be consistently greater than one. To this end, two additional single-cell migration experiments were conducted to produce preliminary data to address these questions (refer to Section 3.1.1 for method). These points will be addressed individually:

1. As seen in Figure 3.5, the cell speed variation of the clones appears to decrease in full media. However, this trend is obvious only in the pooled dataset. In one experiment the parental CA1d cells in full media had a high level of variation, while in the other experiment the parental cell line had similar speed variation as the clones. This experiment needs to be repeated to determine with confidence that CA1d cells maintain their speed variation even upon cloning.
2. Differing mean speeds of the clones were not observed, decreasing the chance that a clonal subpopulation with increased migratory ability exists.

3. All five CA1d clones had a cell speed variability ratio greater than one, suggesting that the increase in speed heterogeneity in serum/EGF-depleted media could not be entirely due to genetic heterogeneity.

In total, preliminary results from the cloning experiment suggest that clonal selection does not eliminate cell speed heterogeneity, or response to serum/EGF-depletion. However, these findings must be validated through additional experimentation.

### 3.4 Significance / Discussion

The study of cell variability at the single-cell level has accelerated in recent years, as computing power, storage capacity, and access to high-content automated microscopy and image processing software has increased (Bullen, 2008; Glory and Murphy, 2007; Pepperkok *et al.*, 2006). It is now possible to obtain hundreds of thousands of images of cells, in a multitude of conditions dynamically over time. With the application of these new techniques, we can begin to answer questions about the nature of phenotypic variability, and how cells alter their variability in response to microenvironmental changes.

In this study, we have compared 3 genetically related cancer cell lines (with increasing invasive potential) with high-resolution motility assays, in order to better understand the similarities and differences between cancer and non-tumorigenic cell motility. In summary, we examined the movement of over 1,500 cells at the individual- and population- levels, to examine two hypotheses:

- *Hypothesis 1: a non-tumorigenic cell line will exhibit less variability of motility than a cancer cell line.*
- *Hypothesis 2: the removal of serum and mitogenic factors (EGF) will decrease the variability of motility of a non-tumorigenic cell line, and will not affect the variability of a cancer cell line.*

The analysis was focused on investigation of individual cell speed, in the presence or absence of key nutrients (serum/EGF). The metrics examined included: individual cell speed fluctuation, variability of speed, population-level speed heterogeneity, day-to-day speed variability, and speed variation in response to serum/EGF-depleted media. Through this analysis, evidence was provided that supports both hypotheses in our cell line panel, and further indicates that the most aggressive cell line tested (CA1d) actually increases variability of speed in serum/EGF-depleted media.

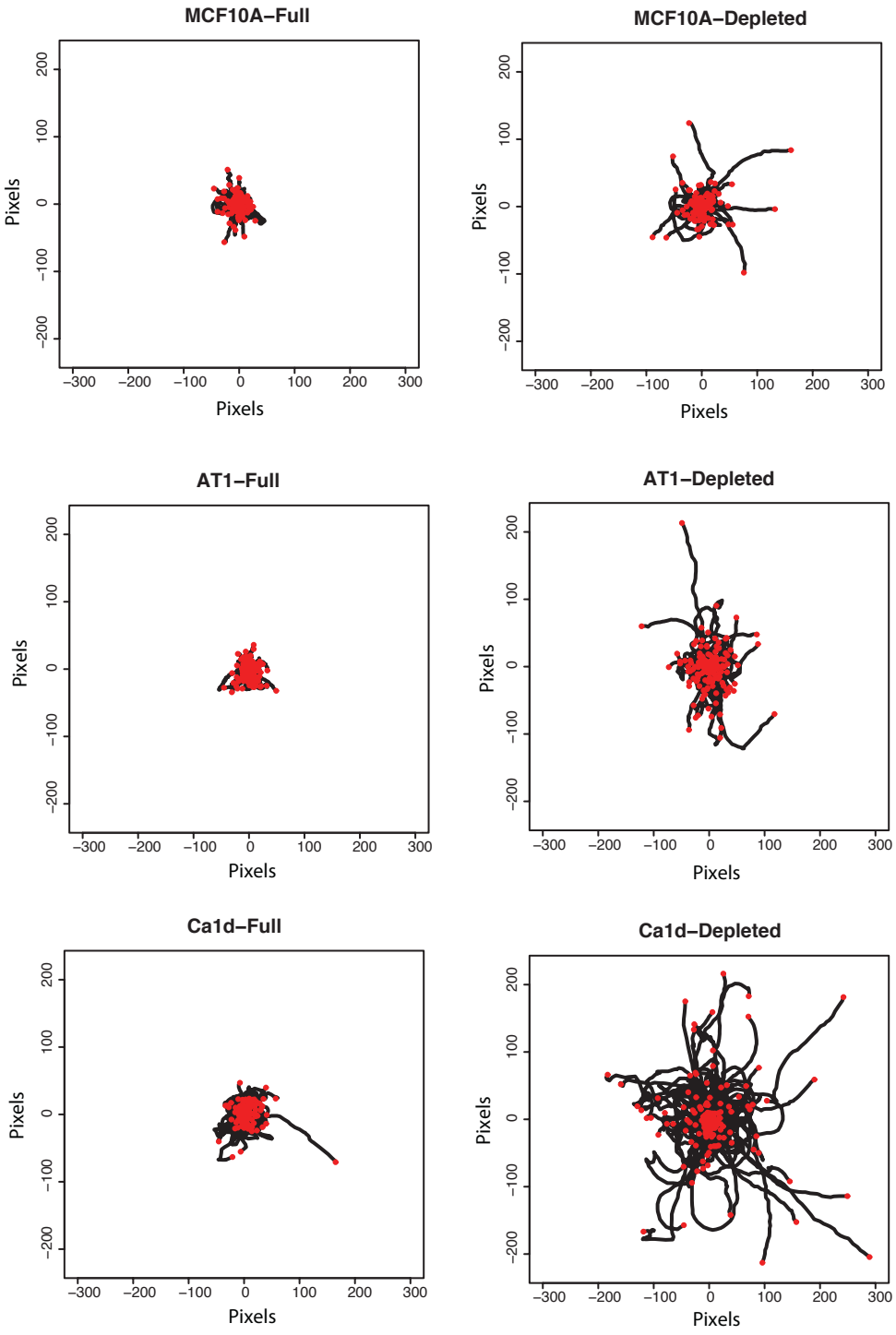



**Figure 3.4 - Cell Tracks.** Cell paths of MCF10A, AT1, and CA1d cells migrating on Ln-332, and in full or serum/EGF-depleted media. Paths are of all cells (1,500+) tracked over a period of 4 h and are replotted such that all paths start from the origin. Red dots indicate the stopping point of each cell at the end of the movie.

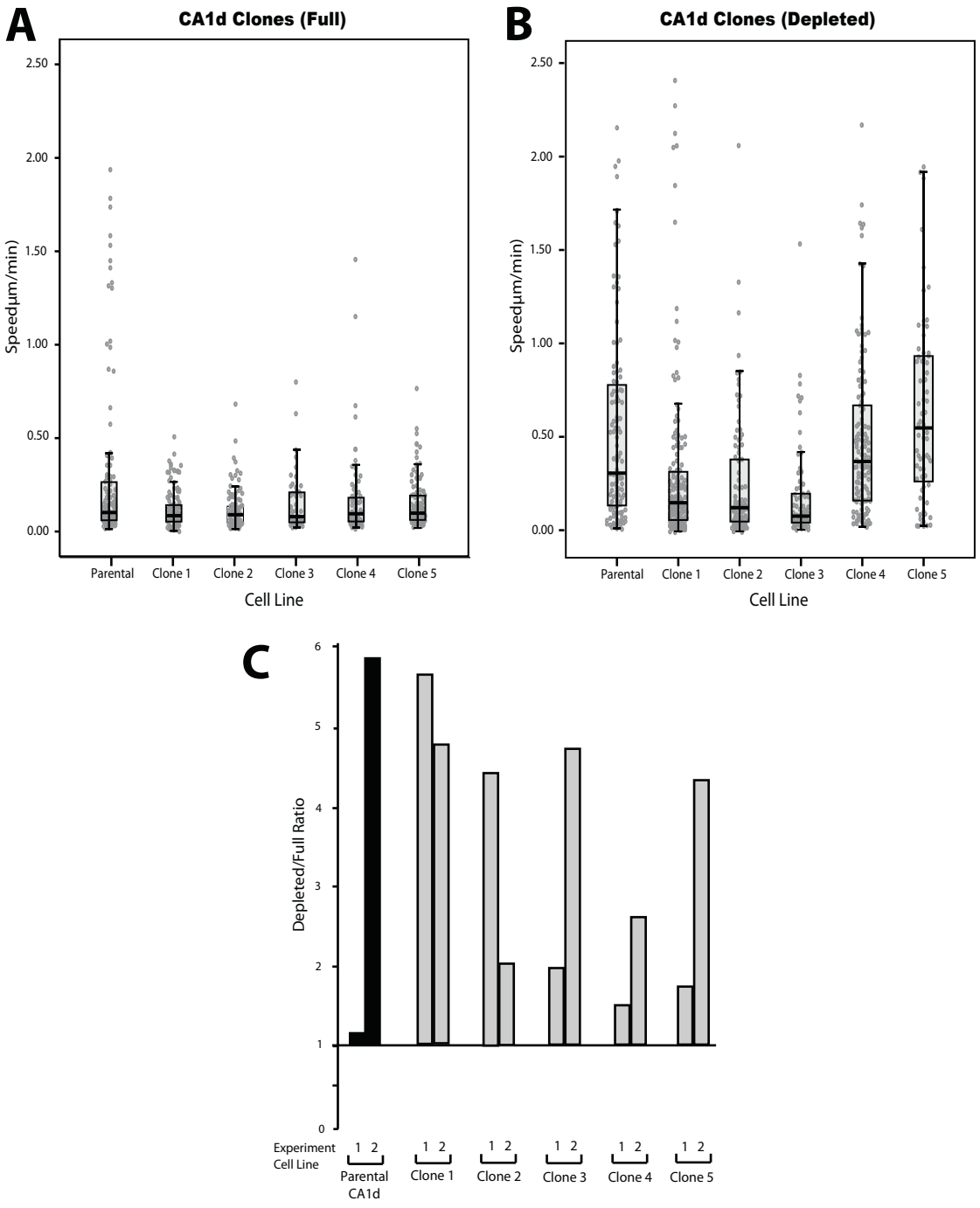



**Figure 3.5 - CA1d clones exhibit similar cell speed variability as the parental line.** (A) Box-and-whisker plots for pooled experiments in full media. Individual cells are represented by scattered points, the box outlines the 25th and 75th quartiles, the horizontal line repreesents the mean, and the whiskers represent the 95% confidence intervals. (B) Box-and-whisker plots for pooled experiments in serum/EGF-depleted media.(C) A ratio was calculated to compare the change in population-level speed heterogeneity between full and serum/EGF-depleted media conditions for all CA1d clones within experiments. A ratio > 1 indicates that cells demonstrated greater speed heterogeneity in serum/EGF-depleted media.

The MCF10A model system was designed to mimic tumor progression. In this system, the parental MCF10A cells' speed variability decreased for all metrics tested. The Ras-transformed and tumorigenic AT1 cells' speed variability showed a mixed phenotype. The variability increased for some experiments and metrics, and decreased for others. The most aggressive CA1d cells, which were passaged several times *in vivo*, were shown to exhibit increased cell speed variability for all metrics tested. Thus it appears that, in this model, increased speed variability was selected for during cancer progression. These findings need to be validated in other cell lines and model systems. However, it is tempting to hypothesize that the variability of other traits may also be increased during cancer progression, leading to a highly heterogeneous cancer cell population, which could increase the adaptive potential of a cancer.

# CHAPTER IV

# DESIGN OF ADDITIONAL METRICS TO QUANTITATE MOTILITY

## 4.1 Background

A migrating cell achieves motion by protruding its leading edge and retracting its rear, resulting in a directional displacement. This method of motility is a complex and dynamic mechanism of focal adhesions, actin polymerization, and dozens of protein-protein interactions and signaling cascades (see Section 2.1.3). However, the most widely used model to describe cellular motion is based on particle physics, and is termed a persistent random walk (PRW; Codling *et al.*, 2008). Mathematical models of cell migration are used to generate hypotheses, test parameters, and advance our understanding of cell motility. These models are one of the main tools of systems biology (see Section 5.1).

A random walk, also known as Brownian motion, was first reported in the movement of pollen in solution by Brown in 1828, and was subsequently turned into the random walk theory (Uhlenbeck and Ornstein 1930). Cells follow a modified version of this random walk, known as a PRW. Cellular movement involves a correlation between successive step orientations (Codling *et al.*, 2008), creating a local directional bias. Currently, PRW is the most widely used model to describe cellular motion (Codling *et al.*, 2008). To produce a PRW model of cell motion, the input usually required is turn angle distribution

(directional bias), and cell speed. From the same data, one can compute the mean squared displacement, diffusion coefficient, and persistence time. These values are used to create the PRW model for that particular cell type and condition. One of the main assumptions made in the PRW model is that the cells are always in motion. In Chapter 3 we saw that cells do not follow this basic assumption of the PRW model. Our data demonstrates that cells pause frequently as they migrate (Figure 3.2). In order to refine this model of cell migration, specifically to include cell pausing, one must accurately determine how many cells are paused at any given time, and how long they travel between pauses.

To address these questions we analyzed three population level metrics: distribution of step-lengths, persistence and moving-to-paused cell ratio. These metrics were chosen to more accurately quantify cell motion, with a goal of refining our models of cell migration to a smaller scale. These metrics can accurately quantitate the dynamics of cell pauses and allow easy integration of tracked cell data into mathematical models of cell motility. Here we present two new metrics (step-length distribution and IMF) and demonstrate how they could be used to refine mathematical models of cell migration, focusing on accurately representing cell movement at the single-cell level based on experimental findings.

## 4.2 Experimental Design

This chapter utilizes the same dataset as in Chapter 3 (see Section 3.2).

### 4.2.1 Persistence

One of the most common measurements of cell motility is persistence time, which assumes that cell motion can be described by a PRW. Within the PRW model, persistence time is defined as the persistence in velocity or motion, since it is a combination of persistence in direction and speed (Dunn and Brown, 1987). The PRW model can be derived from the Langevin equations as described by the Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein 1930). The form of the model takes the shape as described by the Furth equation (Furth, 1920). This equation describes the expected mean squared displacement over time. Initially the motion is super-diffusive, meaning that a ballistic component dominates and the mean squared displacement increases exponentially. This motion then transitions over to a diffusive regime for times far greater than the persistent time (Codling *et al.*, 2008). Thus, to calculate persistence time, one must observe cells for a long enough time interval for them to transition from a ballistic to diffusive movement regime (roughly 3 hours for a 10 min persistence time).

### 4.2.2 Step-length

To accurately add cell pausing into cell migration models it is necessary to experimentally determine the distance a cell travels between consecutive pauses. Step-length, flight length, and flight time are three metrics which are used in ecology to study foraging behavior of birds, bees, and mammals (Viswanathan *et al.*, 1999; Gautestad and Mysterud, 2005). The term step-length has also been used to describe the movement of molecular motors on polymers (Wallin *et al.*, 2007). All three terms are used to quantitate distance or time between pauses in motion, but we are not aware of a previous use of these metrics to quantify the motion of epithelial cells.

### 4.2.3. IMF

Persistence and diffusion coefficients are often used to describe cellular motion. However, both of these representations make a number of assumptions about cellular behavior. In particular, they assume all cells are in motion at all time. The instantaneous motion fraction (IMF) was developed to test this assumption, and to provide an additional metric to monitor differences in migration characteristics between cell lines and conditions. It measures the percent of cells moving at each time point, and thereby also quantifies the fraction of cells paused at any given moment.

### 4.2.4 Materials and Methods

Cell Culture, Cell Preparation, and Time-lapse Microscopy were performed as described in Section 3.2.1.

*4.2.4.1 Persistence Quantitation*

We calculated persistence times by both the traditional Dunn method (as described in Dunn and Brown, 1987), and the updated Kipper method (as described in Kipper *et al.*, 2007). The Kipper method reduces standard error of data to fit by ~50% over the classical Dunn method, and thus persistence times reported in this dissertation were calculated using the Kipper method. For comparison, graphs mean squared displacement *versus* time are shown for both methods (Figure 4.3).

*4.2.4.2 Step-length Quantitation*

We also measured the overall distance traveled between cell pauses in a movie (defined by two consecutive frames at the same coordinate), and discarded all step-lengths (distances between pauses) below our tracking error threshold (lengths < 1 μm). This data was binned in log $2^k$ bins and plotted log/log as previously described (Sims *et al.*, 2008) to determine if the data followed a power-law distribution. In addition, data was plotted directly onto a Pareto distribution (a specific type of power-law distribution) to avoid possible problems with log/log binning (Edwards *et al.*, 2007).

To verify the fits to a power-law distribution, we applied the MLE method, as previously described by Edwards (Edwards *et al.*, 2007). Importantly, this method does not bin the data, but instead uses the raw data, which eliminates false positives from log-log binning. Furthermore, Edwards' approach offers an alternative fit to an exponential distribution, and compares the fits via Akaike weights to see which distribution is a better fit. Thus, the MLE method does not

assume that the data fits a power-law distribution, but instead compares two distributions, and determines which is the best fit.

*4.2.4.3 IMF Quantitation*

IMF was computed as the percentage of cells moving more than one pixel (our measurement error threshold) at each sampling interval.

## 4.3 Results

### 4.3.1 Highly Motile Cells are a Minority

Data from Chapter 3 were further analyzed in terms of their distribution. Individual cell mean speeds were binned (Section 3.2.1.3) and represented as a population frequency histogram. Plots for each cell line (pooled experiments) in both media are shown in Figure 4.1 (individual experiments in Figure 4.2). The plots show that the binned data do not fit with a normal distribution (overlaid in Figure 4.1) neither for pooled nor individual datasets. The mean speed, standard deviation, and sample size (N) for each population is also shown.

In almost all instances, distributions were positively skewed, indicating that individual cell speeds were more concentrated to the left in the plots (Table 4.2). This indicates that, in all cell lines, including the most aggressive ones, a majority of cells are slower moving or non-motile (this statement is true for all ECM substrates tested, see Table 2.2). This information is generally lost in classical migration assays (i.e. Boyden or scratch), because they do not typically examine or quantify non-motile cells. Nonetheless, it may be especially relevant when

establishing correlations between degree of invasiveness of a cancer cell line and its motility properties. Our data suggest that a few cells, located in the high-speed tails, in these aggressive cell lines may be responsible for invasion. This possibility needs to be investigated in more detail.

