MODELING AND TARGETING SIGNAL TRANSDUCTION PATHWAYS GOVERNING CELL MIGRATION

by

Sourabh Prakash Kharait

MD, University of Pune, India, 2001

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH
SCHOOL OF MEDICINE

This dissertation was presented

by
Sourabh Prakash Kharait

It was defended on

April 7th, 2006

and approved by

Robert Getzenberg, PhD, Professor of Oncology and Urology, Johns Hopkins University

Chuanyue Wu, PhD, Associate Professor of Pathology, University of Pittsburgh

Douglas Lauffenburger, PhD, Professor of Biological Engineering, Massachusetts Institute of Technology

Dissertation Committee Chair: Satdarshan Pal Singh Monga, MD, Assistant Professor of Pathology, University of Pittsburgh

Dissertation Advisor: Alan Wells, MD, DMS, Thomas Gill Professor and Vice-Chairman of Pathology

MODELING AND TARGETING SIGNAL TRANSDUCTION PATHWAYS GOVERNING CELL MIGRATION

Sourabh Prakash Kharait, MD, PhD.

University of Pittsburgh, 2006

ABSTRACT

Cell migration is a complex biophysical event that is dysregulated in a variety of human diseases including cancer. The ability of tumor cells to migrate enables cancer dissemination causing significant mortality thus making it an important therapeutic target. Motility is exhibited epigenetically by activation of numerous signaling pathways that transmit extracellular cues to the final effectors of cell movement. Such signaling switches are a part of larger and highly complex signaling (proteomic) networks that are under the control of numerous activators or inhibitors. Although majority of the proteins that are 'required' during cell motility have been identified, it is yet unclear wherein they fit within the signaling network to govern motility. Thus, a 'systems biology' approach is needed to understand the complex interplay of signaling cascades in mediating cell motility so that better therapeutic targets can be defined.

We utilized a mathematical modeling approach, called decision tree analysis to map the interplay between five key signaling proteins known to regulate vital biophysical processes of fibroblast motility downstream of EGF receptor activation. Interestingly, our model identified myosin light chain (MLC) mediated cell contractility as a crucial node for maximal motility. Even more non-intuitively the decision tree model predicted that subtotal inhibition of MLC can actually increase motility. Confirmatory experiments with fibroblasts and cancer cells have shown that to be the case.

Since the model proposed that total abrogation of contractility can limit cell migration, we asked if such an intervention can limit tumor invasion. Since PKCδ is implicated in EGF receptor mediated transcellular contractility, we abrogated PKCδ using pharmacological (Rottlerin) and molecular (RNAi) interventions. Such depletion of PKCδ reduced migration as well as invasiveness of prostate carcinoma cells predominantly by decreasing their contractility through myosin light chain (MLC). Additionally, activation of PKCδ correlated with human prostate cancer progression as assessed by immunohistochemistry of prostate tissue sections.

In summation our studies illustrate the importance of quantitative (total versus subtotal) disruption of key signaling nodes in mediating a desired cell response. Novel computational modeling approaches are needed to identify newer molecular switches from existing proteomic networks that can be explored, using classical experimental methods, as therapeutic targets.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Dedication

My work is dedicated to the countless graduate students and postdoctoral researchers around the globe who are at the forefront of scientific discoveries and yet, so many times, are unappreciated.

Acknowledgements

I am extremely grateful to my guru, **Dr. Alan Wells**, who introduced the world of science to me. It was his patience and guidance that enabled the successful completion of this work. His teachings were not restricted to science alone. I have learnt from him, what it takes to be successful and how true success is really defined. I hope in the future, more students are as lucky as I have been in finding such a mentor.

I am grateful to **Dr. Douglas Lauffenburger**, who has really served as my co-advisor during these four years. He has been extremely receptive to my constant barrage of questions regarding the mathematical modeling experiments. Along with him, other members of his lab, particularly **Kirsty Smith, Sampsa Hautaniemi and Shan Wu** were instrumental in their help during the completion of the first two manuscripts. I will always cherish the previlige of working closely with these fantastic people.