**Table 4.1 - Cell lines populations are non-normal and positively skewed.** Shapiro-Wilks W tests revealed that, in almost all cases, experimental distributions (i.e., for all cell lines, in both conditions) were found to be non-normal ($P<0.05$). Interestingly, in almost all instances, distributions were positively skewed, indicating that individual cell speeds were more concentrated to the left of the mean in the plots. Kurtosis values reflect the "peakedness" of the distributions; higher values reflect that more variance is due to infrequent extreme deviations (outliers), as opposed to modest deviations.

| Cell Line | Medium | Statistic | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 | Experiment 5 | Pooled |
|---|---|---|---|---|---|---|---|---|
| MCF10A | Full | P | <0.001 | 0.003 | <0.001 | <0.001 | <0.001 | <0.001 |
| | | Skewness | 1.954 | 1.054 | 1.475 | 4.261 | 1.427 | 1.772 |
| | | Kurtosis | 4.391 | 0.547 | 1.867 | 20.169 | 1.088 | 3.358 |
| | Depleted | P | 0.104 | <0.001 | <0.001 | <0.001 | 0.002 | <0.001 |
| | | Skewness | 0.579 | 1.436 | 2.031 | 1.469 | 0.774 | 1.434 |
| | | Kurtosis | 0.137 | 1.927 | 5.238 | 1.81 | -0.214 | 1.924 |
| AT1 | Full | P | <0.001 | <0.001 | <0.001 | <0.001 | 0.027 | <0.001 |
| | | Skewness | 1.381 | 1.845 | 2.397 | 1.179 | 0.527 | 1.376 |
| | | Kurtosis | 2.297 | 2.325 | 7.088 | 0.470 | -0.190 | 2.236 |
| | Depleted | P | 0.106 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | | Skewness | 0.537 | 2.575 | 0.923 | 1.797 | 1.729 | 0.890 |
| | | Kurtosis | 0.222 | 8.918 | -0.090 | 2.220 | 4.340 | 1.000 |
| CA1d | Full | P | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | | Skewness | 1.497 | 1.561 | 3.056 | 1.375 | 1.771 | 2.389 |
| | | Kurtosis | 1.775 | 2.626 | 10.174 | 2.551 | 2.782 | 7.169 |
| | Depleted | P | 0.011 | <0.001 | <0.001 | <0.001 | 0.107 | <0.001 |
| | | Skewness | 0.789 | 3.072 | 2.219 | 1.475 | 0.481 | 1.457 |
| | | Kurtosis | 1.076 | 10.000 | 5.898 | 1.745 | -0.513 | 1.810 |

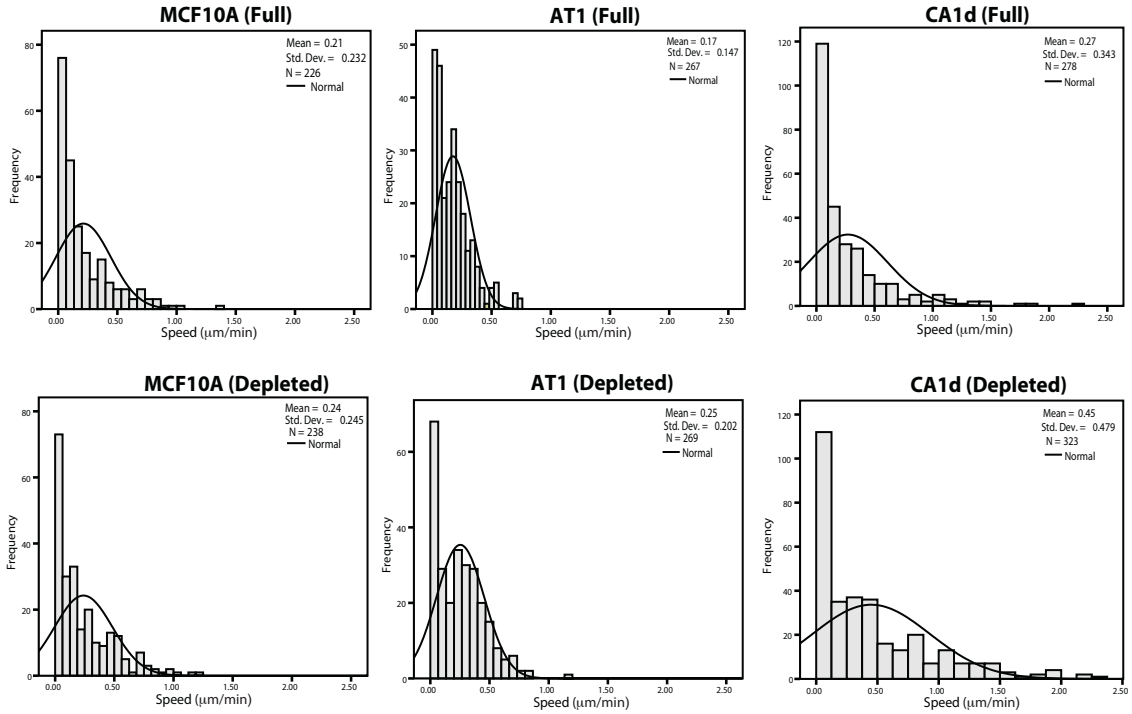



**Figure 4.1 - Cell speed is non-normally distributed.** Indvidual cell speeds were calculated by averaging each cell's speed across a movie (N=49), for further analysis at the population-level. Each cell is represented within a population histogram. Histograms for all individual experiments are shown in Figure 4.2. All graphs include bars that represent the number of cells (frequency) at each speed, and the normal (Gaussian) fit for each set of data (based around the average). It is clear from the overlaid plots that the actual data do not fit well with the projected normal fits, neither for pooled nor individual experimental datasets. Shapiro-Wilks W tests confirmed that, in almost all cases, experimental distributions were found to be non-normal.

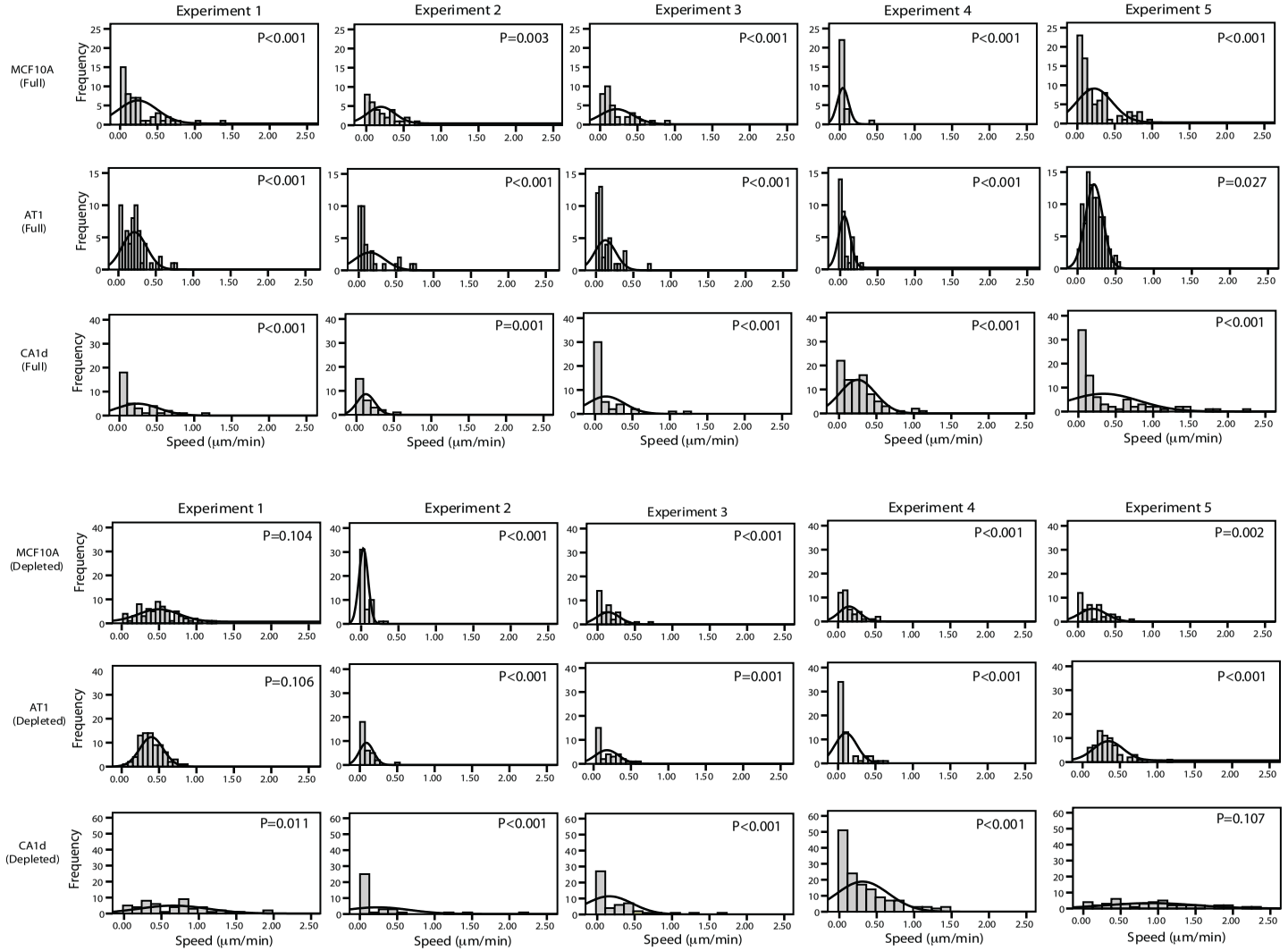



**Figure 4.2 - Cell speed is non-normally distributed (by experiment).** Frequency histograms of cell speed for every experiment. P values represent the probability of each distribution of being normally distributed (Shapiro-Wilks test, non-normal if $p < 0.05$). For skewness and kurtosis, see Table 4.1.

### 4.3.2 Persistence Increases for all Cell Lines in Depleted Media

A well-established measure of motility in a cell line is *persistence*, which assumes a PRW model. In PRW, persistence is defined as the length of time a particle (here, a cell) remains "*persistent in velocity or motion, or simply persistence in motion, since it is a combination of persistence in direction and persistence in speed*" (Dunn and Brown, 1987). Calculation of this parameter reveals important information about cell movement, which is overlooked by representing only speed. Graphs of the classical, non-normalized (Dunn) migration persistence parameter are shown in Figure 4.3A. However, since this method is associated with a high level of error, we also applied an updated method that can reduce error by as much as 50% (Kipper et al., 2007, see also Section 4.2.4.1). Graphs of normalized persistence by experiment are shown in Figure 4.3B, calculated persistence time (P (min)) values are shown graphically in Figure 4.3C, and raw data for Kipper persistence time is shown in Table 4.2. These results indicate that all three cell lines increase their persistence in the absence of serum/EGF. These findings compare well with previous research by Lauffenburger's group and colleagues, who found that in two-dimensions, cell persistence increased when EGF was removed (Kim *et al.*, 2008). This concordance of results shows that our data collection strategy is adequate to be fitted with standard PRW models of motility. However, since PRW persistence time did not distinguish between non-cancer and cancer cell lines, we pursued additional population level metrics.

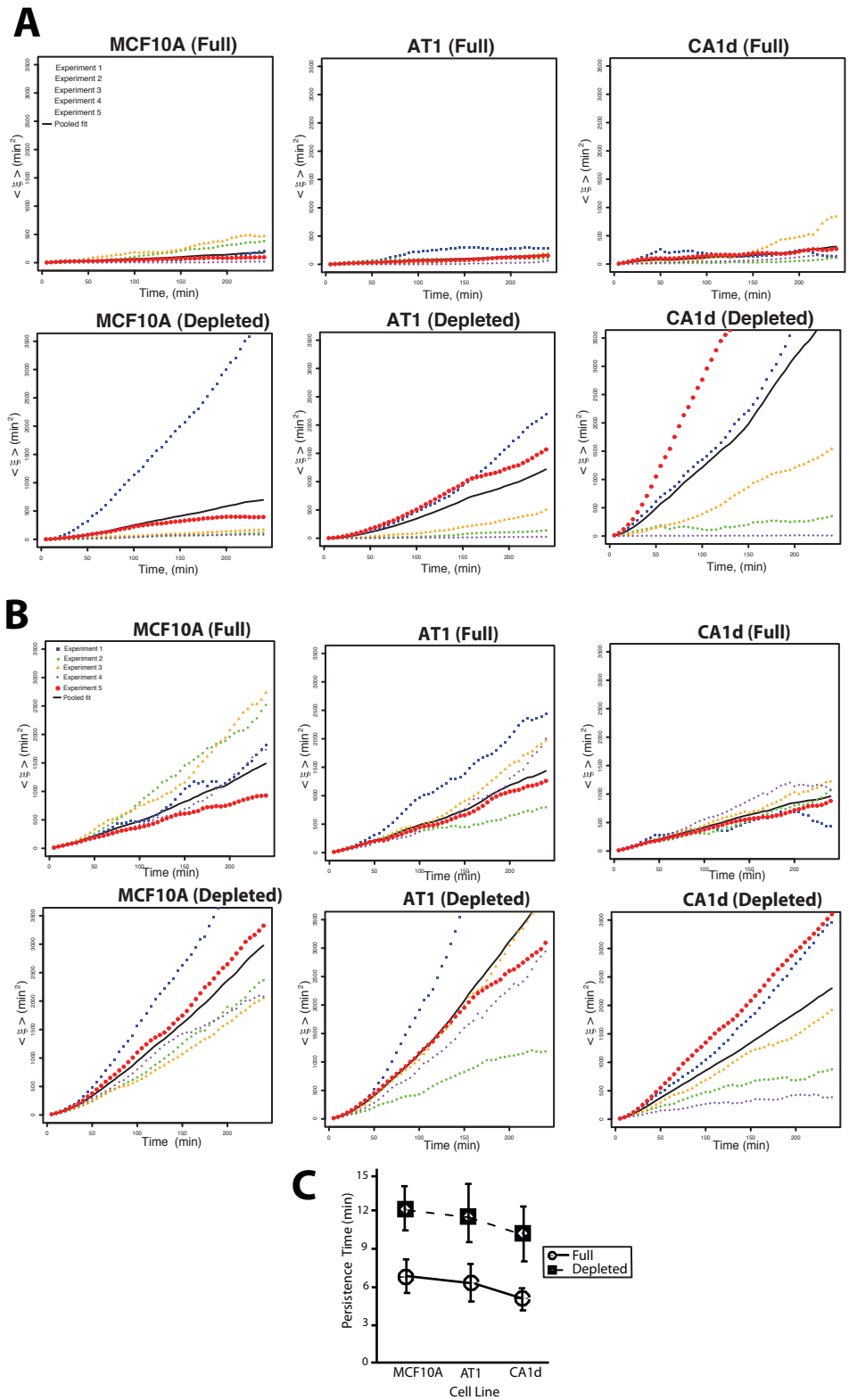



**Figure 4.3 - Persistence time increases for all cell lines in serum/EGF-depleted media.** Cellular persistence was calculated both by the traditional Dunn method (A), and by the updated Kipper method (B). (C) Graphical representation of pooled persistence times from Kipper method (for persistence by experiment, see Table 4.2). These results indicate that all three cell lines increase their persistence in the absence of seum and EGF.

**Table 4.2 – Persistence time calculated using Kipper method.** Curve fits for persistence time (P (min)) and standard deviation are shown. Persistence time was not able to be calculated for Experiment 1 of AT1 in serum/EGF-depleted media, and thus was omitted. These results indicate that all three cell lines increase their persistence in the absence of serum and EGF. These findings compare well with previous research performed independently by another group.

| Cell Line | Medium | Statistic | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 | Experiment 5 | Pooled |
|---|---|---|---|---|---|---|---|---|
| MCF10A | Full | P (min) | 6.660 | 10.200 | 10.290 | 5.940 | 4.040 | 5.890 |
| | | P (SD) | 0.788 | 0.757 | 1.539 | 1.186 | 0.113 | 0.395 |
| | Depleted | P (min) | 21.890 | 9.310 | 8.100 | 9.420 | 13.920 | 12.280 |
| | | P (SD) | 1.760 | 0.998 | 0.774 | 0.528 | 0.994 | 0.919 |
| AT1 | Full | P (min) | 10.580 | 3.380 | 7.320 | 6.370 | 5.120 | 5.750 |
| | | P (SD) | 0.565 | 0.384 | 1.158 | 1.431 | 0.433 | 0.474 |
| | Depleted | P (min) | N/A | 5.500 | 15.930 | 11.770 | 13.850 | 16.420 |
| | | P (SD) | N/A | 0.370 | 1.713 | 1.029 | 0.462 | 1.412 |
| CA1d | Full | P (min) | 3.150 | 3.930 | 5.006 | 5.720 | 3.710 | 4.260 |
| | | P (SD) | 0.804 | 0.351 | 0.198 | 0.553 | 0.161 | 0.125 |
| | Depleted | P (min) | 14.300 | 4.010 | 7.930 | 2.080 | 15.920 | 9.760 |
| | | P (SD) | 1.138 | 0.497 | 0.323 | 0.307 | 0.245 | 0.204 |

### 4.3.3 Step-lengths Appear to Follow a Pareto Distribution

We observed that cells often paused (same coordinates in two consecutive frames, Figure 4.4A), and the distances traveled between pauses (step-lengths) appeared to be highly variable (Figure 3.2). However, a standard log/log binning method revealed a distribution of step-lengths with an underlying pattern, consistent with a power-law (Figure 4.4B). The significance of this fit was demonstrated independently using a maximum likelihood estimation (MLE) method, as detailed in Section 4.2.4.2. The MLE was explicitly used to compare fitting of an exponential distribution *versus* a Pareto distribution (a particular implementation of a power-law distribution). The MLE strongly favored the Pareto distribution for all cell lines under full media conditions (Figure 4.5A) and MCF10A in serum/EGF-depleted conditions. These findings are summarized in bar-graph form in Figure 4.5B (pooled Akaike weight confidence; full analysis in Table 4.3).

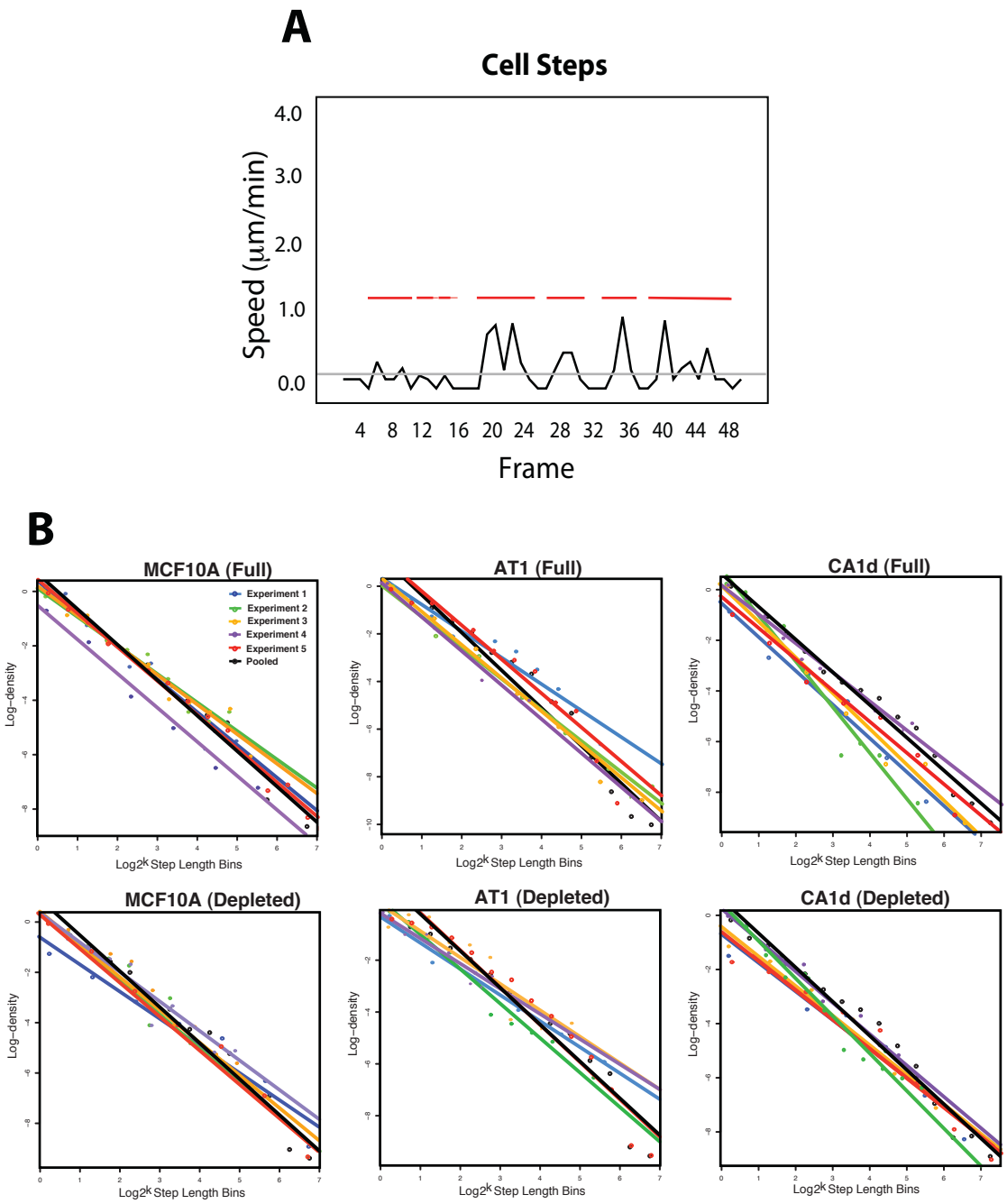



**Figure 4.4 - Step-length distribution is linear on a log-log plot.** (A) Red lines highlight the steps in a randomly chosen single-cell. A step is the time between cell pauses. A step-length is the distance traveled by the cell during a step. (B) Step-length data was binned in log 2^k bins and plotted log/log as described in Sims et al., 2008, in order to determine if the data followed a power-law distribution. A straight line in log/log is indicative of a power-law distribution.

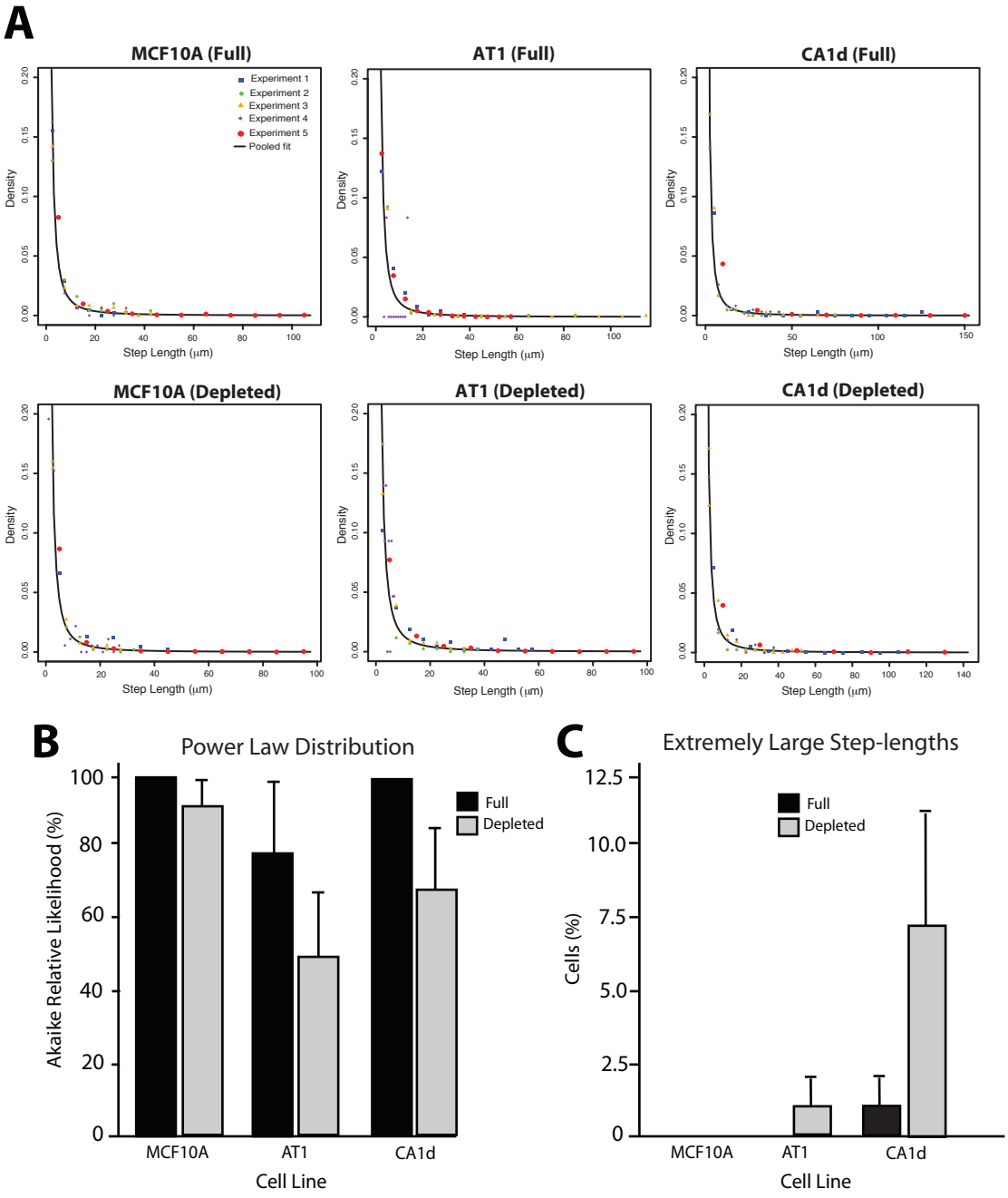



**Figure 4.5 - Step-lengths are organized around a power-law distribution.** We measured the overall distance traveled between cell pauses in a movie (defined by two consecutive frames at the same coordinate). (A) Data was plotted directly onto a Paretto distribution (a type of power-law distribution) to avoid possible problems with log/log binning. To verify the fits to a power-law distribution, we applied the MLE method (B). This approach offers an alternative fit to an exponential distribution, and compares the fits *via* Akaike weights to see which distribution is a better fit. A 100% Akaike weight value demonstrates the highest certainty that data fits the Pareto distribution, rather than an exponential. Akaike weights for individual experiments can be found in Table 4.3. (C) Percentage of cells with step-lengths too large to quantify (cells that did not pause, or paused only once during the 4 hour movies).

**Table 4.3 - MLE analysis of step-lengths.** We measured the overall distance traveled between cell pauses in a movie (defined by two consecutive frames at the same coordinate). This data was fit to a Pareto distribution (as shown in Figure 4.5A). We demonstrated the significance of the Pareto fit using a maximum likelihood estimator (MLE) method. 100% Akaike weight is indicative of a perfect Pareto curve fit. Grayed boxes indicate rejected Pareto fits due to low Akaike weight.

| Cell Line | Medium | Statistic | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 | Experiment 5 | Pooled |
|---|---|---|---|---|---|---|---|---|
| MCF10A | Full | Pareto Exp | 2.01 | 1.74 | 1.82 | 2.05 | 1.78 | 1.81 |
| | | Lambda | 0.32 | 0.172 | 0.196 | 0.3 | 0.16 | 0.178 |
| | | Akaike Weight | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| | | 95% CI | (1.82,2.24) | (1.61,1.90) | (1.68,1.98) | (1.73,2.46) | (1.72,1.85) | (1.76,1.86) |
| | Depleted | Pareto Exp | 1.58 | 1.96 | 1.91 | 1.96 | 1.89 | 1.85 |
| | | Lambda | 0.099 | 0.323 | 0.3 | 0.312 | 0.217 | 0.21 |
| | | Akaike Weight | 96.98% | 97.95 | 78.58% | 99.90% | 100.00% | 100.00% |
| | | 95% CI | (1.47,1.71) | (1.78,2.16) | (1.75,2.09) | (1.77,2.17) | (1.79,2.01) | (1.78,1.91) |
| AT1 | Full | Pareto Exp | 1.78 | 2.08 | 1.92 | 2.39 | 1.8 | 1.82 |
| | | Lambda | 0.209 | 0.227 | 0.225 | 0.524 | 0.235 | 0.232 |
| | | Akaike Weight | 95.98% | 100.00% | 100.00% | 95.70% | 0.00% | 100.00% |
| | | 95% CI | (1.65,1.93) | (1.84,2.36) | (1.76,2.11) | (1.75,3.33) | (1.75,1.85) | (1.78,1.87) |
| | Depleted | Pareto Exp | 1.62 | 2.09 | 1.74 | 2.33 | 1.65 | 1.71 |
| | | Lambda | 0.106 | 0.372 | 0.213 | 0.669 | 0.142 | 0.159 |
| | | Akaike Weight | 100.00% | 100.00% | 0.01% | 5.98% | 0.04% | 100.00% |
| | | 95% CI | (1.50,1.75) | (1.88,2.34) | (1.59,1.91) | (1.97,2.77) | (1.59,1.71) | (1.66,1.76) |
| CA1d | Full | Pareto Exp | 1.71 | 2.06 | 2.12 | 1.84 | 1.71 | 1.81 |
| | | Lambda | 0.113 | 0.356 | 0.274 | 0.19 | 0.117 | 0.15 |
| | | Akaike Weight | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| | | 95% CI | (1.50, 1.97) | (1.85,2.30) | (1.93,2.33) | (1.70,2.00) | (1.64,1.78) | (1.75,1.87) |
| | Depleted | Pareto Exp | 1.62 | 2.42 | 1.72 | 1.83 | 1.5 | 1.68 |
| | | Lambda | 0.116 | 0.445 | 0.185 | 0.214 | 0.072 | 0.13 |
| | | Akaike Weight | 100.00% | 100.00% | 15.65% | 100.00% | 0.00% | 100.00% |
| | | 95% CI | (1.52,1.73) | (2.14,2.74) | (1.60,1.87) | (1.67,2.02) | (1.42,1.59) | (1.63,1.74) |

The fitting of a power-law distribution indicates that the vast majority of cell step-lengths are short, and that long steps are infrequent (Figure 4.5A). The implication of these findings is that, even though the speed of moving cells fluctuates, the distance cells can cover between pauses (step-lengths) appears to be regulated in an orderly fashion. It will be interesting to further investigate the relationship between length of steps and possible underlying molecular mechanisms that provide cells with the ability to cover distances of shorter versus longer length. Along these lines, it appears that in non-cancer cells, under serum/EGF-depleted conditions, the step-length is maintained within the power-law distribution, whereas neither AT1 nor CA1d step-lengths fit the Pareto

distribution consistently (Figure 4.5B), as their step-lengths became very large. Thus, it is tempting to speculate that cancer cells have lost intrinsic and/or extrinsic mechanisms that dampen length of steps.

### 4.3.4 IMF Increases in Cancer Cell Lines in Depleted Media

To follow up on the analysis of persistence and step-length distribution, we sought to quantify the percentage of motile cells at any given time. The motile cell fraction was a metric used previously (Kim *et al.*, 2008), but which was of limited use to us due to its extreme variability (Fig 4.6A). So here we introduce and utilize IMF, computed as the percentage of cells moving more than one pixel (our measurement error threshold) at each sampling interval (Figure 4.6B). The IMF metric is quite robust and associated with low error across many experiments for both non-cancer and cancer cell lines. It was perhaps the metric with the least variation of all measured, suggesting that the percentage of cells trying to initiate motion at any instant is an intrinsic property of cell lines. It is interesting that the IMF for both non-cancer and cancer cell lines is very close under full media conditions. In contrast, under serum/EGF-depleted media, the non-cancer cell IMF remains the same, whereas the cancer cells' IMF increases by about 40% (Figure 4.6B), suggesting that IMF may be sensitive to extrinsic factors in cancer but not in non-cancer cells. This conclusion, obviously, needs confirmation from larger scale experimentation.

We also measured the "Motile Cell Fraction" calculated as previously reported (Kim *et al.*, 2008). The total number of cells traveling a distance of more

than one cell diameter over the course of the entire experiment (distance greater than 20 μm) was divided by the total number of tracked cells. By this method, we observed large error bars in Motile Cell Fraction for the more aggressive cell lines (AT1 and CA1d) (Figure 4.6A). It is tempting to speculate that the outcome (measured by Motile Cell Fraction) of attempts to initiate motion (measured by IMF) is tightly constrained in the non-cancer cell line, whereas that outcome is more irregular in the cancer cell lines possibly because of loss of regulation.

## 4.4 Significance and Discussion

In the pursuit of motility metrics that can quantify variability, especially in cancer cell lines, we have described analytical methods based on time-lapse microscopy data. The conclusions from our studies indicate that in the three cell lines observed, highly motile cells were in the minority, even for our aggressive cancer cell line. In the midst of high levels of variability, three stable motility metrics were observed: persistence, step-length distribution, and IMF. Cell step-lengths appeared to be organized around a power-law distribution. Further, IMF, but not persistence, is a metric that can distinguish the change in cancer cell motility properties under conditions of serum/EGF depletion. More studies are needed to verify our findings that step-length distribution and IMF are useful metrics for analyzing altered motility characteristics in multiple cell lines on varying microenvironments.

When modeling cell migration at the population level, it is fairly standard in the field to use the PRW. The PRW was developed to model particle movements

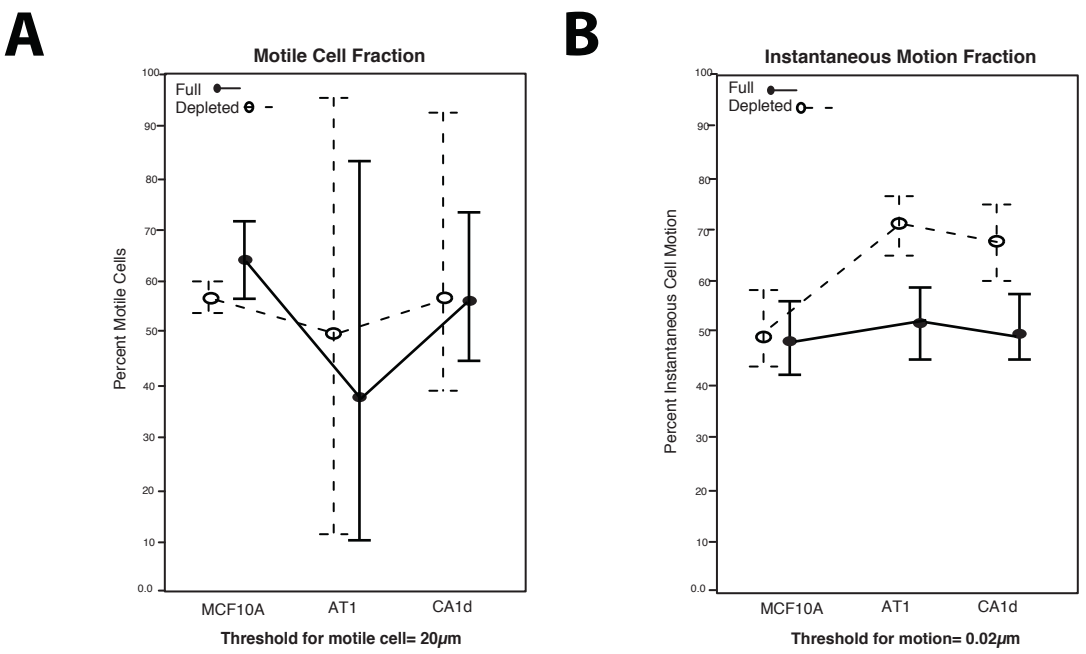



**Figure 4.6 - IMF of cancer cells increases upon serum/EGF-depletion.** (A) To calculate the motile cell fraction, the total number of motile cells was divided by the total number of tracked cells. Here, we have defined "motile cells" as any cell traveling more than 20μm (approximately one cell diameter) during the four hour period of our movies. (B) We calculated the percentage of cells that are moving at any given moment. We termed this parameter "instantaneous motile fraction, " and define its ratio as the number of cells that have moved more than one pixel (measurement error threshold), divided by the total number of cells, for every frame. By examining the IMF, it is clear that the cancer cell lines, AT1 and CA1d, become more likely to move in serum/EGF-depleted media, while there is no change in the IMF for the non-cancer MCF10A cells.

in physics, and includes both a random and persistent component of direction, with a resultant feature of a biased turn angle distribution towards continuing in the direction the particle was already traveling. One of the key assumptions in the PRW model is that all particles are in motion. Importantly, our data demonstrate that these assumptions are not accurate. In particular, we have shown that individual cells move at a non-constant speed (Figure 3.2), only a subset of cells are motile (Figure 4.1), and even motile cells stop moving frequently (Figure 4.5A) The PRW method can accurately model the average behavior of a large number of cells over a long time period, but is not necessarily acceptable for modeling motility at the cellular level, since its most basic assumptions do not reflect the biology.

To aid in the more accurate modeling of cellular migration, we have included the motile cell fraction, as presented in Kim et al., 2008 (Figure 4.6A). We have also added a two additional metrics, the step-length distribution (Figure 4.5A), and the instantaneous motile fraction (Figure 3C), which provides the percentage of cells that are in motion at any given time point. These three measurements can be experimentally determined, and subsequently integrated into probability density functions to produce models of cellular migration that are more accurate at the single-cell level (Figure 4.7).

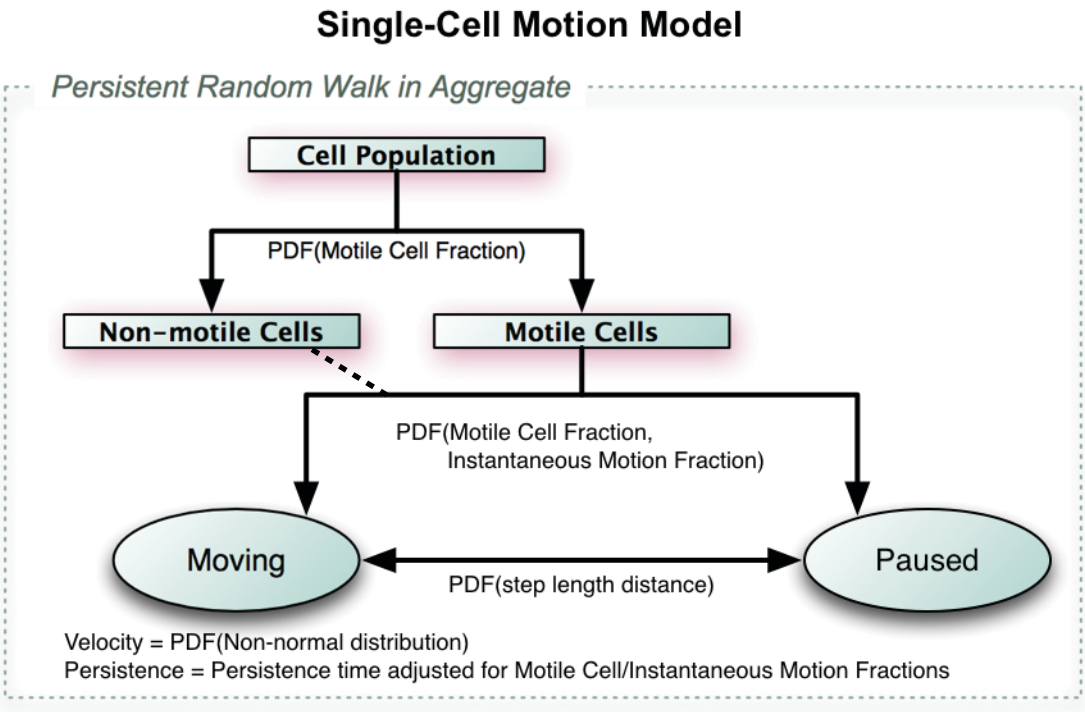



**Figure 4.7 - Single-cell motion model.** Outline of a method to mathematically model single-cell and population-level motility utilizing motile cell fraction, IMF, persistence, and step-length distance. Probability density functions (PDF) could be used to create accurate representations of heterogeneity mathematically. The dashed line indicates a small contribution towards the "moving" cell population from "non-motile" cells. This model would agree with the PRW model in aggregate, but at lower levels would more accurately model single-cell behavior.

# CHAPTER V

# DEVELOPMENT OF ADDITIONAL DYNAMIC ASSAYS TO STUDY VARIABILITY

## 5.1 Systems Biology – Multivariate Analysis

Systems Biology is a paradigm within biological sciences that seeks to understand the complex nature of biological systems by integrating information, rather than reducing it into component parts. The term systems biology was coined in 1948 by Norbert Weiner (Wiener, 1948), and since then systems biology has sought to bring together vast amounts of information to understand biology at a systems level. Systems biology research is by nature multivariate, meaning that experiments measure many metrics under numerous conditions. Computer algorithms are then used to tease apart meaning from these large datasets.

The systems biology method is necessary to understand complex regulatory systems that cannot be solved by a reductionist approach. In 2002, Kitano summed up the systems biology method by stating, "*To understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism*" (Kitano, 2002). This model can be utilized to enhance our knowledge of cellular migration, and to understand the adaptability and heterogeneity inherent in the system.

### 5.1.1 System's Approach to Studying Cellular Migration

The systems biology approach can be applied to cell migration in order to gain further understanding of the interrelations between many facets of motility, including: morphology, speed, adhesion, and spreading. Typically, the process of motility is broken down into component parts, which are often studied independently. In contrast, the study of migration *via* the systems approach seeks to capture large amounts of data from many metrics, and to analyze this multivariate data to understand the whole process at a systems level. This approach allows integration and may lead to enhanced understanding of higher ordered states of migration (Kitano, 2001).

With the addition of microscopes with automated stages, and the massive increases in processing power and data storage, it is now possible to produce enormous amounts of image data in an automated fashion. The bottleneck in imaging research is no longer data acquisition; it is now data analysis (Zimmer and Olivo-Marin, 2005). Using computer-aided image analysis or semi-automated quantitative approaches is ideal to efficiently cope with complex processes such as cell behavior (Soll, 1995; Soll *et al.*, 1988). Many researchers have utilized computer-assisted analysis, but semi- and fully-automated systems are still beyond most cell motility laboratories.