My thesis committee members, **Drs. Satdarshan Monga, Cary Wu** and **Robert Getzenberg** were some of the most helpful people I have ever worked with. Their guidance helped steer my thesis work in the right direction. I am very grateful for their time and support.

This work would not have been possible without the support of my laboratory members. I would like to thank **Drs. Akihiro Iwabu, Latha Satish and Asmaa Mamoune** for guiding me during my first year in the laboratory. I would also like to thank our lab manager **Diana Whaley** for being the backbone of our lab since I have been here. I would also like to acknowledge the

support of all the Wells' lab members including **Clayton Yates, Cecelia Yates, Anand Iyer, Richard Bodnar, Weixin Zhou, Kien Tran, Jeff Chou, Christopher Shepard, Priya Kulasekaran, Kenichi Tamama, Ranjith Babu and Hanshuang Shao**. These friends of mine made the lab an attractive and fun place to work.

My parents **Ms. Nutan Kharait** and **Mr. Prakash Kharait** have played a profound role in my life. Their unconditional love and constant support during my 'long' educational journey is the key to my success, if any. I feel lucky to have such wonderful parents.

Last but not the least, I would like to thank my wife **Ishita Shah** who has made my life so much comfortable and fun. This work was impossible without her love and support.

Organization of the thesis

The thesis is written as follows: The first chapter deals with the introduction including the progress within this scientific field and the problems it posits. The second chapter describes a novel modeling approach towards understanding cell motility. This chapter predominantly describes the construction of a new methodology, Decision trees, and its application to cell migration. In the third chapter, we describe the utilization of the model to predict cellular biophysical events based on the activation status of key biochemical signals. We extract important predictions from the model and test them experimentally to validate such predictions. In the fourth chapter we apply such predictions to prostate cancer cells and show that abrogating cellular contractility, as predicted by the model, can limit prostate cancer invasiveness. We include some final perspectives in chapter 5.

1.0 INTRODUCTION

1.1 OVERVIEW OF CELL MIGRATION

The homeostasis of the human body depends on a variety of complex processes, cell migration being one of them. In a variety of physiological states, migration of specific cells is key to a biological response; e.g. during inflammation, cytokines and chemokines enable chemotaxis of white blood cells in tissues to help ward off any infection by micro-organisms. During wound healing, fibroblast and keratinocyte migration into the denuded area is necessary for an appropriate wound healing response [1, 2]. In pathology too, migration of cells determines the fate of the organism. Indeed, one of the key phenomenon responsible for mortality and morbidity in cancer is its ability to disseminate into surrounding tissue forming metastatic foci. Motility of tumor cells is a pre-requisite factor in metastasis [3]. While cell migration has been best studied in the fields mentioned above, its role is obvious in numerous other disease states including Rheumatoid arthritis and systemic lupus erythematosus [4]. In both these ailments, inflammation of synovial membranes (and other vital organs) is responsible for much of the morbidity. Many medications like colchicine and indomethacin, used commonly in treating acute arthritis mediate their effects by retarding migration of inflammatory cells into joint space [5].

Targeting cell motility has been a long-standing challenge. Numerous signaling proteins governing redundant biophysical responses that are crucial for motility have been discovered in the last decade [6, 7]. While each of these proteins is necessary for cell movement, it is insufficient alone in mediating migration. This begs some fundamental questions that still need to be answered in cell biology -- where does a certain protein fit in dictating certain biophysical responses? How crucial is it to target such a protein if cell motility needs to be manipulated? Which proteins need to be inhibited together for the most optimal response?[8]. Answering some of these questions needs a novel methodology that can integrate the enormous data generated

during the last decades into fundamental testable hypotheses [9, 10]. While traditional experimental laboratory methods still remain the most widely employed tools for such data measurements, advanced computational methods enable expansion of such data sets by simulation techniques that can further be useful in deriving any significant biological conclusions. Such observations have burgeoned the concept of a "systems biology" approach to studying major biological phenomena. We have employed one such technique, Decision tree analysis, to answer some of the intriguing questions related to cell migration [11]. The following sections describe in detail, some fundamental concepts in cell migration.