Other types of image analysis, such as kymography, currently have no automation beyond stacking images through time, along a one-dimensional line (Hinz *et al.*, 1999). Fully automated systems must be developed to analyze the

vast number of conditions necessary to use a systems biology approach to understanding cell motility in the context of the microenvironment.

### 5.1.2 Multivariate Analysis

Cell speed is just one of the single-cell motility metrics that can be measured by looking at single cell migration movies. Other metrics include: focal adhesion turnover, lamellapodial dynamics, surface area, cellular morphology, and invadapodia formation. Some of these measurements require special treatments such as stains, or fluorescent matrix deposited below the cells, and thus have not been combined into a single experiment. However, understanding the interplay of these metrics is critical for fully understanding cell migration in greater detail.

As a part of this dissertation, a method of multivariate analysis was designed and implemented as a proof-of-principle experiment. This analysis measures three metrics per cell from a standard phase-contrast single-cell migration movie. The analysis measures cell speed, surface area, and a novel measurement: Dynamic Expansion and Contraction of Cell Area (DECCA) (see Figure 5.1 for flowchart).

Although our understanding of individual processes underlying cell migration continues to increase, major gaps in information concerning how they are coordinated spatially and temporally still remain (Lauffenburger and Horwitz, 1996). New techniques need to be developed that can bring insight into how these individual processes interact by quantifying dynamic cell movements and

analyzing single cells in an automated manner. Computer-assisted, quantitative analysis of migrating cells provides an objective means of comparing migration properties of cells and yields insight into the underlying mechanisms of cell motility. Here, we report a dynamic multivariate analysis of single-cell motility that includes a combination of both novel algorithms/image analysis methods (surface area; DECCA) and existing techniques (cell speed).

## 5.2 Experimental Design

Single-cell migration assays (Section 2.4) were used to analyze the three metrics of two cell lines (HT-1080 and A431) on two types of matrix (Ln-332 and Fn) to demonstrate that a multivariate analysis at the single-cell level was feasible, and to examine the interrelatedness of the variables based on cell lines and different ECM conditions. All data analyzed within this chapter was compiled from 2 experiments (see Table 2.2).

A new dynamic multivariate single cell assay was developed to analyze three metrics of cellular migration: cell speed, surface area, and DECCA (Figure 5.1). The first measurement we present, cell speed, was captured using a standard manual cell-tracking technique (Metamorph) from movies generated by time-lapse, phase contrast microscopy. The second measurement utilized a custom-written MATLAB algorithm designed to threshold images to calculate cell surface area. Obtaining surface area measurements gives us insight into the shape and overall health of cells. For example, epithelial cells tend to decrease their surface area when unhealthy or stressed and many cell lines change their

shape upon differentiation, which is often reflected by a change in surface area. The third measurement, DECCA, is a novel measurement of total cell activity. That is, DECCA captures the pixel intensity change from one frame to the next, and averages these changes over the length of the movie. Thus, a higher DECCA value represents a higher amount of cell activity. This measurement is not necessarily a measurement of cell migration, as membrane protrusion without translocation, can also lead to high DECCA values. These three values were then combined using a unique identifier system to obtain three measurements per cell (for 1,000+ individual cells, from 2 cells lines, on 2 substrates).

New methods of image analysis were designed for the surface area and DECCA measurements. These calculations were performed *via* custom written algorithms. This multivariate assay was used to test two cell lines (HT-1080 and A431), on two types of extracellular matrix (FN and Ln-332).

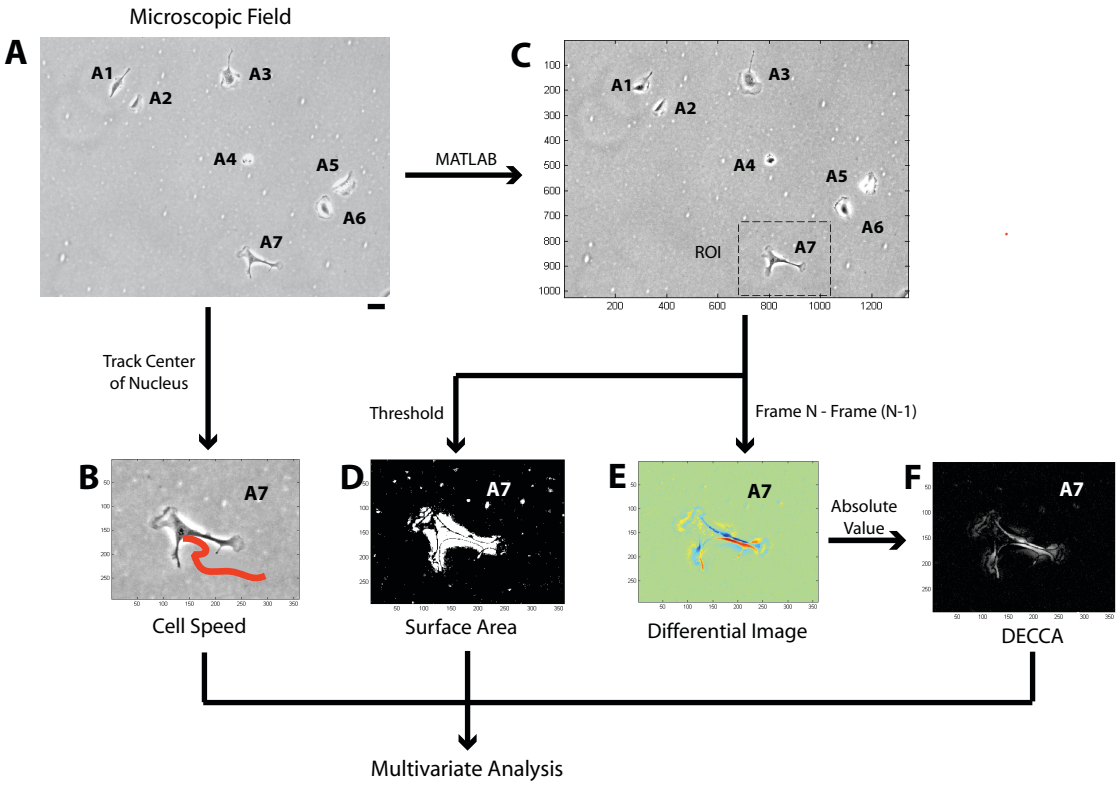



**Figure 5.1 - Overview of mulivariate profiling of single-cells.** This flowchart represents the step-wise progression of our image analysis technique, including: (A) phase-contrast image capture (6 random fields per well, in duplicate) and manual application of unique identifying numbers to all cells (i.e., A1-A7); (B) tracking the center of each cell nucleus manually using Metamorph software to quantify cell speed; (C) selection of regions of interest (ROI) manually from original phase contrast images using MATLAB; (D) creating computer-generated thresholded images in MATLAB to calculate cell surface area, and (E) creating computer-generated differential images in MATLAB by subtracting the pixel intensities from one frame to the next. (F) Differential images were further processed by taking the absolute value of the pixel intensities to obtain the DECCA measurement. The scale bar seen in part (A) is equal to 100μm. Image axes in (B-F) are MATLAB generated coordinates for each image and ROI.

### 5.2.1 Surface Area

A number of methods have also been used to quantify morphological changes at the cellular level, including a diagonal measurement of cell elongation (Mueller-Rath *et al.*, 2007), human interpretation of shapes (Jokhadar *et al.*, 2007), and most commonly surface area measurements (Alexopoulos *et al.*, 2002; Toraason *et al.*, 1990; Carpenter *et al.*, 2006; Opstal *et al.*, 1994). These studies have looked at morphology as an indicator of differentiation, apoptosis, and various other processes (Ray *et al.*, 2005; Mukherjee *et al.*, 2004). Traditional surface area measurements tend to rely on two types of analyses: edge detection through thresholding and edge-based segmentation (Alexopoulos *et al.*, 2002). Some weaknesses of these techniques include the need for fluorescent cellular markers and an assumption of a particular cell shape (e.g., a spherical cell)—both of which can lead to complications, depending on your model system (Alexopoulos *et al.*, 2002; Truskey and Proulx, 1990; Ionescu-Zanetti *et al.*, 2005). For these reasons, surface area measurements must be optimized for specific cell types (Alexopoulos *et al.*, 2002). Many epithelial cell lines spread out and lie flat against the surface they are seeded on, and have a wide variety of shapes, with multiple leading edges at once; this variety of shapes and decrease in contrast makes edge detection, by traditional means, much less accurate (Alexopoulos *et al.*, 2002). Although a number of assays currently exist to examine cell morphology, both at the cellular and sub-cellular levels, there is still certainly room for new, quantitative techniques that can be used for accurate analyses of cell shape and other motility parameters.

**A** **SURFACE AREA COMMANDS**

```
stdev_m = std (image_3d, 0,3);
stdev_v = reshape (stdev_m,1,[ ]);
[num,stdev] = hist (stdev_v,20);
[pks, locs] = findpeaks (num);
ii = locs(1);
midstd_background = stdev(ii);
mask_background = stdev_m<=midstd_background;
mask3d_background = zeros (nRows,nCols,nTimes);
for time = 1:nTimes
        mask3d_background(:,:,time) = mask_background;
end
background_3d = mask3d_background.*image_3d;
length = sum(mask3d_background(:));
background_v = zeros (1,length);
ii = 0
for time = 1:nTimes
        for row = 1:nRows
                for col =1:nCols
                        if background_3d(row,col,time)~=0
                                ii = ii +1;
                                background_v(ii) = background_3d(row,col,time);
                        end
                end
        end
end
background_mean = mean(background_v);
background_std = std(background_v);
background_leveled_v = background_v -background_mean;
mean (background_leveled_v)
background1_std = std(background_leveled_v)
abs_unmasked_3d = image_3d-background_mean;
mask_abs3d = abs_unmasked_3d>=1.5*background1_std|abs_unmasked_3d<=-1.7*background1_std;
abs_3d = abs_unmasked_3d.*mask_abs3d;

abs_v = Generate1d(abs_3d);
count_3d = abs_3d ~= 0;
count_v = sum(sum(count_3d,1),2);
count_v = squeeze(count_v);
```

**B**

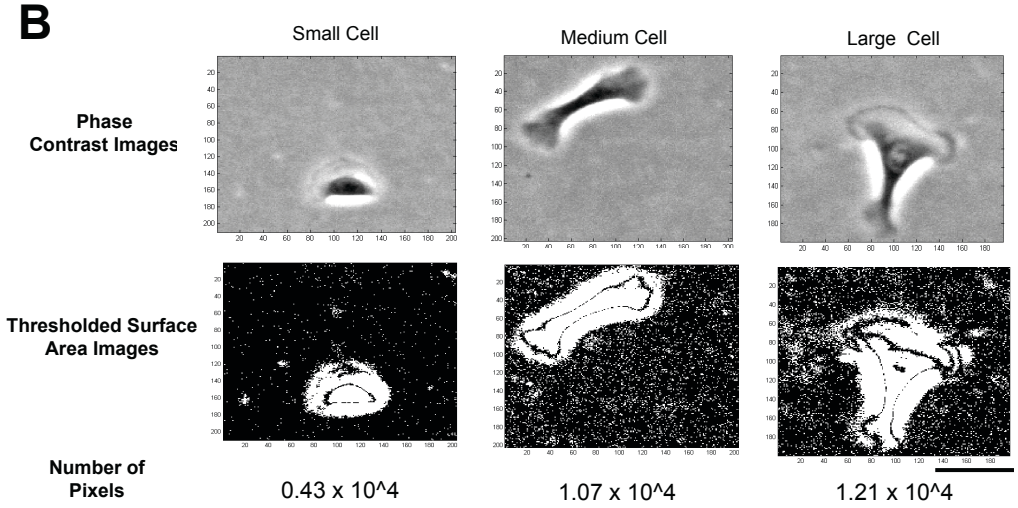



**Figure 5.2 - Surface area algorithm for image analysis.** (A) Custom-written, MATLAB algorithms were developed to measure surface area of cells (in pixels). Blue text indicates an integrated MATLAB function and black is used for all other text and commands. (B) Sample phase contrast images (top row) and corresponding micrographs of thresholded surface area images (middle row) are presented for a small, medium, and large cell. The calculated number of pixels that corresponds to each cell is also listed for reference (bottom row).

### 5.2.2 DECCA

Migration of epithelial cells follows a number of specific processes that are typically associated with specific changes in cell size and shape. Thus, one can gain insight into the specific mechanisms of cell movement by studying these morphological changes. Kymography is one method that has been used to gain insight into the mechanisms of actin, cortactin, and various other molecules involvement in membrane protrusion (Bryce *et al.*, 2005; Cai *et al.*, 2007). This technique involves high-resolution time lapse microscopy to capture subcellular motion. Kymography is used for relatively small sample sizes (due to highly magnified imaging), during relatively short periods of time, for the study of events such as lamellipodial dynamics, microtubule polymerization, and many other motility events contributing to shape change (Cai *et al.*, 2007; Bear *et al.*, 2002; Kellermayer *et al.*, 2008).

In this manner, DECCA measures the protrusive activity of cells, whether or not they actually migrate processively. A high DECCA value means that a cell has moved during the movie, possibly due to migration, but it may also be due to the additive effect of membrane ruffling and cell shape change over time. When combined with cell speed data, one can determine if the DECCA number corresponds to cell migration or cell shape change.

**A**

Dynamic Expansion/Contraction Cell Area (DECCA) Commands

```
for time = 2:nTimes
        for row = 1:nRows
                for col = 1:nCols
                        unit = image_3d(row,col,time)-image_3d(row,col,time-1);
                    if unit >= limit
                        difpos_3d(row,col,time) = unit;
                    elseif unit <= unit <= -limit
                        difneg_3d(row,col,time) = unit
                    elseif unit < limit && unit > -limit
                        difpos_3d(row,col,time) =  0;
                        difneg_3d(row,col,time) = 0;
                    else
                        display ('Menu Error in Masking, line 119')
                    end
                end
        end
end

difinc_3d = difpos_3d + difneg_3d
adifinc_3d = abs(difinc_3d);
```

**B**

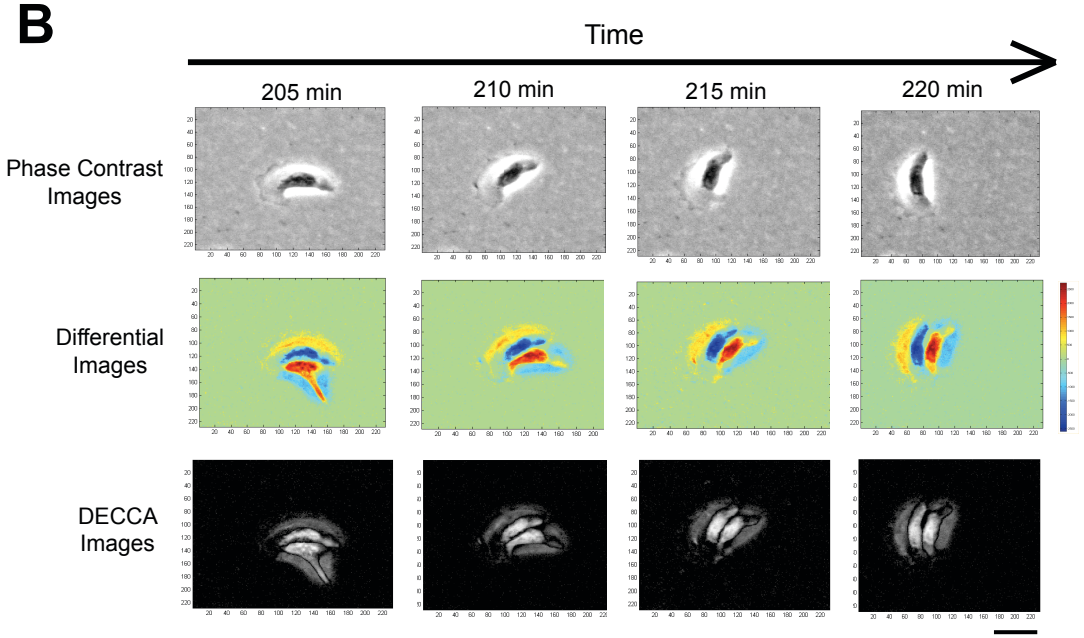



**Figure 5.3 - DECCA algorithm for image analysis.** (A) Custom-written, MATLAB algorithms were developed to measure DECCA. Blue text indicates an integrated MATLAB function and black is used for all other text and commands. (B) Sample phase contrast images (top row) and corresponding micrographs of differential (middle row), and DECCA images (bottom row) are presented for a sample motile cell over a short time series (205-220 min). In differential images, color is representative of the change in intensity between phase contrast images from frame-to-frame (blue is a negative intensity change, green is no change, and yellow/red is a positive change). DECCA images are obtained by calculating the absolute value of pixels present in corresponding differential images, and this value is averaged across an entire movie to produce a cell's DECCA measurement. The solid scale bar represents 100µm.

### 5.2.3 Materials and Methods

*5.2.3.1 Cell Culture*

Human fibrosarcoma HT-1080 cells (CCL-121) and human epidermoid carcinoma A431 cells (CRL-1555) were purchased from American Type Culture Collection (Manassas, VA). Both lines were maintained in Dulbecco's modified Eagle's medium (DMEM; Life Technologies, Inc., Rockville, MD) supplemented with 10 % fetal bovine serum (FBS; Gemini, Irvine, CA) and 1 % glutamine / penicillin / streptomycin antibiotics (Life Technologies), and incubated at 37 °C in a humidified, 5 % $CO_2$, 95 % air atmosphere.

*5.2.3.2 Cell Preparation*

The laminin (Ln) isoform Ln-332 (1µg/mL; purified in-house) or human plasma fibronectin (Fn; 10 µg/mL; Millipore, Billerica, MA) was coated on Nunc™ polystyrene, non-tissue culture treated, 6-well microplate dishes (Cole Parmer, Vernon Hills, IL) for 1 h at room temperature (RT). The dishes were then blocked with 5% milk (Regilait, France) in phosphate buffered saline solution (PBS, Invitrogen) for 1 h at 37ºC.

Cell lines were trypsinized (TrypLE Select, Invitrogen, Sunnyvale, CA), neutralized with L-15 media (Invitrogen) supplemented with 10% FBS, washed three times in PBS, and resuspended at a density of $2x10^4$ in L-15 media supplemented with 10% FBS in 6-well microplates. Cells were allowed to adhere for 1 h within the heated (37ºC) microscope chamber.

*5.2.3.3 Time-lapse Microscopy*

Time-lapse microscopy was conducted using a Zeiss Axiovert 200M microscope (Zeiss, Thornwood, NY) equipped with a temperature-controlled chamber and an automated x-y-z stage (0.2 μm repeatability). Microscopy was under the control of MetaMorph software (Molecular Devices, Sunnyvale, CA). At the beginning of each experiment (0 h), six fields were manually selected at random from within each well (in duplicate). Each region was focused manually, and the specific x, y, and z coordinates for each was saved using MetaMorph's "Multi-dimensional Acquisition" tool. Phase-contrast images were captured automatically every 5 min for 4 h. Following image capture, all 49 individual images from each particular coordinate were combined using MetaMorph to produce image stacks.

*5.2.3.4 Cell Speed Quantitation*

Image stacks (with 49 slices; sample slice seen in Figure 5.1A) were opened in MetaMorph, and the "Track Points" function was used to manually track cells (Figure 5.1B). All cells that remained within the field were tracked by following the center of their nucleus. Cells that collided with other cells and dividing cells were included in our analysis (there was no significant difference between touching, non-touching, and dividing cells' speed, results not shown). Tracking data was exported to a Microsoft Excel spreadsheet for storage. The cell speed parameter was finally calculated by averaging all data collected for each cell first, followed by averaging all cells for that sample population, and final presentation of data includes mean ± standard deviation (SD) of that particular

set. After manually tracking cell speed, a unique identifier (based on date, cell line, ECM component, microscopy field, and cell number) was assigned to every cell, by manually inserting text next to each cell in the final frame of each movie (see Figure 5.1A). This ID system was applied for simplification of binning cells prior to multivariate analysis.

*5.2.3.5 Computer-assisted Quantitative Analysis (Surface Area and DECCA)*

All subsequent image analysis was performed on non-compressed, 16-bit, TIFF image stacks. Computational and programming support was provided by MathWorks™ MATLAB® (Natick, MA). We used both custom-written algorithms and several advanced MATLAB image-processing toolbox functions, further described below. For each field's stack of microscopic images, a user-defined rectangular region of interest (ROI) was manually drawn around each cell to be analyzed, and all pixels in each ROI were systematically processed one at a time by the software (Figure 5.1C). Subsequently, using two separate, custom-written algorithms for MATLAB, two processed image sequences were produced: 1) intensity-weighted thresholded images (Figure 5.1D, used to calculate surface area) and 2) differential intensity images (Figure 5.1E). These two image sequences were then used to derive measurements for cellular surface area and DECCA (Figure 5.1F), respectively, as described in detail below.

*5.2.3.6 Surface Area Quantitation*

Intensity-weighted, thresholded images were generated using a custom-written MATLAB thresholding algorithm that separates pixels of the background from pixels within the cells based on intensity (Figure 5.2A). Temporal SD's-were

calculated for each pixel, which resulted in an array of SD's for each image stack. A histogram of these values was automatically created and the peaks of each determined by a standard MATLAB "Imaging Toolbox" function. Each peak of the histogram represents the mean for each different pixel class, with the lowest peak representing the background of each image, which was applied as the threshold value for our images. All pixels at the same row and column (i.e., in the same image) that were an SD of intensity lower than the background peak were subsequently saved to a list in Microsoft Excel. When complete, this list represents the intensity characteristic of the background for each image stack. This mean value of each stack (taken from the list) was then subtracted from each pixel in each movie, in order to set the new mean of the background to approximately zero. This leveled image stack was further thresholded and all pixel values within 1.7 SD of the background mean were again reduced to zero. This particular number (1.7) was optimized for our cell types and images, and may vary considerably if the user includes different cell lines or image-capture techniques. In other words, algorithms have been fine-tuned for our model, to remove background noise and appropriately normalize all phase contrast images, prior to further quantitative analysis of surface area.

After images were thresholded to remove background using the above technique, all remaining non-zero pixels were taken to represent the selected cell (Figure 5.2B); the number of pixels quantitated for each ROI represented the SA measurement for the cell of interest. All 49 frames for each individual cell were then averaged to produce a mean surface area measurement ± SD for that cell

over time. This measurement was taken for all adherent, healthy cells of all frames that were in focus and recognized by the software; the occurrence of cells not recognized by the software was negligible (results not shown).

*5.2.3.7 DECCA Quantitation*

In brief, DECCA is an index number for the total amount of cell motion over time. This motion is not correlated directly with cell migration, as a non-moving cell can ruffle its membrane and produce a DECCA value without physically translocating across a surface or substrate. In this way, the DECCA measurement incorporates both cell motility, and cell shape change. This parameter is calculated as described below.

Differential intensity images (see Figure 5.1E) were generated using a custom-written arithmetic algorithm in MATLAB that subtracts the pixel intensity value of each pixel from its counterpart in the same row and column in the next frame (Figure 5.3A). These differential intensity images show the relative change in pixel intensity (with color-coded scale) from frame to frame (Figure 5.3B), and in this way highlight dynamic cell motion. A non-zero value for a differential pixel indicates an intensity change for that particular pixel from the last frame. Images were thresholded by setting all differential pixels with a value lower than 250 to zero. It is important to note that this value (250) was optimized for our cell types and microscopy technique, and may vary if the user includes different cell lines or image-capture techniques. This analysis results in the creation of an image stack with colored pixels representing the change in pixel intensity from one frame to another. The color and pattern of this differential image stack

represents the magnitude and area of cellular motion. These differential image stacks can be viewed as movies to observe the dynamics of cellular motion over time. We have also developed an index number to quantify the total amount of cellular motion within these movies. To create the index number, the absolute value of each pixel was taken and the differential values for all pixels in each frame were summed to produce the total absolute differential intensity for each frame. This parameter was averaged across each stack to obtain the DECCA measurement presented for each cell (displayed as intensity units).

*5.2.3.8 Data Analysis and Statistics*

Data analysis was performed using SPSS, version 16 (SPSS, Inc., Chicago, IL). The Shapiro-Wilks test for normality was applied to all data sets for distribution analysis. The Kolmogorov-Smirnov two-sample non-parametric test was subsequently applied to data to check for significant differences ($P < 0.05$) across various groups (i.e., by cell line and substrate) for all measurements. For follow-up correlation studies, the unique cell identifying numbers were used to manually combine all three sets of data, so that all parameters for each cell were grouped together in analysis. To analyze relationships between the three measurements, Spearman's R correlation coefficients were calculated for all pairs of variables. All data are presented in terms of mean ± standard deviation (with 95% confidence intervals where indicated).

## 5.3 Results

### 5.3.1 Cell Speed Quantitation

A431 cells were plated in microplates coated with either 1 µg/mL Ln-332 or 10 µg/mL Fn, and time-lapse microscopy experiments were performed in duplicate. As displayed in Figure 5.4A, A431 cells exhibited a mean cell speed of 0.82 ± 0.44 µm/min on Ln-332 (grey; N=3 (415 cells)) and 0.41 ± 0.21 µm/min on Fn (white; N=2 (272 cells)). Cell speed on Ln-332 substrate was found to be significantly faster than on Fn ($p < 0.001$). Furthermore, there was no evidence of a difference ($p > 0.05$) between A431 cell speed measurements within duplicates performed each day, nor between repeated experiments, demonstrating the repeatability of our manual cell tracking method (results not shown).

HT-1080 cell speed was also examined by seeding cells either on 1 µg/mL Ln-332 or 10 µg/mL Fn-coated microplates, and experiments were performed, in duplicate, as above. As depicted in Figure 5.4A, the mean speed of HT-1080 cells was calculated to be 1.07± 0.45 µm/min on Ln-332 (grey; N = 2 (254 cells)) and 0.91 ± 0.37 µm/min on Fn substrate (white; N = 1 (84 cells)), which was found to be a significant difference ($p < 0.05$). Furthermore, there was no evidence of a difference ($p > 0.05$) found between duplicates, or across days of experimentation, further demonstrating the repeatability of our manual cell tracking results. In summary, both A431 and HT-1080 cells showed a

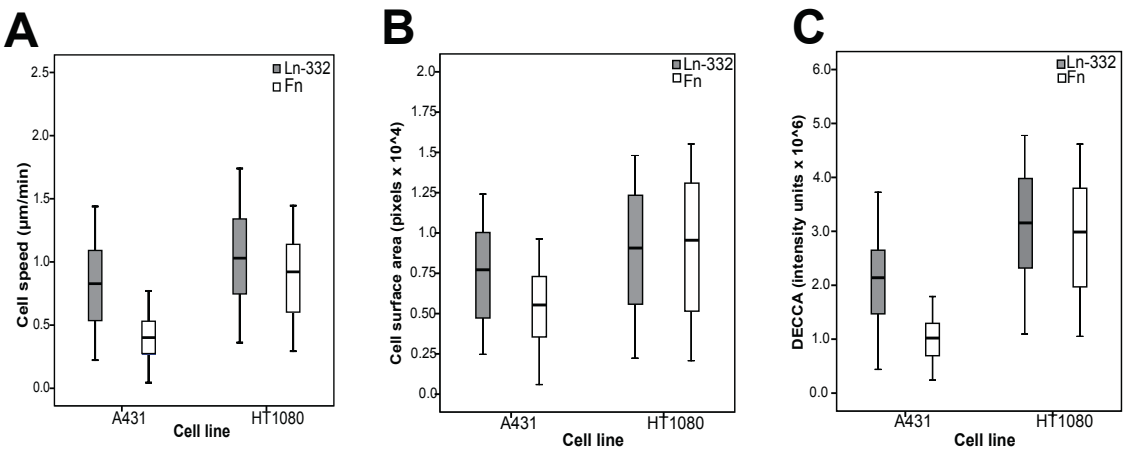



**Figure 5.4 - Quantitation of cell speed, surface area, and DECCA.** A431 or HT-1080 cells were allowed to adhere to laminin-332 (Ln-332) or fibronectin (Fn) coated microplates for 1 h at RT. Time-lapse microscopy was used to capture cell motility for 4 h. (A) For cell speed quantification, cells' paths were tracked manually using MetaMorph software. On Ln-332 (grey), A431 (N = 415 cells) and HT-1080 (N = 254 cells) speed was found to be significantly different ($p < 0.05$). Similarly, on Fn (white), A431 (N = 272 cells) and HT-1080 (N = 84 cells) speed was also found to be significantly different ($p < 0.05$). Each cell lines' speed on the two substrates was also significantly different ($p < 0.05$). (B) Cell surface area measurements were captured using custom-written MATLAB algorithms, which removed background pixels via thresholding. All remaining pixels were taken to represent the cell. On Ln-332 (grey), A431 and HT-1080 cell surface area measurements were significantly different ($p < 0.05$); on Fn (white), measurements were also significantly different ($p < 0.05$). (C). DECCA measurements were obtained using custom-written MATLAB algorithms, which took the absolute value of subtracted pixel intensities frame to frame, to produce a cell activity index. There was a significant difference between cell lines, with HT-1080 cells having a higher DECCA on both matrices ($p < 0.001$). There was also a significant difference on A431 DECCA on different matrices ($p < 0.001$); however, there was no difference between matrices for HT-1080 cells ($p > 0.05$). All plots represent mean (—), 95% confidence interval (box), and standard deviations (whiskers).

reproducible and significant difference in mean cell speed, with HT-1080 cells migrating faster than A431 cells across both matrices ($p < 0.05$). In addition, both cell lines migrated faster on Ln-332 than on Fn ($p < 0.001$).

### 5.3.2 Surface Area Quantitation

The same movies used to quantify cell speed in the previous results were also used to determine cell surface area measurements. A blinded investigator produced this data, with no knowledge of cell speed results. Of all the cells captured during time-lapse microscopy for initial cell speed analysis, only 3.1% (34 / 1080 total cells) were lost during this surface area analysis, due to unfocused images or cells leaving frames mid-movie.

As depicted in Figure 5.4B, the mean cell surface area of A431 cells on Ln-332 and Fn substrates was calculated to be $0.78 \times 10^4 \pm 0.46 \times 10^4$ and $0.59 \times 10^4 \pm 0.32 \times 10^4$ pixels, respectively. Mean cell surface area of HT-1080 cells on Ln-332 (grey) and Fn (white) was $0.94 \times 10^4 \pm 0.51 \times 10^4$ and $0.99 \times 10^4 \pm 0.51 \times 10^4$, respectively. There was a significant difference between surface area measurements of the two cell lines, with HT-1080 cells having a higher surface area on both matrices ($p < 0.001$). This computer-assisted analysis (that HT-1080 cells were larger than A431 cells) was confirmed by the researcher who performed the manual cell tracking data analysis, demonstrating the accuracy of the computed-assisted surface area measurements. There was also a significant difference on A431 cell surface area on different matrices ($p < 0.001$), however, there was no difference between matrices for HT-1080 cells ($p > 0.05$).

### 5.3.3 DECCA Quantitation

The same movies used to quantify cell speed and surface area were also used to determine DECCA values (see Figure 5.1 for clarification), for the analysis of total cellular motion (as described in sections 5.2.2 and 5.2.3). Again, a blinded researcher produced this data, with no knowledge of cell speed results. The DECCA of A431 cells on Ln-332 and Fn was calculated to be $2.22 \times 10^6 \pm 1.33 \times 10^6$ intensity units and $1.05 \times 10^6 \pm 0.45 \times 10^6$ intensity units, respectively. The DECCA of HT-1080 cells on Ln-332 and Fn was $3.20 \times 10^6 \pm 1.35 \times 10^6$ intensity units and $3.04 \times 10^6 \pm 1.33 \times 10^6$ intensity units, respectively (Figure 5.4C). There was a significant difference between cell lines, with HT-1080 cells having a higher DECCA on both matrices ($p < 0.001$). There was also a significant difference on A431 DECCA on different matrices ($p < 0.001$), however there was no evidence of a difference between matrices for HT-1080 cells ($p > 0.05$).

### 5.3.4 Multivariate Correlations

Cell speed, surface area, and DECCA measurements correlated significantly for all pairs, in both A431 and HT-1080 cell lines, and on both matrices (Figure 5.5). Spearman's correlation coefficients ($\rho$) were calculated for the data grouped by cell line, matrix, and in its entirety. Speed vs. surface area typically showed a low correlation ($\rho = 0.163$-$0.349$), while speed vs. DECCA

(ρ = 0.394-0.594) and surface area vs. DECCA (ρ = 0.434-0.545) both showed a medium level of correlation (Figure 5.5). This trend indicates that larger cells tended to migrate faster, and have a higher cell activity level than smaller cells, as anticipated for these particular cell types. These results again demonstrate the validity of our method. Furthermore, since the trend was present in two cell lines, and on two different matrices, our method has been demonstrated to be reproducible for a variety of experimental conditions.

## 5.4 Significance / Discussion

Our experimental results demonstrate the repeatability and reliability of our technique. Our data indicate low to medium-high correlations between all three metrics, depending on the particular cell line and substrate combinations. This range of relationships was anticipated for these cells, due in part to their high migration rates, and various shape changes observed in culture, when plated on Ln-332 or Fn (Winterwood *et al.*, 2006). However, given different experimental guidelines (i.e., cell lines, ECM components, or introduction of mutations), these trends may certainly change. For example, NRK49F cells with defects in *Rho* or adducin have been shown to have active lamellipodial ruffling, while being unable to migrate (Dove, 1999). Based on these findings, we hypothesize that these mutant cells would have an unchanged cell surface area and DECCA (compared to wild type), but their cell speed would decrease drastically. Furthermore, inclusion of leukocytes, or various other immune cells,

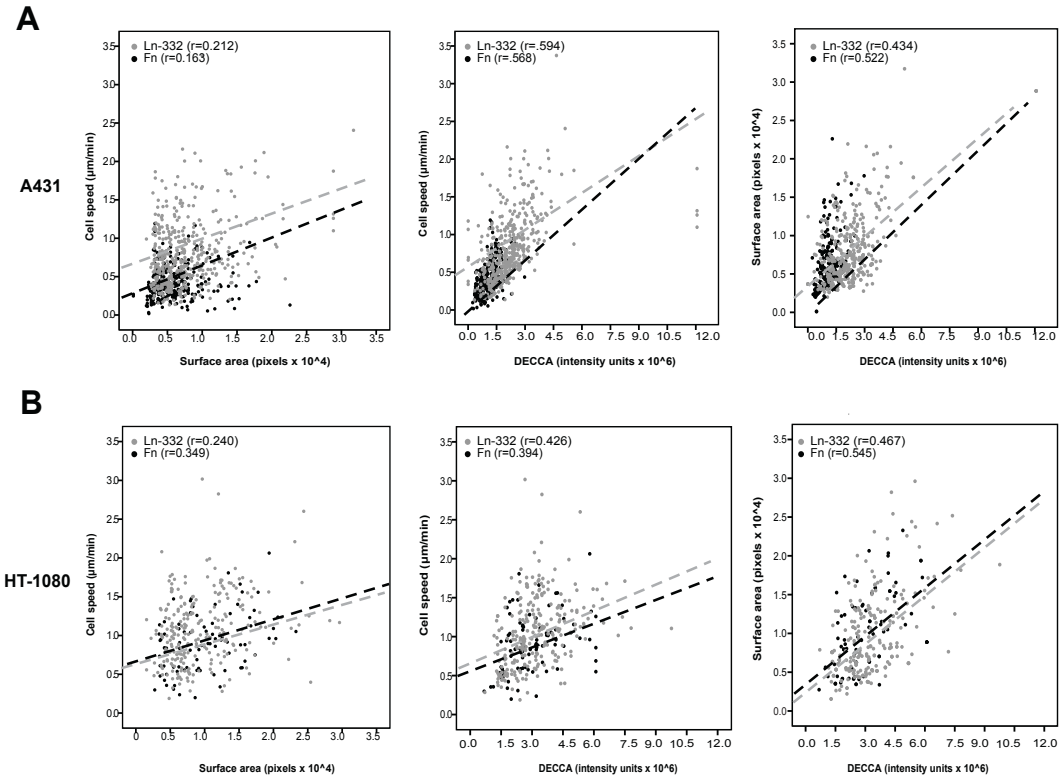



**Figure 5.5 - Correlation between variables.** Regression plots for A431 cells (A) and HT-1080 (B) cells on both Ln-332 (gray markers) and FN (black markers). All Spearman's correlation coefficients (r) demonstrated various positive levels of correlation between the three measurements when grouping by both cell line and ECM substrate.

may also significantly alter results, as these cells are commonly very motile, but much smaller and produce less dynamic shape changes.

As outlined in Chapter 2, cell speed analysis is one important component of studying cell migration. While methods of automated cell tracking exist in commercial software programs, they are not widely used in the field because either they require labeling of cells, or their accuracy and reproducibility (compared to manual tracking) is lacking (Chon, et al, 1997). Many researchers prefer to track unlabeled cells using phase contrast microscopy, both for ease of use and to eliminate added variables. A fully automated system, termed the 2D DIAS, has been developed to study the motility of *Dictyostelium amoebae* (Wessels *et al.*, 2006; Wessels *et al.*, 2008), but thus far it has been more difficult to develop such a system for epithelial cells, due to their complex behavior and irregular cell shape (Zimmer, et al 2005). Ultimately, one of our immediate goals is to update our current manual speed tracking method to include a similar automated system, but not at the expense of accuracy.

Surface area analysis is also an important component of cell migration studies because cell size can be linked to cell shape and health. In general, cells that have suffered mild insults shrink in size as one of the first steps in the apoptotic pathway (Elmore 2007; Kerr *et al.*, 1972). Differentiation of cell lines is also often associated with a change in cell size that may be reflected in our surface area measurements (Aharon and Bar-Shavit, 2006; Zouboulis *et al.*, 1994). There are currently a number of available methods to obtain surface area measurements through image analysis. However, many of these methods rely on

either the use of fluorescent labeling of cells, differential interference contrast (DIC) microscopy images, and edge detection methods that require heavy computing power, and also often making assumptions about a general cell shape. For some applications, our method of surface area estimation will work well for eukaryotic cells. It is not as accurate as some methods referenced above, but it shows relative changes in surface area very well for phase contrast images, and with very little processing power needed for our algorithm. In addition, our method allows a researcher to follow surface area changes over time (results not shown).

The introduction of the DECCA measurement is a significant contribution, as this technique captures cell activity in a way that no other applications have demonstrated previously. It is important to note that DECCA measures the protrusive activity of cells, whether or not they actually move in a processive manner (*i.e.*, across a substrate). In fact, a DECCA index need not correlate positively with movement; any positive correlation with cell speed is an indication of how efficient cell activity is towards actual migration. A simple example includes comparison of a motile cell that physically moves across a field, compared to a non-motile cell. The moving cell will always have a DECCA value, since its “footprint” changes from frame to frame. However, the non-moving cell may have a low or high DECCA, depending on the protrusion activity of particular cell. The cell may be completely inactive (if all images are the same, DECCA = 0), or it may change shape without moving its nucleus by lamellapodial ruffling or creating numerous cell protrusions that lead to shape change. As a result, the

image pixel intensity changes, even though their nucleus does not move. A manual version of differential imaging was previously shown in a publication by Fukui *et al*., and was referred to as producing “difference pictures” (Fukui *et al.*, 1991). However, we are unaware of any other algorithms/computational methods that reflect the same activity as DECCA. Originally, DECCA was developed with the intent to distinguish between two cell types that have the same migration speed, but very different membrane protrusion dynamics. For example, our analysis demonstrated that although HT-1080 cell speed was significantly altered ($p < 0.01$) by changing the matrix, the surface area and DECCA of these same cells were essentially unaltered ($p > 0.05$). This data may indicate that HT-1080 cells have the ability to spread and become activated by both matrices, but for reasons yet to be determined, the cells have a significantly slower migration rate on Fn. Although we cannot explain these differences based on our preliminary analysis, our assay was able to provide additional insight, which would have been missed using population-based cell migration techniques or classical motility tracking assays. By understanding the interplay between cell speed, surface area, and DECCA measurements, our method may lead to additional cell migration hypotheses and findings.