1.2 BIOPHYSICS AND BIOCHEMISTRY OF CELL MIGRATION

Cell migration is a net result of a series of coordinated biophysical events [6, 12, 13]. While each of these events described individually, is necessary for migration, they are not independent of each other. In other words, the optimal migratory response is a result of both the spatial as well as temporal coordination as well as repetitive cycling of such events.

Various simplistic models for cell migration have now been elucidated [6, 12, 13]. In order for a cell to move, it first has to be polarized or attain a sense of directionality. This leaves the cell with a tail or a rear end and a front end upon stimulation with a motogenic agent like epidermal growth factor (EGF). Attainment of polarity is accompanied by extension of dominant lamellipodia at the 'front end'. This is followed by detachment of the cell's rear end and subsequent transcellular contractility that provides the necessary force for locomotion (Figure 1). Each of these processes are controlled by numerous intracellular signaling molecules, some of them being involved in other cellular responses as well. Substantial evidence suggests that disruption of these individual biophysical events like lamellipodal protrusion or rear cell-membrane detachment by targeting the underlying signaling switches offers a very promising approach to novel drug discovery [14-16].

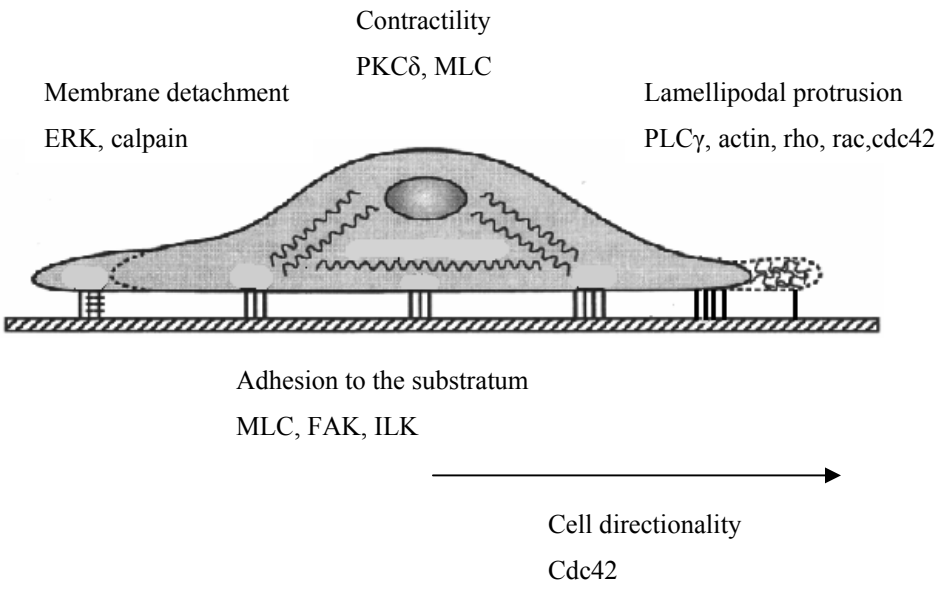


Figure 1. Biophysical processes and underlying biochemistry of cell motility.
Different signaling molecules that govern one or more of the biophysical processes described below are shown. *Adapted with permission from Cell motility in cancer invasion and metastasis; 2005, Edited by Alan Wells; Elsevier publications.*