Knowledge of the fundamental biological mechanisms of cell motility is currently spurring the development of novel pharmacological and genetic approaches that attempt to harness this process, in order to ultimately overcome pathological events such as cancer metastasis. Researchers have screened thousands of compounds for the ability to inhibit cell migration, in hopes of

developing new drug targets (Yarrow *et al.*, 2005). However, commonly used assays that study these interactions cannot distinguish off-target effects. For example, adding formaldehyde to cells would surely halt their cell migration, but will do so by fixing and killing the cells, not because it is a specific inhibitor of migration. In some instances, applying a DECCA measurement may be a useful control for cell health, because of its high resolution and focus on individual cell parameters. However, in order to use our method for large scale screens such as these, some parameters of our experiment will need to be improved upon. We are actively developing many other aspects of this method that can facilitate more efficient data collection and analysis. The most notable addition needed is fully-automated cell tracking software as we mention earlier. We are currently working on a new cell tracking system using our differential imaging method to develop a completely automated technique for epithelial cells similar to those for amoebae.

A method of auto-selection could be implimented to narrow the ROI to the minimum window size. With our current program, and operator must manually select the ROI. Future versions of the program will auto-select all ROIs by means of a fast, movie-spanning analysis of the SD of pixel intensity both spatially and temporally. By auto-selecting ROIs, we will decrease both image processing time and non-specific background by applying minimally sized ROIs. Other modifications to our image processing programs are also being adopted, including changes to noise reduction, and image normalization methods.

Another powerful addition to this method is the ability to look at speed, surface area, and DECCA at all time points over the course of the movies. For our initial proof-of-principle analysis presented here, all data points were averaged over the course of the movie (e.g. one cell speed measurement per cell per movie). In fact, there were 49 individual measurements each for cell speed, surface area, and DECCA. We are keenly aware that our multivariate method has the ability to study the stop-and-go pattern of cell locomotion or the change of surface area and DECCA over time using the same statistical techniques introduced in chapters 3 and 4.

Here, we present a method that produces a multivariate profile for individual cells based on three metrics: cell speed, surface area, and DECCA. In this regard, we can generate three dimensional plots, where each data point represents an individual cell (Figure 5.6). In the future, we plan to use this technique to separate interesting sub-populations within specific cell lines using similar statistical techniques that are used for statistical analysis of cell sorting data (Bindschadler and McGrath, 2007; Mochizuki *et al.*, 1996). In this manner, we can further dissect the complex mechanisms of cell migration utilizing the systems biology method, and improve our understanding of cellular adaptability and heterogeneity.

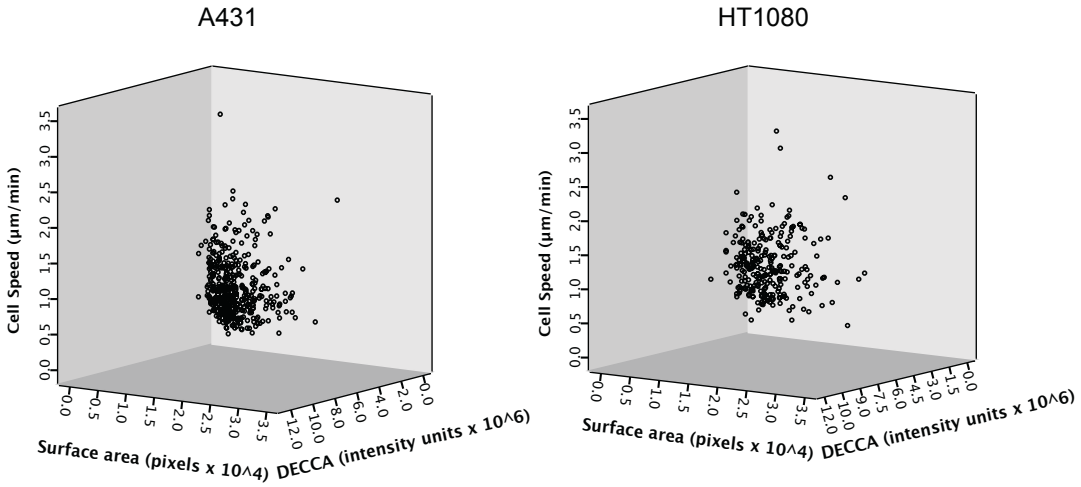



**Figure 5.6 - 3-D Graphs.** Multivariate analysis plotted in three dimensions. Every point represents one cell's averaged speed, surface area, and DECCA metrics.

# CHAPTER VI

# SUMMARY AND DISCUSSION

## 6.1 Project Summary

An analysis of cell line heterogeneity by image-based migration assays was undertaken with the goals of: 1) testing and quantitating heterogeneity of cell speeds, 2) exploring the differences in cell speed between non-tumorigenic and cancer cell lines, and 3) developing novel tools and metrics for the study of dynamic characteristics of cancer cells. To this end, over 7,300 cells were manually tracked, from a variety of cell lines, in an assortment of microenvironmental conditions.

### 6.1.1 Demonstration of Heterogeneity in Cell Speeds

Cell speed heterogeneity is routinely overlooked or ignored in the literature by presenting average speeds. Many models of cellular migration also ignore the presence of speed heterogeneity. For example, the commonly used PRW model assumes all cells are in motion at all times. To address these assumptions, single-cell motility assays were performed (Section 2.4) for thousands of cells in many environmental conditions (Table 2.1). Results from these data unambiguously demonstrate cell speed heterogeneity, and further demonstrate that cell speed is non-normally distributed (Figures 4.1 and 4.2). Thus, by

studying only cell speed averages a large quantity of data is overlooked, and furthermore, results may be misleading.

### 6.1.2 Development of Metrics to Further Quantify the Motility of Epithelial Cells

As this project progressed, it became obvious that a number of experimental observations could not be quantified by traditional cell motility metrics. Therefore, new metrics were designed to accurately quantify both single-cell motility characteristics, as well as population-level motion dynamics, with the goal of quantifying cell parameters that were previously unmeasured.

At the single-cell level, the metrics IMF and DECCA were designed as a part of this dissertation. IMF is used to measure the percent of time a cell is in motion. For the IMF analysis presented in Figure 4.6B, this metric showed more consistency from experiment-to-experiment as compared to the previously published metric, motile cell fraction. IMF can be thought of as attempts by the cell to initiate motion, whether or not it leads to effective motion. DECCA was another metric designed as a part of this project. DECCA is a measurement of total cellular movement activity as measured by changes in light intensity over time in phase-contrast microscopic images. This metric allows the quantification of cell motion which does not necessarily translate into translocation. Thus, two cells with identical cell speeds can often be distinguished by their DECCA values. Both of these metrics provide researchers with quantitative tools that describe single-cell motion in new ways.

Analysis at the population-level lead to the development of a number of novel statistical measurements for quantifying heterogeneity, including: variability of cell speed fluctuations, step-length distributions, and speed variation in response to microenvironmental changes. The first, variability of cell speed fluctuations, quantitates the heterogeneity of cells within a population in terms of cell speed fluctuation. If all the cells within a population fluctuate their speeds in the same fashion, the value of this metric will be low, indicating the cells are acting similarly, and by extension, are not heterogeneous. In this fashion, the metric can quantitate motility heterogeneity within a population. The second metric, step-length distributions was borrowed from ecological analyses of foraging behavior (Viswanathan *et al.*, 1996). The metric quantifies the distance traveled by cells between consecutive pauses. To our knowledge, this type of analysis has not been undertaken for epithelial cells. This metric was shown to follow a Pareto distribution, which may be altered for cancer cells in serum/EGF-depleted media. The final metric developed was speed variation in response to microenvironmental change. This metric quantitates the change in speed variation of a cell line when exposed to different conditions. The metric utilizes experimental controls to reduce experiment-to-experiment variability. The development of these three population-level metrics allow the quantitation of heterogeneity across cell lines and conditions.

In total, the metrics developed through this work have expanded the toolset available to the researcher interested in cell heterogeneity; providing us with the ability to study cancer at a whole new level.

### 6.1.3 Exploration of Changes in Cell Motility Between Cancer and Non-tumorigenic Cell Lines

Using the metrics outlined in Section 6.1.2, we were able to more accurately probe for differences between the motility of non-tumorigenic and cancer cell lines. Results indicated that the motility characteristics of cancer and non-tumorigenic cell lines in the presence of full media were virtually indistinguishable (Figure 6.1A and Figure 3.4). Cells of all three cell lines moved with similar speed, persistence, and IMF, and their step-lengths followed a power-law distribution. However, when serum and EGF were removed from the media, there was a strong divergence in phenotypic response. The non-tumorigenic MCF cells decreased their speed, variation, variability, and heterogeneity, while persistence increased (Figure 6.1B). In contrast, cancer cells increased their speed, variation, variability and heterogeneity, as well as their persistence and IMF. Furthermore, cancer cells may have a breakdown in their power-law distribution of step-lengths. In total, it appears that in cancer cells, the cells are more likely to move, pause less frequently, and move in a faster and more persistent fashion than non-cancer cells (Figure 6.1C).

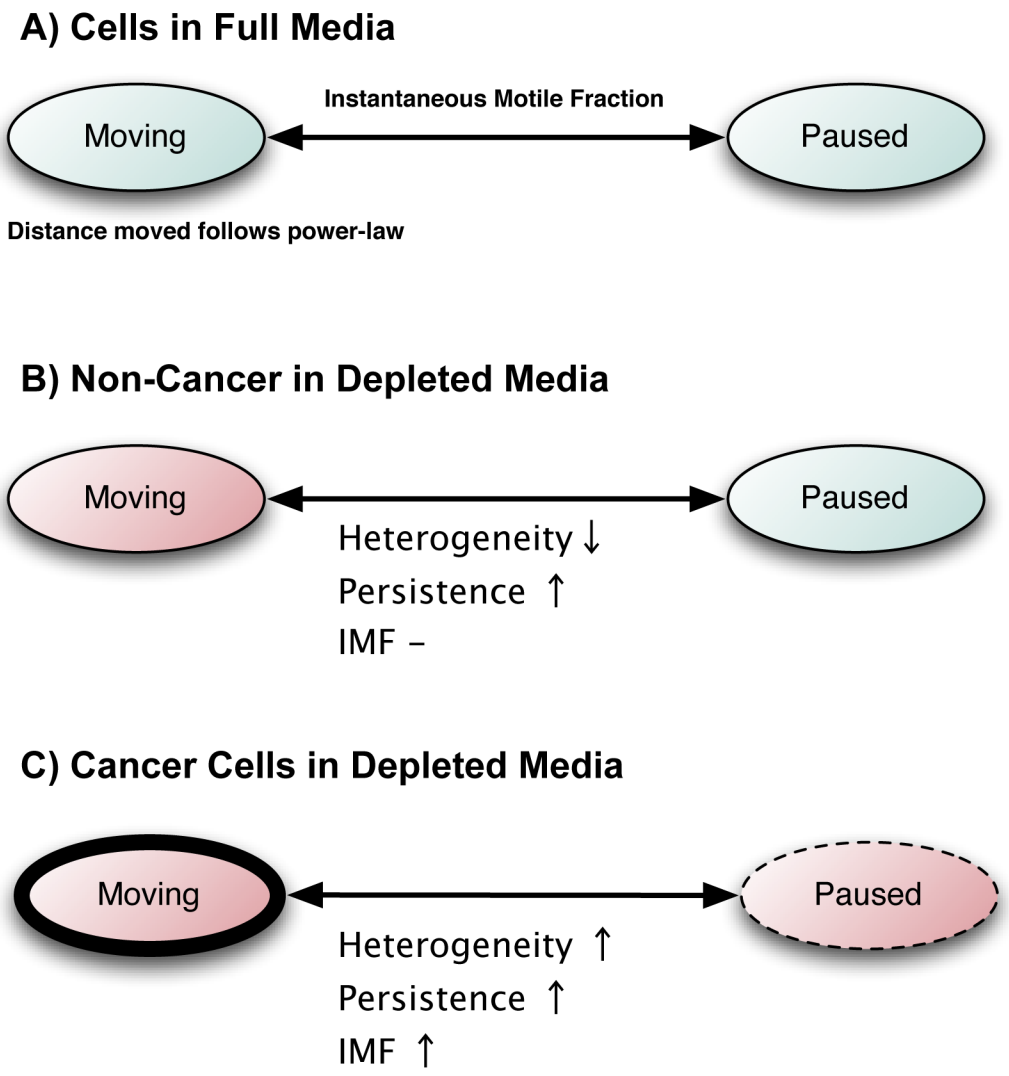



**Figure 6.1 - Cancer Cell Motility changes.** Summary of changes to cell motility metrics studied in Chapters 3 and 4. Blue denotes no change, and red indicates altered values. (A) In full media, cancer and non-cancer cells in this model system behave in a nearly identical fashion. Cells are either "Moving" or "Paused." Individual cells switch from moving to paused, or *vice versa*, based on the IMF and step-length distribution. (B) Non-cancer cells, when exposed to serum/EGF-depleted media, exhibit a speed decrease, and all metrics of variation, variability, and heterogeneity also decrease. In addition, the IMF is decreased and persistence is increased. Step-length distribution is unchanged. Thus, moving cells are affected by the change in media. (C) Cancer cells, when exposed to serum/EGF-depleted media increase their speed, variation, variability, and heterogeneity, as well as their persistence and IMF. Furthermore, there is a breakdown of the powerlaw distribution of step-lengths. In total, cancer cells in depelted media are more likely to move, move in a more persistent manner, and with increased speed. However, the variability of all metrics observed is also increased.

## 6.2 Conclusions

The analysis of the cell lines in the MCF10A model system in Chapters 3 and 4 led to several conclusions:

- Single-cells exhibit greater speed fluctuation in cancer cell lines
- Cell-to-cell variability of speed is greater in cancer cell lines
- Population-level speed heterogeneity is greater in cancer cell lines
- Day-to-day speed heterogeneity is greater in cancer cell lines
- Cancer cell speed variability increases in serum/EGF-depleted media, while non-tumorigenic cell speed variability decreases.
- Cloned cancer cells maintain speed variability and response to serum/EGF-depleted media
- Persistence increases for all cell lines in depleted media
- Cell step-lengths appear to follow a Paretto distribution
- IMF of cancer cells increases in depleted media

Based on metrics described in Chapter 3 and 4, MCF, AT1, and CA1d cells are not easily distinguishable by traditional motility analyses in full media. However, a switch to serum/EGF-depleted media creates an environment where the motility characteristics of non-tumorigenic and cancer cell lines diverge. These differences are summarized graphically in Figure 6.1. Furthermore, these differences would have been missed by traditional PRW models of migration, due to the fact that all three cell lines displayed similar levels of persistence. Thus,

the novel metrics developed in this dissertation allow one to distinguish cancer from non-cancer based on motility assays, and to quantify their differences.

The development of a multivariate assay in Chapter 5 led to the following conclusions:

- multivariate dynamic assays at single-cell resolution are possible
- A431 and HT10-80 cells, in full media, display positive correlation between cell speed, surface area, and DECCA
- HT10-80 cell speed, surface area, and DECCA was significantly higher than that of A431 cells
- both cell types displayed higher cell speed and surface area on Ln-332, as compared to Fn
- A431 cells displayed increased DECCA on Ln-332, while HT-1080 DECCA values were not significantly changed.

This dynamic multivariate analysis of single-cell motility was a proof-of-principle analysis, demonstrating, how the future of single-cell analysis may progress towards multivariate strategies to be analyzed by the systems biology approach.

## 6.3 Significance

### 6.3.1 Single-cell Assays

The data produced to support this dissertation represents the first analysis and quantitation of epithelial cell speed heterogeneity. Single-cell heterogeneity

is currently of great interest in the field of cell biology (Feinerman *et al.*, 2008; Spencer *et al.*, 2009; Brock *et al.*, 2009; Fraser *et al.*, 2009; Colman-Lerner *et al.*, 2005; Gascoigne *et al.*, 2008; Cang *et al.*, 2008). Many researchers are realizing that the future of cellular biology will be defined by the ability to understand what is going on at the single-cell level, since "*biology at the single-cell level sharply diverges from expectations*" (Levsky and Singer, 2003). This body of work outlines the first effort to quantify the inherent motility heterogeneity of epithelial cells. As single-cell videomicroscopy-obtained metrics become automatically quantifiable by computer algorithms, multivariate analysis will become a standard practice, much as wound healing and Boyden chamber assays have been for decades. Commercially-available software currently exists which is capable of obtaining multiple parameters per cell, such as: Surface area, eccentricity, circumference, length, and intensity (Carpenter *et al.*, 2006). The component missing from this software for true single-cell, dynamic multivariate analysis is accurate cell tracking.

The bottleneck of single-cell epithelial motility analysis has always been tracking the cells manually. However, we are on the verge of being able to automatically track single-cells using H2B-RPF labeled nuclei for florescent images, and using advanced cell segmenting techniques for phase contrast images (Walter Georgescu, personal communication). When tracking becomes fully automated, there will be a vast increase in data of all cell lines and a multitude of media conditions, drugs, and ECMs. It is my intent that this work serve as a benchmark for analyzing the heterogeneity of motility behavior, and

that the metrics developed in this dissertation will be commonly used to quantitate motility in the near future.

As technology progresses, it will be necessary to integrate numerous assays and techniques into a powerful software application allowing acquisition of multivariate data. One example of such a technique can be found in CellProfiler, an open source, MATLAB based program designed for high-content microscopy images (Carpenter *et al.*, 2006). Currently this software is limited to end-point assays due to a lack of tracking software, but this hurdle will soon be overcome. Software for data visualization is also necessary. Looking at 6 cell lines, in 6 conditions, with 20 metrics produces data with 720 dimensions. A challenge of these datasets will be to make them human-interpretable. A few possibilities exist. The Althchuler and Wu laboratory currently employ the principal-component analysis method to reduce dimensionality of their data to isolate subpopulations. They have successfully reduced a 1,536 dimension dataset to 25 dimensions, without losing predictive power (Slack *et al.*, 2008). Thus, the software is becoming available to search for subpopulations in enormous datasets.

### 6.3.2 Influence of Variability on Other Phenotypic Traits

The discovery of non-genetic phenotypic heterogeneity of cell speeds suggests that other traits may also demonstrate non-genetic heterogeneity. In fact, research on single-cell variability of mRNA and protein levels indicates that heterogeneity of phenotypic traits may be widespread (Bar-

Even *et al.*, 2006; Raser and O'Shea, 2005; Samoilov *et al.*, 2006). Colman-Lerner *et al.* have demonstrated cell-to-cell variation of yeast mating pheromone response. It appears likely that other quantitative single-cell metrics will also be highly heterogeneous in cell populations. These other metrics could be studied with the statistical analyses presented here to quantitate their heterogeneity. Thus, the statistical analyses and multivariate techniques developed here can be applied to many types of analyses from other fields, including proliferation, metabolism, and signaling.

### 6.3.3 Phenotypic Plasticity

Quantitation of plasticity (demonstrated by the ability to maintain variation in a variety of microenvironments) is a clear next step of this research. One could hypothesize that cancer cells demonstrate a higher level of plasticity than non-tumorigenic cells. Increased plasticity would allow cancer cells to thrive in many microenvironmental conditions, while non-tumorigenic cells would lose their phenotypic hetereogeneity. This hypothesis could be tested by quantitation of a number of single-cell metrics for multiple cell lines, in many microenvironmental conditions. Cells demonstrating high variation of quantitative metrics in a large percentage of conditions would be considered highly plastic. Quantitation would be achieved by a frequency histogram of binned metric ranges. The more the histogram is shifted to the right, the greater a cell line's plasticity. It would be interesting to compare the plasticity of several cancer cell lines, to determine if plasticity is cell line dependant, or if it is a trait of all cancers.

Other questions could be analyzed by this type of analysis: 1) does the plasticity of cancer cell lines increase as the number of passages *in vivo* increases? 2) does plasticity increase upon multiple exposures to drugs or other forms of a harsh microenvironment? 3) does plasticity correlate with invasive/metastatic potential?

Going one step further, plasticity could be calculated and visualized in a human-interpretable manner for dozens of metrics, for dozens of cell lines, in hundreds of microenvironmental conditions. All of this data (hundreds or thousands of dimensions), could be visualized in a single histogram per cell line. Average values from the frequency histograms mentioned above (for metric ranges) would then themselves be represented on a frequency histogram. Peaks at a high range indicate high plasticity, while a distribution at lower ranges is indicative of low plasticity.

An analysis of plasticity would bring my multivariate, single-cell, dynamic analysis to the next level, and would lead to advanced understanding of how heterogeneity plays a role in cancer progression.