1.2.1 Lamellipodal protrusion

Upon stimulation with a motogenic agent like EGF, the first step towards cell locomotion involves extensions of the cell membrane, as lamellipodia, in the direction of movement. Lamellipodia are principally composed of cytoskeletal elements like actin and myosin along with integrin receptors and signaling molecules [17]. The Phospholipase C-γ (PLCγ) signaling pathway is shown to be crucial in lamellipodial extension in fibroblasts and cancer cells upon activation by growth factors, though it likely plays little, if any, role in integrin-mediated extension [12]. Phosphoinositide 3-OH kinase (PI3K) is also active at the front of the cell during adhesion receptor-induced as well as growth factor-induced motility but in epithelial and not fibroblastoid cells [18]. Both of these enzymes alter the phosphoinositide face of the inner membrane, with PI3K creating phosphoinositol 3,4,5-trisphosphate (PIP3) and PLCγ hydrolyzing and removing the PIP2. This alters the docking sites for a variety of molecules that

impinge on the actin cytoskeleton. New sites are created by PIP3 and the loss of PIP2, particularly by PLCγ hydrolysis which releases prebound molecules including vinculin, gelsolin, cofilin and profilin [19-21]. All of these molecules then act to disassemble focal adhesions and accelerate actin polymerization, enabling protrusion of lamellipodia [20].

Emergence of a dominant lamellipod marks the state of polarity of the cell [13]. The small GTPase cdc42 is required to either establish or maintain a persistent cell polarity and directionality that leads to productive locomotion [22-24]. This is tightly balanced in that either under- or over-activation of cdc42 tips the balance and either no lamellipodia are formed or none is established as dominant and the cell 'dances' in place [25]. These membrane extensions or lamellipodia are then stabilized by attachment of the protruding lamellipodia to the substratum, thereby enabling new focal contacts between cell and extracellular matrix. Thus, cell-substratum adhesion is a crucial element for migration of a variety of cells and is achieved principally by the action of integrins and other adhesion receptors [13].

1.2.2 Cell – substratum adhesion

Motility of cells require adhesion of its membrane with the underlying substratum. Sites of cell-substratum adhesion are not merely sites of passive contact between the cell and the extracellular matrix but are active in cell signaling. Focal adhesions are rich not only in cytoskeletal proteins, integrins, other adhesion receptors and linker proteins, but also various signaling kinases [26, 27]. The non-receptor tyrosine kinases focal adhesion kinase (FAK), src, and integrin-linked kinase (ILK), all modulate cell-adhesion dependent growth, survival and motility by activating numerous signaling pathways [28]. These have been shown to interact directly with the adhesion complex and to be activated therein. Additionally, the adhesiveness of substratum also dictates the biophysical response [29]. While too much surface adhesion strength prevents detachment of the cell, too little adhesiveness provides minimal force for contraction, retarding motility at both these extremes [30, 31]. Active cell movement depends on formation of new focal adhesions along with breakage of others. Thus, the net movement is a product of cell-substratum adhesion strength that is controlled in turn by surface ligand concentration, receptor number and ligand-receptor affinity [30]. During high surface adhesiveness, calpain cleaves the integrin cytoplasmic

edges enabling rear detachment. However, during lower surface adhesiveness, breakage of focal adhesions via contraction is adequate for motility [32, 33]. Although a variety of different signaling proteins are implicated in this event, the principal effect is for the cell to "walk-on" a surface, with the surface providing the necessary scaffold as well as signaling from the extracellular mileu.

1.2.3 Transcellular contractility

Upon sustaining contacts with the substratum, the cell contracts and generates sufficient force needed for translocation. Adhesion sites are vital as a fulcrum for mechanical forces. Transcellular contractility is achieved via a consorted action of actin-myosin cytoskeletal machinery. In fibroblasts and prostate cancer cells at least, growth factors can activate the regulatory element within myosin light chain (MLC) proteins of myosin II via Protein kinase Cδ (PKCδ) [34]. Such contractility is needed by tumor cells for motility and by fibroblasts during contraction of the wound edges during later stages of wound healing [35, 36]. Studies from our laboratory have shown that invasion of DU145 and PC3 prostate cancer cells through a thin layer of Matrigel is substantially reduced by pharmacological and molecular inhibition of the PKCδ - MLC signaling pathway (Kharait et al, Biochem & Biophy Res Commun., 2006, In press). The reduction in invasiveness was predominantly a result of reduced cellular motility. Transcellular tension also leads to membrane detachment by fracturing focal adhesions during lower substratum adhesiveness [32, 33]. Overall, the phosphorylation/activation state of MLC is controlled by a balance between the activating MLC Kinase (MLCK) and deactivating MLC Phosphatase (MLCP) enzymes with such inputs utilized by a majority of cells. A second pathway, upstream of MLCK, occurs via Rho Kinase which directly activates MLCK and inhibits MLCP, thereby stimulating cell contractility and motility; this pathway is likely used by adhesion receptor signaling [37, 38]. Additionally, Rho Kinase can directly activate MLC by phosphorylating the 19^{th} serine residue of regulatory light chain [39]. However, both MLCK and Rho Kinase can distinctly and exclusively regulate MLC phosphorylation at differential subcellular locations, with MLCK regulating MLC activation at the cell periphery whereas ROCK controlling it more at the center [40]. Sequential activation and subcellular localization of these two distinct MLC pools occurs during different cellular biophysical events such as cell

motility, spreading and adhesion. Abrogation of Rho Kinase in PC3 prostate carcinoma cells substantially inhibited their migration and invasion *in vitro* and *in vivo* [41]. To add to the list, other enzymes have recently shown to be involved in regulating MLC activation. These include zipper-interacting protein (ZIP) kinase and Integrin linked kinase (ILK) [42, 43]. While both of these are responsible for MLC activation during integrin mediated haptokinetic motility, PKCδ mediated contractility occurs downstream of growth factor (mainly EGF) receptor signaling.

1.2.4 Detachment of the cell membrane at the rear end

Adhesion of the lamellipodia to the substratum at the front end is accompanied by the detachment of the cell membrane at the rear that enables the cell to move forwards [12]. Rear detachment is both passive in response to transcellular tension driven by MLC [32, 33], and active in that adhesion sites in the rear are weakened through active signaling. Transcellular tension is sufficient only under regimes of low adhesivity [33], under which overall locomotion of fibroblasts is actually decreased. Active deadhesion is achieved largely by the intracellular protease calpain, two isoforms of which are found within motile fibroblasts and prostate cancer cells [44]; Calpain-I (μ-calpain) and calpain-II (M-calpain) activated *in vitro* by micromolar and millimolar levels calcium respectively. Calpain II has been shown to be activated by direct phosphorylation by ERK/MAP kinase downstream of signaling from EGFR and other receptor tyrosine kinases [45]. In the absence of receptor tyrosine kinase activity, it is possible that calpain I subsumes this role being activated by calcium influx upon triggering stretch-activated calcium channels [46]. Calpain cleaves the cytoplasmic tails of integrins and/or linker proteins such as talin, loosening the attachment to substratum [44]. Coupled with transcellular contractility this leads to membrane de-adhesion at the rear end of the cell. Rear membrane detachment primarily occurs via either a rapid mechanism seen in conditions of lower surface adhesiveness and involves the dissociation of integrin receptors with the extracellular matrix or a slow detachment mechanism seen during conditions of higher surface adhesiveness involving breakage of integrin receptor-cytoskeletal linkages [47]. During such dissociations, a large number of integrins is ripped off from the cell membrane and are found left at the site of broken adhesions [47]. The importance of this biochemical event in cancer progression was shown by

abrogation of calpain using molecular and pharmacological agents substantially inhibiting migration and invasion of prostate cancer cells [14]. In addition, calpains may be involved in lessening the cell-cell adhesions required for tumor cell dissemination from the primary localized mass by targeting E-cadherins [48].

Careful orchestration of these individual biophysical events enables productive cell locomotion. By conceptually segregating these discrete events, we have begun to understand how intracellular switches act in concert to produce the optimal response. Many of these biophysical events, as mentioned earlier, are now potential targets in anti-cancer drug research.