### 6.3.4 Relevance to Cancer Progression

For decades, the clonal theory of cancer has predominated (see Section 1.1.1). However, recent research suggests that non-genetic heterogeneity may also factor into cancer progression (Brock *et al.*, 2009). From this research, the non-genetic heterogeneity theory of cancer has arisen. This theory suggests that cancer cells may evolve the ability to adapt to their environment through

phenotypic heterogeneity brought about by gene expression variability (Fraser and Kaern, 2009; Brock *et al.*, 2009). In essence, this theory proposes that cancer cells can evolve through both genetic and non-genetic means into a highly variable state. Those cells that are more variable are then more likely to survive the variety of conditions they meet within a tumor, which can include: hypoxia, fluctuating growth factors, immune system response, ECM alteration, tissue restructuring, and drug treatment. By evolving the ability to obtain a highly heterogeneous population from a single genotype, a clonal population of cancer cells will become more likely to survive a selective environment for a limited duration (Brock, *et al.*, 2009). A comparison between the clonal theory and the non-genetic heterogeneity theory is outlined in Figure 6.2.

It is likely that both the clonal theory and the non-genetic heterogeneity theory play a role in cancer development. For example, continuous, sustained environmental changes may select for rare mutants and lead to a clonal population (*e.g.* for the case of chemotherapy regimes), while transient changes can be handled by the non-genetic heterogeneity present within the tumor. It is important for cancer therapy to determine which theory predominates during cancer progression for different disease models. This end is achievable by continued development of single-cell heterogeneity analyses.

### 6.3.5 Distinguishing cancer from non-cancer

New metrics have been developed which, in this system, can distinguish cancer cells from non-tumorigenic cells *via* motility assays. This has broad

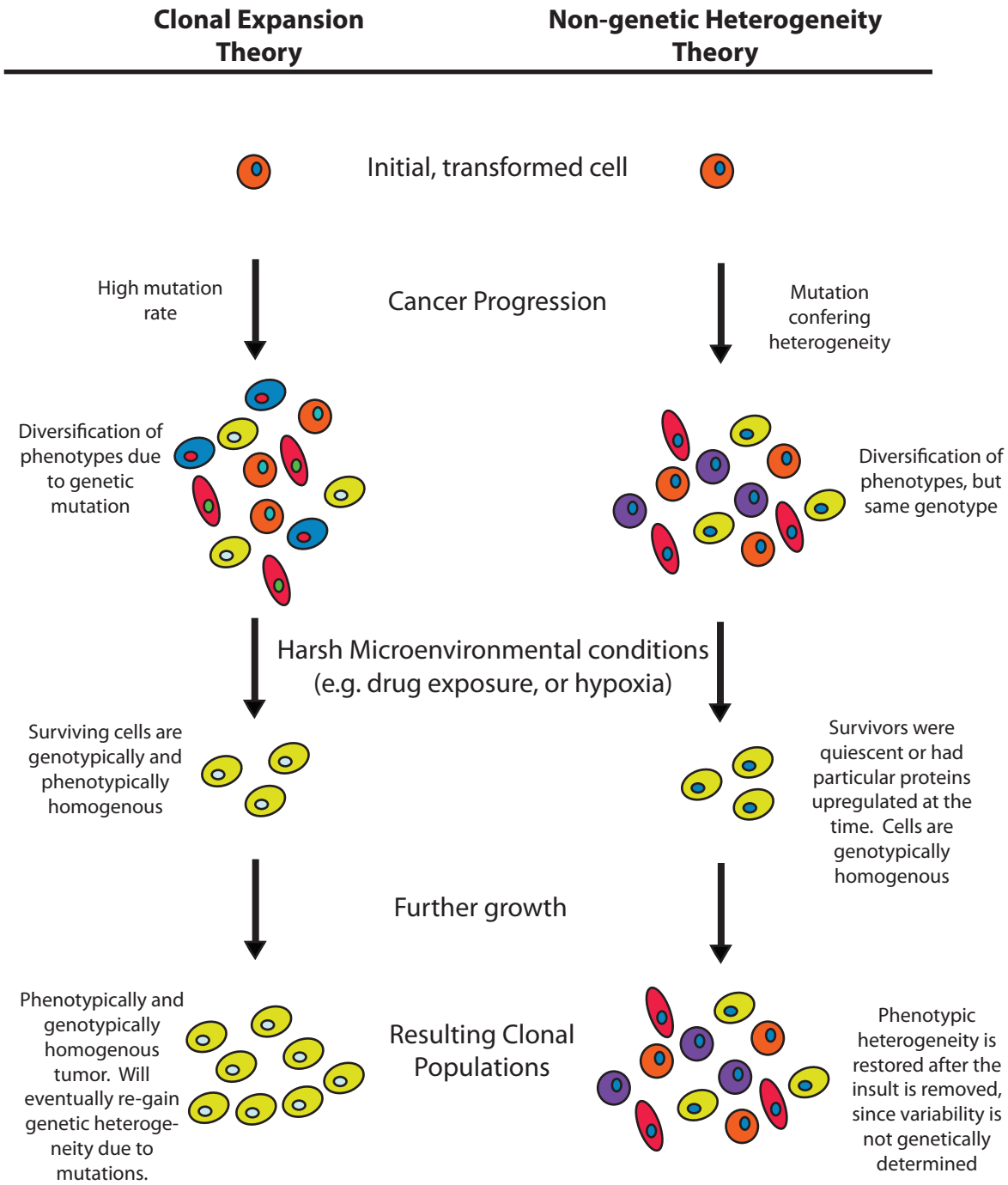



**Figure 6.1- Non-genetic heterogeneity theory of cancer.**

implications for cancer modeling, and suggests we may be able to determine the mechanism driving the alteration of cancer cell motility, since we now have the assays and metrics to measure these differences. For example, drug inhibitors of different areas of the migration pathway could be utilized to pinpoint the mechanism behind heterogeneity of motility.

The researched presented here indicates that variability and heterogeneity may be a stronger predictor of cancer *versus* non-cancer than other methods such as cell speed, proliferation, or ability to form colonies in soft agar. Thus, single-cell analysis may one day be used for clinical diagnosis and/or prognosis. Phenotypic signatures may exist which could be found to correlate with disease progression, metastasis, or specific mutations. Alternatively, single-cell analyses could be used to screen patient tumor biopsies for drug sensitivity. Significant challenges would need to be overcome to move this type of analysis into the pathology laboratory, including:

- clear phenotypic signatures that are predictive of clinical outcome
- automated data analysis
- ability to extract and efficiently culture cancer cells from a tumor
- faster turnaround time for results

In total, results from Chapters 3 and 4 have broad implications in cancer research:

- there is significant heterogeneity of phenotypic response even upon transient exposure to changing environments (response was tested from 2-6 hours)
- Cancer cell variability and heterogeneity of motility is altered upon micro-environmental change
- The changes in variability and heterogeneity in response to serum and EGF depletion of CA1d cells are not due to genetic heterogeneity

### 6.4 Questions Raised

This project has raised a number of questions which can be addressed in the near future:

- *Do phenotypic subpopulations exist, or is there a continuum of cell states?*

Qualitative assumptions must be made in order to separate out subpopulations. There must be a reason for stating that one group of cells is different from another. Separating out subpopulations due to speed, surface area, or any other metric is ineffective unless there is a phenotypic difference that is experimentally or clinically relevant. Thus, the test of whether a subpopulation is truly present is the presence of a clear reason behind differentiating two groups (*i.e.* drug sensitivity, patient life expectancy, or disease progression). Any subpopulations determined by algorithms must be validated through experimental observations.

- *Can cells switch from one subpopulation to another over time?*

Most studies aimed at distinguishing subpopulations do so through endpoint assays. However, it is unknown if an individual cell belonging to one subpopulation is destined to remain within that particular subpopulation for its lifetime. It is possible that cells can switch from one group to another. This question can be answered through dynamic multivariate single-cell assays.

- *what happens at longer time points?*

The research presented here only looks at a short term response to microenvironmental change. It is currently unknown if this response will be maintained over longer time periods. This question can be addressed in the near-term by utilizing automatic tracking software currently being developed.

- *Is the study of variability applicable in 3D systems or in vivo?*

Condeelis *et al.,* have demonstrated that single-cell analysis is possible *in vivo* by the use of multi-photon microscopy (Condeelis and Segall, 2003). This type of analysis may be a valuable tool for cancer research, but requires the development of algorithms and tools to automate quantification in their model in a much more high-throughput manner. Current quantitation methods for 3D or *in vivo* motility studies are much less advanced than those for 2D.

- *What happens at higher cell concentrations?*

The research presented looks only at sparsely plated cells, to minimize the effect of cell-cell contact, and to reduce the dataset to a reasonable size. Cell segmentation and tracking becomes more complicated as the cell density increases. However, the use of H2B-RFP to visualize the nucleus, coupled with advanced cell segmentation algorithms, will allow us to address trends between motility metrics and cell density.

## 6.5 Concluding Remarks

Effective treatment of cancer will require both the development of new methods of therapeutic intervention, and also the ability to determine which patients will receive the most benefit from specific types of intervention. Development of targeted treatments for cancer will be aided through technological and scientific insights into the biological mechanism driving cancer cell heterogeneity. The work presented in this dissertation provides important additions to the study of single-cell variability, and provides robust tools for future investigations in this field.

# REFERENCES


Aharon, R. and Z. Bar-Shavit, *Involvement of aquaporin 9 in osteoclast differentiation.* J Biol Chem, 2006. **281**(28): p. 19305-9.

Alexopoulos, L.G., G.R. Erickson, and F. Guilak, *A method for quantifying cell size from differential interference contrast images: validation and application to osmotically stressed chondrocytes.* J Microsc, 2002. **205**(Pt 2): p. 125-35.

Anderson, A.R. and V. Quaranta, *Integrative mathematical oncology.* Nat Rev Cancer, 2008. **8**(3): p. 227-34.

Anderson, A.R., et al., *Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment.* Cell, 2006. **127**(5): p. 905-15.

Anderson, G.R., et al., *Intrachromosomal genomic instability in human sporadic colorectal cancer measured by genome-wide allelotyping and inter-(simple sequence repeat) PCR.* Cancer Res, 2001. **61**(22): p. 8274-83.

Aznavoorian, S., et al., *Molecular aspects of tumor cell invasion and metastasis.* Cancer, 1993. **71**(4): p. 1368-83.

Bahnson, A., et al., *Automated measurement of cell motility and proliferation.* BMC Cell Biol, 2005. **6**(1): p. 19.

Bar-Even, A., et al., *Noise in protein expression scales with natural protein abundance.* Nat Genet, 2006. **38**(6): p. 636-43.

Bear, J.E., et al., *Antagonism between Ena/VASP proteins and actin filament capping regulates fibroblast motility.* Cell, 2002. **109**(4): p. 509-21.

Beltman, J.B., et al., *Lymph node topology dictates T cell migration behavior.* J Exp Med, 2007. **204**(4): p. 771-80.

Bindschadler, M. and J.L. McGrath, *Sheet migration by wounded monolayers as an emergent property of single-cell dynamics.* J Cell Sci, 2007. **120**(Pt 5): p. 876-84.

Bissell, M.J., P.A. Kenny, and D.C. Radisky, *Microenvironmental regulators of tissue structure and function also regulate tumor induction and progression: the role of extracellular matrix and its degrading enzymes.* Cold Spring Harb Symp Quant Biol, 2005. **70**: p. 343-56.

Bissell, M.J. and M.A. Labarge, *Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment?* Cancer Cell, 2005. **7**(1): p. 17-23.

Blake, W.J., et al., *Noise in eukaryotic gene expression.* Nature, 2003. **422**(6932): p. 633-7.

Brock, A., H. Chang, and S. Huang, *Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours.* Nat Rev Genet, 2009. **10**(5): p. 336-42.

Brodt, P., et al., *Studies on clonal heterogeneity in two spontaneously metastasizing mammary carcinomas of recent origin.* Int J Cancer, 1985. **35**(2): p. 265-73.

Bronchud, M.H., *Is cancer really a 'local' cellular clonal disease?* Med Hypotheses, 2002. **59**(5): p. 560-5.

Bryce, N.S., et al., *Cortactin promotes cell motility by enhancing lamellipodial persistence.* Curr Biol, 2005. **15**(14): p. 1276-85.

Bullen, A., *Microscopic imaging techniques for drug discovery.* Nat Rev Drug Discov, 2008. **7**(1): p. 54-67.

Cahill, D.P., et al., *Genetic instability and darwinian selection in tumours.* Trends Cell Biol, 1999. **9**(12): p. M57-60.

Cai, L., et al., *Coronin 1B coordinates Arp2/3 complex and cofilin activities at the leading edge.* Cell, 2007. **128**(5): p. 915-29.

Carpenter, A.E., et al., *CellProfiler: image analysis software for identifying and quantifying cell phenotypes.* Genome Biol, 2006. **7**(10): p. R100.

Casanovas, O., et al., *Drug resistance by evasion of antiangiogenic targeting of VEGF signaling in late-stage pancreatic islet tumors.* Cancer Cell, 2005. **8**(4): p. 299-309.

Chambers, A.F., A.C. Groom, and I.C. MacDonald, *Dissemination and growth of cancer cells in metastatic sites.* Nat Rev Cancer, 2002. **2**(8): p. 563-72.

Chambers, A.F., et al., *Clinical targets for anti-metastasis therapy.* Adv Cancer Res, 2000. **79**: p. 91-121.

Chambers, A.F. and L.M. Matrisian, *Changing views of the role of matrix metalloproteinases in metastasis.* J Natl Cancer Inst, 1997. **89**(17): p. 1260-70.

Chambers, A.F., et al., *Molecular biology of breast cancer metastasis. Clinical implications of experimental studies on metastatic inefficiency.* Breast Cancer Res, 2000. **2**(6): p. 400-7.

Champernowne, D., *A model of income distribution.* Economic Journal, 1953. **63**: p. 318-351.

Chang, H.H., et al., *Transcriptome-wide noise controls lineage choice in mammalian progenitor cells.* Nature, 2008. **453**(7194): p. 544-7.

Chicurel, M., *Cell biology. Cell migration research is on the move.* Science, 2002. **295**(5555): p. 606-9.

Chon, J.H., et al., *Characterization of single-cell migration using a computer-aided fluorescence time-lapse videomicroscopy system.* Anal Biochem, 1997. **252**(2): p. 246-54.

Chow, M. and H. Rubin, *Clonal dynamics of progressive neoplastic transformation.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6976-81.

Christofori, G., *New signals from the invasive front.* Nature, 2006. **441**(7092): p. 444-50.

Cifone, M.A. and I.J. Fidler, *Increasing metastatic potential is associated with increasing genetic instability of clones isolated from murine neoplasms.* Proc Natl Acad Sci U S A, 1981. **78**(11): p. 6949-52.

Clark, E.A., et al., *Genomic analysis of metastasis reveals an essential role for RhoC.* Nature, 2000. **406**(6795): p. 532-5.

Codling, E.A., M.J. Plank, and S. Benhamou, *Random walk models in biology.* J R Soc Interface, 2008. **5**(25): p. 813-34.

Colman-Lerner, A., et al., *Regulated cell-to-cell variation in a cell-fate decision system.* Nature, 2005. **437**(7059): p. 699-706.

Condeelis, J. and J.E. Segall, *Intravital imaging of cell movement in tumours.* Nat Rev Cancer, 2003. **3**(12): p. 921-30.

Condeelis, J., R.H. Singer, and J.E. Segall, *The great escape: when cancer cells hijack the genes for chemotaxis and motility.* Annu Rev Cell Dev Biol, 2005. **21**: p. 695-718.

Cook, T., et al., *Molecular cloning and characterization of TIEG2 reveals a new subfamily of transforming growth factor-beta-inducible Sp1-like zinc finger-encoding genes involved in the regulation of cell growth.* J Biol Chem, 1998. **273**(40): p. 25929-36.

Cook, W.D. and B.J. Alexander, *Phenotypic effect correlating with loss of a novel tumor suppressor gene: towards cloning by complementation.* Leukemia, 1998. **12**(12): p. 1937-43.

Cram, L.S., et al., *Quantitation of one aspect of karyotype instability associated with neoplastic transformation in Chinese hamster cells.* Prog Nucleic Acid Res Mol Biol, 1983. **29**: p. 39-42.

Cristini, V., et al., *Morphologic instability and cancer invasion.* Clin Cancer Res, 2005. **11**(19 Pt 1): p. 6772-9.

Dawson, P.J., et al., *MCF10AT: a model for the evolution of cancer from proliferative breast disease.* Am J Pathol, 1996. **148**(1): p. 313-9.

Dexter, D.L. and J.T. Leith, *Tumor heterogeneity and drug resistance.* J Clin Oncol, 1986. **4**(2): p. 244-57.

DiMilla, P.A., et al., *Maximal migration of human smooth muscle cells on fibronectin and type IV collagen occurs at an intermediate attachment strength.* J Cell Biol, 1993. **122**(3): p. 729-37.

Dormann, D. and C.J. Weijer, *Imaging of cell migration.* Embo J, 2006. **25**(15): p. 3480-93.

Dove, A., *Membrane Specialization.* J Cell Biol, 1999. **145**(2).

Dunn, G.A. and A.F. Brown, *A unified approach to analysing cell motility.* J Cell Sci Suppl, 1987. **8**: p. 81-102.

Dutrillaux, B., *Pathways of chromosome alteration in human epithelial cancers.* Adv Cancer Res, 1995. **67**: p. 59-82.

Edwards, A.M., et al., *Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer.* Nature, 2007. **449**(7165): p. 1044-8.

Egger G., et al., *Epigenetics in human disease and prospects for epigenetic therapy.* Nature, 2004. **429**: p457-63.

Elmore, S., *Apoptosis: a review of programmed cell death.* Toxicol Pathol, 2007. **35**(4): p. 495-516.

Elowitz, M.B., et al., *Stochastic gene expression in a single cell.* Science, 2002. **297**(5584): p. 1183-6.
Feinerman, O., et al., *Variability and robustness in T cell activation from regulated heterogeneity in protein levels.* Science, 2008. **321**(5892): p. 1081-4.

Feinberg AP, Tycko B, *The history of cancer epigenetics.* Nat. Rev. Cancer, 2004. **4**: 143-153.

Fidler, I.J., *The organ microenvironment and cancer metastasis.* Differentiation, 2002. **70**(9-10): p. 498-505.

Fidler, I.J., *Critical determinants of metastasis.* Semin Cancer Biol, 2002. **12**(2): p. 89-96.

Fidler, I.J., et al., *Demonstration of multiple phenotypic diversity in a murine melanoma of recent origin.* J Natl Cancer Inst, 1981. **67**(4): p. 947-56.

Fidler, I.J. and I.R. Hart, *Biological diversity in metastatic neoplasms: origins and implications.* Science, 1982. **217**(4564): p. 998-1003.

Fidler, I.J. and I.R. Hart, *The development of biological diversity and metastatic potential in malignant neoplasms.* Oncodev Biol Med, 1982. **4**(1-2): p. 161-76.

Fidler, I.J. and I.R. Hart, *Recent observations on the pathogenesis of cancer metastasis.* Prog Clin Biol Res, 1982. **85 Pt B**: p. 601-19.

Fidler, I.J., et al., *The seed and soil hypothesis: vascularisation and brain metastases.* Lancet Oncol, 2002. **3**(1): p. 53-7.

Fraser, D. and M. Kaern, *A chance at survival: gene expression noise and phenotypic diversification strategies.* Mol Microbiol, 2009. **71**(6): p. 1333-40.

Fraser, H.B., et al., *Noise minimization in eukaryotic gene expression.* PLoS Biol, 2004. **2**(6): p. e137.

Fukui, Y., et al., *Cell behavior and actomyosin organization in Dictyostelium during substrate exploration.* Cell Struct Funct, 1991. **16**(4): p. 289-301.

Fürth, R., *Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung, Mit Anvendunen auf die Bewegung lebender Infusorien. Z*. Physik, 1920. **2**: p. 244-256.

Gabaix, X., *Zipf's law for cities: an explanation.* Quarterly Journal of Economics, 1999. **114**: p. 739-767.

Gascoigne, K.E. and S.S. Taylor, *Cancer cells display profound intra- and interline variation following prolonged exposure to antimitotic drugs.* Cancer Cell, 2008. **14**(2): p. 111-22.