1.3 CELL MOTILITY IN TUMOR INVASION

The ability of tumor cells to burrow through the stroma enables formation of metastatic foci. The process of invasion is a highly coordinated one [3]. As compared to normal cells, tumor cells have an increased capability to loosen their connections to the substratum and break cell-cell adhesions. Integrin receptors that mediate the cell-substratum adhesion and cadherins that mediate cell-cell cohesion, are pivotal in this behavior. Integrins serve a double role, as they interact as adhesion receptors with the substratum during tumor cell migration. Migration through the extracellular matrix (ECM) barriers, while mainly occuring via natural cleavage planes, also requires matrix remodeling accomplished by various proteases including matrix metalloproteinases [49-51]. This motile strategy is used both to invade local adnexia and gain access to conduits for distant dissemination.

Tumor cells disseminate mainly via bloodstream or lymphatics (Figure 2). During this process, the tumor cells must survive the de-adhesion induced anoikes either by cell intrinsic changes rendering them resistant or by forming clumps to recreate the ‘attached’ cell signals. These individual cells or clusters of cells reach target organs, extravasate, and migrate to appropriate sites within the tissues. If the target tissue provides the missing extrinsic signals, the cells will proliferate to form a metastatic focus. Both local invasion and metastasis are significant issues in the early dissemination of prostate carcinoma, as opposed to breast cancer in which

invasiveness is a later event or ovarian carcinoma in which peritoneal spread predates demonstrable metastases. Thus, maximum benefit may accrue from deciphering and targeting a tumor cell acquired property that is critical for both, such as motility.

Recent research in tumor biology has bolstered the concept that tumor invasion is the net result of dysregulated cell motility [3]. Invasiveness of tumors can directly be attributed to their migratory potential since epithelial (carcinomatous) cells that have transitioned to a mesenchymal phenotype are highly motile and possess higher capability to penetrate through the matrix barrier. Such a transition is achieved via genetic as well as epigenetic perturbations within the transformed cell. Overexpression of certain growth factor receptors (including the EGF receptor) is pivotal in oncogenesis as well as tumor progression.

The ErbB family of growth factor receptors has been implicated in the progression of variety of tumors including that of the breast, brain, lung, urinary bladder and prostate [52-57]. These tumors secrete a variety of ligands for growth factor receptors in an autocrine manner establishing intrinsically self-stimulating autocrine loops [57-59]. Such aberrant and continual signaling drives tumor cell proliferation, survival, migration and invasion. Thus, tumors acquire invasive properties, atleast in part by boosting autocrine systems that keep the intracellular signaling cascades in an active state. Indeed EGFR overexpressing prostate cancer cells (DU145) were shown to exhibit higher invasiveness as compared to parental cells [60]. Targeting EGFR induced motility reduced their invasiveness.