Gautestad, A.O. and I. Mysterud, *Intrinsic scaling complexity in animal dispersion and abundance.* Am Nat, 2005. **165**(1): p. 44-55.
Geho, D.H., et al., *Physiological mechanisms of tumor-cell invasion and migration.* Physiology (Bethesda), 2005. **20**: p. 194-200.

Georgescu, W., et al., *Model-controlled hydrodynamic focusing to generate multiple overlapping gradients of surface-immobilized proteins in microfluidic devices.* Lab Chip, 2008. **8**(2): p. 238-44.

Giaretti, W., et al., *Near-diploid and near-triploid human sporadic colorectal adenocarcinomas differ for KRAS2 and TP53 mutational status.* Genes Chromosomes Cancer, 2003. **37**(2): p. 207-13.

Glory, E. and R.F. Murphy, *Automated subcellular location determination and high-throughput microscopy.* Dev Cell, 2007. **12**(1): p. 7-16.

Goswami, S., et al., *Identification of invasion specific splice variants of the cytoskeletal protein Mena present in mammary tumor cells during invasion in vivo.* Clin Exp Metastasis, 2009. **26**(2): p. 153-9.

Grossmann, M.E. and D.J. Tindall, *The androgen receptor is transcriptionally suppressed by proteins that bind single-stranded DNA.* J Biol Chem, 1995. **270**(18): p. 10968-75.

Guan, J.L., *Cell biology. Integrins, rafts, Rac, and Rho.* Science, 2004. **303**(5659): p. 773-4.

Gupta, C., A.K. Malani, and S. Rangineni, *Breast metastasis of ilial carcinoid tumor: case report and literature review.* World J Surg Oncol, 2006. **4**: p. 15.

Gupta, G.P., et al., *Identifying site-specific metastasis genes and functions.* Cold Spring Harb Symp Quant Biol, 2005. **70**: p. 149-58.

Gupta, S.G., et al., *Sentinel lymph node biopsy for evaluation and treatment of patients with Merkel cell carcinoma: The Dana-Farber experience and meta-analysis of the literature.* Arch Dermatol, 2006. **142**(6): p. 685-90.

Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

Harris, M.P., et al., *Migration of isogenic cell lines quantified by dynamic multivariate analysis of single-cell motility.* Cell Adh Migr, 2008. **2**(2): p. 127-36.

Hermsen, M., et al., *Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability.* Gastroenterology, 2002. **123**(4): p. 1109-19.

Hinz, B., et al., *Quantifying lamella dynamics of cultured cells by SACED, a new computer-assisted motion analysis.* Exp Cell Res, 1999. **251**(1): p. 234-43.

Horwitz, R. and D. Webb, *Cell migration.* Curr Biol, 2003. **13**(19): p. R756-9.

Ingber, D.E., *Fibronectin controls capillary endothelial cell growth by modulating cell shape.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3579-83.

Ionescu-Zanetti, C., et al., *Alkaline hemolysis fragility is dependent on cell shape: results from a morphology tracker.* Cytometry A, 2005. **65**(2): p. 116-23.

Ionescu-Zanetti, C., et al., *Alkaline hemolysis fragility is dependent on cell shape: results from a morphology tracker.* Cytometry A, 2005. **65**(2): p. 116-23.

Jackson, D. and Y. Kagan, *Testable earthquake forecasts for 1999.* Seismological Research Letters, 1999. **70**(4): p. 393-403.

Jones, T.R., et al., *Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning.* Proc Natl Acad Sci U S A, 2009. **106**(6): p. 1826-31.

Joyce, J.A., *Therapeutic targeting of the tumor microenvironment.* Cancer Cell, 2005. **7**(6): p. 513-20.

Juhasz, I., et al., *Growth and invasion of human melanomas in human skin grafted to immunodeficient mice.* Am J Pathol, 1993. **143**(2): p. 528-37.

Keese, C.R., et al., *Electrical wound-healing assay for cells in vitro.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1554-9.

Kellermayer, M.S., et al., *Stepwise dynamics of epitaxially growing single amyloid fibrils.* Proc Natl Acad Sci U S A, 2008. **105**(1): p. 141-4.

Kemkemer, R., et al., *Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13783-8.

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

Kim, H.D., et al., *Epidermal growth factor-induced enhancement of glioblastoma cell migration in 3D arises from an intrinsic increase in speed but an extrinsic matrix- and proteolysis-dependent increase in persistence.* Mol Biol Cell, 2008. **19**(10): p. 4249-59.

Kipper, M.J., H.K. Kleinman, and F.W. Wang, *New method for modeling connective-tissue cell migration: improved accuracy on motility parameters.* Biophys J, 2007. **93**(5): p. 1797-808.

Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.

Kraemer, P.M., et al., *Spontaneous neoplastic evolution of Chinese hamster cells in culture: multistep progression of phenotype.* Cancer Res, 1983. **43**(10): p. 4822-7.

Kumar, Abbas, and Fausto, *Robbins & Cotran Pathologic Basis of Disease*. 7th ed. 2004: Saunders.

Lamprecht, M.R., D.M. Sabatini, and A.E. Carpenter, *CellProfiler: free, versatile software for automated biological image analysis.* Biotechniques, 2007. **42**(1): p. 71-5.

Lauffenburger, D.A., *Models for receptor-mediated cell phenomena: adhesion and migration.* Annu Rev Biophys Biophys Chem, 1991. **20**: p. 387-414.

Lauffenburger, D.A. and A.F. Horwitz, *Cell migration: a physically integrated molecular process.* Cell, 1996. **84**(3): p. 359-69.

Lee, Y., et al., *Characterization of endothelial cell locomotion using a Markov chain model.* Biochem Cell Biol, 1995. **73**(7-8): p. 461-72.

Levsky, J.M. and R.H. Singer, *Gene expression and the myth of the average cell.* Trends Cell Biol, 2003. **13**(1): p. 4-6.

Lin, D.M., et al., *Characterization of mRNA expression in single neurons.* Methods Mol Biol, 2007. **399**: p. 133-52.

Liotta, L.A. and E.C. Kohn, *The microenvironment of the tumour-host interface.* Nature, 2001. **411**(6835): p. 375-9.

Liotta, L.A., C.N. Rao, and S.H. Barsky, *Tumor invasion and the extracellular matrix.* Lab Invest, 1983. **49**(6): p. 636-49.

Liotta, L.A., et al., *Tumor cell attachment and degradation of basement membranes.* Symp Fundam Cancer Res, 1983. **36**: p. 169-76.

Loo, L.H., L.F. Wu, and S.J. Altschuler, *Image-based multivariate profiling of drug responses from single cells.* Nat Methods, 2007. **4**(5): p. 445-53.

Lovelady, D.C., et al., *Distinguishing cancerous from noncancerous cells through analysis of electrical noise.* Phys Rev E Stat Nonlin Soft Matter Phys, 2007. **76**(4 Pt 1): p. 041908.

Marquet, P.A., et al., *Scaling and power-laws in ecological systems.* J Exp Biol, 2005. **208**(Pt 9): p. 1749-69.

McCarty, M.F., et al., *Evidence for the causal role of endogenous interferon-alpha/beta in the regulation of angiogenesis, tumorigenicity, and metastasis of cutaneous neoplasms.* Clin Exp Metastasis, 2002. **19**(7): p. 609-15.

Miller, F.R., et al., *Xenograft model of progressive human proliferative breast disease.* J Natl Cancer Inst, 1993. **85**(21): p. 1725-32.

Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung.* Nature, 2005. **436**(7050): p. 518-24.

Mitsumoto, M., et al., *Emergence of higher levels of invasive and metastatic properties in the drug resistant cancer cell lines after the repeated administration of cisplatin in tumor-bearing mice.* J Cancer Res Clin Oncol, 1998. **124**(11): p. 607-14.

Mueller-Rath, R., et al., *In vivo cultivation of human articular chondrocytes in a nude mouse-based contained defect organ culture model.* Biomed Mater Eng, 2007. **17**(6): p. 357-66.

Mukherjee, D.P., N. Ray, and S.T. Acton, *Level set analysis for leukocyte detection and tracking.* IEEE Trans Image Process, 2004. **13**(4): p. 562-72.

Nixon, M. and A. Aguado, *Feature Extraction and Image Processing*. 2nd ed. 2007: Academic Press.

Nowell, P.C., *The clonal evolution of tumor cell populations.* Science, 1976. **194**(4260): p. 23-8.

Pedraza, J.M. and A. van Oudenaarden, *Noise propagation in gene networks.* Science, 2005. **307**(5717): p. 1965-9.

Pennacchietti, S., et al., *Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene.* Cancer Cell, 2003. **3**(4): p. 347-61.

Pepperkok, R. and J. Ellenberg, *High-throughput fluorescence microscopy for systems biology.* Nat Rev Mol Cell Biol, 2006. **7**(9): p. 690-6.

Poste, G., J. Doll, and I.J. Fidler, *Interactions among clonal subpopulations affect stability of the metastatic phenotype in polyclonal populations of B16 melanoma cells.* Proc Natl Acad Sci U S A, 1981. **78**(10): p. 6226-30.

Potdar, A.A., et al., *Bimodal analysis of mammary epithelial cell migration in two dimensions.* Ann Biomed Eng, 2009. **37**(1): p. 230-45.

Quaranta, V., *Motility cues in the tumor microenvironment.* Differentiation, 2002. **70**(9-10): p. 590-8.

Quaranta, V. and G. Giannelli, *Cancer invasion: watch your neighbourhood!* Tumori, 2003. **89**(4): p. 343-8.

Quaranta, V., et al., *Invasion emerges from cancer cell adaptation to competitive microenvironments: quantitative predictions from multiscale mathematical models.* Semin Cancer Biol, 2008. **18**(5): p. 338-48.

Quaranta, V., et al., *Mathematical modeling of cancer: the future of prognosis and treatment.* Clin Chim Acta, 2005. **357**(2): p. 173-9.

Rabinovitch, P.S., et al., *Pancolonic chromosomal instability precedes dysplasia and cancer in ulcerative colitis.* Cancer Res, 1999. **59**(20): p. 5148-53.

Raser, J.M. and E.K. O'Shea, *Control of stochasticity in eukaryotic gene expression.* Science, 2004. **304**(5678): p. 1811-4.

Raser, J.M. and E.K. O'Shea, *Noise in gene expression: origins, consequences, and control.* Science, 2005. **309**(5743): p. 2010-3.

Ray, N. and S.T. Acton, *Data acceptance for automated leukocyte tracking through segmentation of spatiotemporal images.* IEEE Trans Biomed Eng, 2005. **52**(10): p. 1702-12.

Reed, W., *The Pareto, Zipf and other power laws.* Economics Letters, 2001. **74**(1): p. 15-19.

Risques, R.A., M. Ribas, and M.A. Peinado, *Assessment of cumulated genetic alterations in colorectal cancer.* Histol Histopathol, 2003. **18**(4): p. 1289-99.

Rosenfeld, N., et al., *Gene regulation at the single-cell level.* Science, 2005. **307**(5717): p. 1962-5.

Roy, P., et al., *Microscope-based techniques to study cell adhesion and migration.* Nat Cell Biol, 2002. **4**(4): p. E91-6.

Samoilov, M.S., G. Price, and A.P. Arkin, *From fluctuations to phenotypes: the physiology of noise.* Sci STKE, 2006. **2006**(366): p. re17.

Santos-Rosa H, Caldas C, *Chromatin modifier enzymes, the histone code and cancer*. Eur. J. Cancer, 2005. **41**: 2381-2402.

Santner, S.J., et al., *Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells.* Breast Cancer Res Treat, 2001. **65**(2): p. 101-10.

Schoenberg , F., R. Roger Peng, and J. Woods, *On the distribution of wildfire sizes.* Environmetrics, 2003. **14**(6): p. p583-592.

Sheetz, M.P., *Cell migration by graded attachment to substrates and contraction.* Semin Cell Biol, 1994. **5**(3): p. 149-55.

Sherwood, D.R., *Cell invasion through basement membranes: an anchor of understanding.* Trends Cell Biol, 2006. **16**(5): p. 250-6.

Shlesinger, M.F., *Mathematical physics: search research.* Nature, 2006. **443**(7109): p. 281-2.

Sieber, O.M., et al., *Analysis of chromosomal instability in human colorectal adenomas with two mutational hits at APC.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16910-5.

Sims, D.W., et al., *Scaling laws of marine predator search behaviour.* Nature, 2008. **451**(7182): p. 1098-102.

Slack, M.D., et al., *Characterizing heterogeneous cellular responses to perturbations.* Proc Natl Acad Sci U S A, 2008. **105**(49): p. 19306-11.

Smalley, K.S., P.A. Brafford, and M. Herlyn, *Selective evolutionary pressure from the tissue microenvironment drives tumor progression.* Semin Cancer Biol, 2005. **15**(6): p. 451-9.

Soll, D.R., *The use of computers in understanding how animal cells crawl.* Int Rev Cytol, 1995. **163**: p. 43-104.

Soll, D.R., et al., *"Dynamic Morphology System": a method for quantitating changes in shape, pseudopod formation, and motion in normal and mutant amoebae of Dictyostelium discoideum.* J Cell Biochem, 1988. **37**(2): p. 177-92.

Sporn, M.B., *The war on cancer: a review.* Ann N Y Acad Sci, 1997. **833**: p. 137-46.

Stephens, D.J. and V.J. Allan, *Light microscopy techniques for live cell imaging.* Science, 2003. **300**(5616): p. 82-6.

Stephenson, P., *Back to the drawing board.* Health Serv J, 2003. **113**(5842): p. 14.

Strauss, D. and L. Bednar, *Do one percent of forest fires cause ninety-nine percent
of the damage?* Forest Science, 1989. **35**(2): p. 319–328.

Subkhankulova, T., M.J. Gilchrist, and F.J. Livesey, *Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells.* BMC Genomics, 2008. **9**: p. 268.

Swain, P.S., M.B. Elowitz, and E.D. Siggia, *Intrinsic and extrinsic contributions to stochasticity in gene expression.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12795-800.

Tomelleri, C., et al., *A quantitative study of growth variability of tumour cell clones in vitro.* Cell Prolif, 2008. **41**(1): p. 177-91.

Toraason, M., et al., *Automated surface area measurement of cultured cardiac myocytes.* Cytotechnology, 1990. **4**(2): p. 155-61.

Truskey, G.A. and T.L. Proulx, *Quantitation of cell area on glass and fibronectin-coated surfaces by digital image analysis.* Biotechnol Prog, 1990. **6**(6): p. 513-9.

Uhlenbeck, G.E., Ornstein, L.S., *On the Theory of the Brownian Motion.* Physical Review, 1930. **36**: p. p823-841.

Vere-Jones, D., R. Robinson, and Y. W, *Remarks on the accelerated moment release
model: problems of model formulation, simulation and estimation.* Geophysics Journal, 2001. **144**: p. 517–531.

Viswanathan, G.M., et al., *Levy flight search patterns of wandering albatrosses.* Nature, 1996. **381**(5481): p. 413-415.

Viswanathan, G.M., et al., *Optimizing the success of random searches.* Nature, 1999. **401**(6756): p. 911-4.

Wallin, A.E., A. Salmi, and R. Tuma, *Step length measurement--theory and simulation for tethered bead constant-force single molecule assay.* Biophys J, 2007. **93**(3): p. 795-805.

Wang, N. and D.E. Ingber, *Control of cytoskeletal mechanics by extracellular matrix, cell shape, and mechanical tension.* Biophys J, 1994. **66**(6): p. 2181-9.

Wang, W., et al., *Identification and testing of a gene expression signature of invasive carcinoma cells within primary mammary tumors.* Cancer Res, 2004. **64**(23): p. 8585-94.

Watanabe, S., et al., *A new model to study repair of gastric mucosa using primary cultured rabbit gastric epithelial cells.* J Clin Gastroenterol, 1995. **21 Suppl 1**: p. S40-4.

Weaver, A.K., V.C. Bomben, and H. Sontheimer, *Expression and function of calcium-activated potassium channels in human glioma cells.* Glia, 2006. **54**(3): p. 223-33.

Weaver, A.M., *Invadopodia: specialized cell structures for cancer invasion.* Clin Exp Metastasis, 2006. **23**(2): p. 97-105.

Webb, D.J., C.M. Brown, and A.F. Horwitz, *Illuminating adhesion complexes in migrating cells: moving toward a bright future.* Curr Opin Cell Biol, 2003. **15**(5): p. 614-20.

Webb, D.J. and A.F. Horwitz, *New dimensions in cell migration.* Nat Cell Biol, 2003. **5**(8): p. 690-2.

Weidt, C., et al., *Differential effects of culture conditions on the migration pattern of stromal cell-derived factor-stimulated hematopoietic stem cells.* Stem Cells, 2004. **22**(6): p. 890-6.

Wessels, D., S. Kuhl, and D.R. Soll, *Application of 2D and 3D DIAS to motion analysis of live cells in transmission and confocal microscopy imaging.* Methods Mol Biol, 2006. **346**: p. 261-79.

Wessels, J.T., R.M. Hoffman, and F.S. Wouters, *The use of transgenic fluorescent mouse strains, fluorescent protein coding vectors, and innovative imaging techniques in the life sciences.* Cytometry A, 2008. **73**(6): p. 490-1.

West, G.B., J.H. Brown, and B.J. Enquist, *A general model for the origin of allometric scaling laws in biology.* Science, 1997. **276**(5309): p. 122-6.

Wiener, N., *Cybernetics: Or the Control and Communication in the Animal and the Machine*. 1948, Cambridge, MA: The MIT Press.

Winterwood, N.E., et al., *A critical role for tetraspanin CD151 in alpha3beta1 and alpha6beta4 integrin-dependent tumor cell functions on laminin-5.* Mol Biol Cell, 2006. **17**(6): p. 2707-21.

Wolf, K., et al., *Compensation mechanism in tumor cell migration: mesenchymal-amoeboid transition after blocking of pericellular proteolysis.* J Cell Biol, 2003. **160**(2): p. 267-77.

Woodhouse, E.C., R.F. Chuaqui, and L.A. Liotta, *General mechanisms of metastasis.* Cancer, 1997. **80**(8 Suppl): p. 1529-37.

Xu-van Opstal, W.Y., et al., *Automatic cell culture quantitation with TRAKCELL: application to cell toxicology and differentiation.* Cell Biol Toxicol, 1994. **10**(5-6): p. 387-92.

Xu-van Opstal, W.Y., et al., *Automatic cell culture quantitation with TRAKCELL: application to cell toxicology and differentiation.* Cell Biol Toxicol, 1994. **10**(5-6): p. 387-92.

Yarrow, J.C., et al., *Screening for cell migration inhibitors via automated microscopy reveals a Rho-kinase inhibitor.* Chem Biol, 2005. **12**(3): p. 385-95.

Yu, J., et al., *Probing gene expression in live cells, one protein molecule at a time.* Science, 2006. **311**(5767): p. 1600-3.

Zemljic Jokhadar, S., et al., *The effect of substrate and adsorbed proteins on adhesion, growth and shape of CaCo-2 cells.* Cell Biol Int, 2007. **31**(10): p. 1097-108.

Zemljic Jokhadar, S., et al., *The effect of substrate and adsorbed proteins on adhesion, growth and shape of CaCo-2 cells.* Cell Biol Int, 2007. **31**(10): p. 1097-108.

Zemljic Jokhadar, S., et al., *The effect of substrate and adsorbed proteins on adhesion, growth and shape of CaCo-2 cells.* Cell Biol Int, 2007. **31**(10): p. 1097-108.

Zhang K, Dent SY. *Histone modifying enzymes and cancer: going beyond histones.* J. Cell. Biochem, 2005. **96**: 1137-44.

Zimmer, C. and J.C. Olivo-Marin, *Coupled parametric active contours.* IEEE Trans Pattern Anal Mach Intell, 2005. **27**(11): p. 1838-42.

Zimmer, D., et al., *Determination of deltamethrin residues in plant materials by liquid chromatography/tandem mass spectrometry with electrospray ionization.* J AOAC Int, 2006. **89**(3): p. 786-96.

Zouboulis, C.C., et al., *Progressive differentiation of human sebocytes in vitro is characterized by increasing cell size and altering antigen expression and is regulated by culture duration and retinoids.* Exp Dermatol, 1994. **3**(4): p. 151-60.

Zygourakis, K. and P.A. Markenscoff, *Computer-aided design of bioerodible devices with optimal release characteristics: a cellular automata approach.* Biomaterials, 1996. **17**(2): p. 125-35.