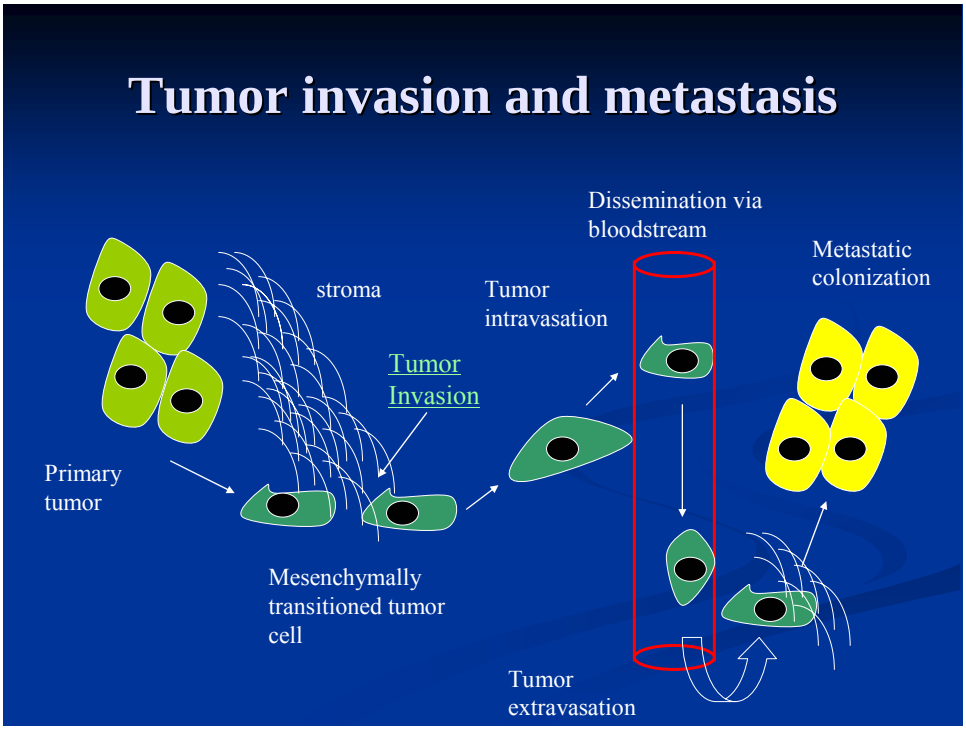


Figure 2. Overview of the metastatic process.
The various steps during metastatic progression are shown. Motility of tumor cells is vital during invasion through the stroma, extravasation into the blood vessels and final 'homing' of tumor cells into a distant focus.

While conventional chemotherapy targets population of cells undergoing proliferation, it has substantial adverse effects like gastrointestinal intolerance and alopecia often leading to discontinuation of treatment due patient compliance issues. Motility-specific chemotherapy can be proposed to circumvent some of these adverse effects if developed and utilized clinically. In a variety of laboratory based experiments, targeting motility of tumor cells left proliferation unaffected [14, 15]. Such an approach can render a tumor less invasive but more amenable to surgical removal without exhibiting deleterious adverse effects. Thus, a thorough mapping of intracellular 'motility-specific' signaling switches that can be 'turned off' is the key to future anti-cancer drug development that can be better tolerated by patients.

1.3.1 Cell motility in Prostate carcinoma invasion

Cancer is the second leading cause of death in the United States, second only to cardiovascular disease. Prostate cancer is the most commonly diagnosed cancer in men, with more than 70 % of prostate cancers being diagnosed in men over 65. Prostate cancer is rare in young men and the probability rises steeply with increasing age with the risk being 1 in 7 after the age of 60 as compared to a risk of 1 in 44 between the age 40 and 59 years. Organ-confined prostate cancer advances slowly with survival lasting over a decade even in untreated cases. This creates a therapeutic dilemma in older patients, as surgical and radiological ablation of the tumor (and the prostate) carry significant morbidity and even subsequent mortality [61]. The current therapeutic management of advanced prostate cancer involves chemical or physical castration to induce an androgen-withdrawal apoptosis of the tumor cells. However, this prolongs survival by only about 10% on average with the relapse being androgen-independent. The progression to invasion and metastasis is thus only slowed and not blocked.

Local extracapsular invasion of prostate cancer continues to be a significant problem especially in elderly patients, with limited options of radiotherapy and combination chemotherapy [62, 63]. Invasion of the adnexia results in compromised function of the renal and genital systems with significant physiological and psychological sequelae. Additionally, metastatic spread carries a high mortality burden. A recent study on autopsy of patients that died of hormone refractory prostate cancer analyzed the proteomic and genomic signatures as well as the metastatic sequelae that led to mortal complications [64]. In this cluster of 30 patients autopsied, bone was the most common site of metastasis (approximately 83 %) with majority of the lesions affecting the vertebral column. Once in bone, the prostate carcinoma cells induced an osteoblastic response that was responsible for debilitating bone pain in majority of these patients. Other frequent sites of metastasis included the liver (66 %), lymph nodes (63 %) and lungs (50 %). Brain metastasis, initially thought of as a rare (1-2 %) complication, was identified in 23 % of these cases and probably carried the worst prognosis with a maximum life expectancy of 6 – 7 months.

Prostate cancer mortality results mainly from progression to the invasion and metastatic state. Significant morbidity ensues from surgical or radiological ablation of non-invasive, organ-confined carcinomas, and these adverse effects are greatly amplified once the tumor breaches the

prostate capsule or migrates out along neurovascular conduits [61]. The approach to organ-confined carcinomas is still uncertain as development of clinically insignificant prostate adenocarcinoma appears to be a consequence of aging with small carcinomas found upon autopsy examination of most men dying in their 90s. Still, the current practice is to remove carcinomas as advanced prostate cancer responds poorly to the current generation of therapies. This is due in large part to these agents being optimized to kill growing cells while prostate cancer present mitogenic indices far lower than many normal tissues. As such, future interventions need to focus on the cell properties that enable prostate carcinomas to migrate from an organ-confined situation to invade adnexia and metastasize to distant organs. Recent studies have highlighted a central role for cell motility in acquiring the ability of prostate tumor cells to reach a metastatic focus [65].

Increase in migratory potential of prostate cancer cells is impelled by vigorous upregulation of certain growth factor receptors including the EGF receptor and the HGF receptor, c-Met [66-68]. In prostate cancer, signaling through EGFR is increased mainly via increased secretion of its ligand TGF-α by transformed epithelial (autocrine) as well as stromal (paracrine) cells [68]. Thus, the net result is sustained activation of intracellular signaling cascades, downstream of EGFR signaling axis, that promote tumor progressive properties. Targeting such signaling switches in laboratory experimental settings has been effective in limiting prostate cancer cell motility and invasion [14, 15, 69]. In addition to epigenetic alterations, upregulated growth factor receptor signaling causes increased expression of certain pro-metastatic genes like the urokinase plasminogen activator (uPA) and its receptor (uPAR) [69] and matrix metalloproteinases [70, 71]. These signaling systems are one of the many that enable degradation of the stroma and tumor dissemination through the matrix. Thus, limiting tumor cell motility by targeting growth factor receptor signaling can prevent cancer dissemination and its progression.

1.4 THE EGF RECEPTOR (EGFR) SYSTEM IN CELL MOTILITY

1.4.1 Introduction

The type 1 growth factor receptor tyrosine kinase (RTK) family, also known as erbB or Human Epidermal growth factor receptor family comprises of four well characterized receptors [72]. These include c-erbB1 (or HER-1) or EGF receptor (EGFR), c-erbB2 (or HER-2/neu), c-erbB3 (HER-3) and c-erbB4 (or HER-4). Several ligands have been identified for these receptors except for HER-2, for which a ligand is yet to be found. However, much of the signaling from this family of receptors involves cross activation and hetero-aggregation of members. Specific ligands to EGFR (HER-1) include epidermal growth factor, transforming growth factor-α, heparin-binding EGF and amphiregulin; the high affinity ligands for HER-3 and HER-4 are heregulins, with the first member also known as neu differentiation factor [73]. Upon binding of the ligand to the extracellular domain, these RTKs are activated by homo- or hetero-dimerization thereby leading to tyrosine phosphorylation at multiple residues within the long cytoplasmic tails [74]. These phosphotyrosine residues then serve as multiple docking sites for various SH2- or PTB-containing adaptor proteins. Direct activation of these docking proteins often by phosphorylation, or indirect activation of their downstream effectors elicits the multiple cell responses.

EGFR was the first receptor identified as a proto-oncogene and, thus, the first to be implicated by upregulation in numerous human cancers. Interestingly, the level of EGFR activation correlated with tumor progression and not carcinogenesis [3]. Various genetic alterations such as gene amplification or alternatively spliced variants have been found in glioblastomas [75] and other carcinomas with this correlating with worse clinical prognosis. In prostate cancer the relationship appears primarily epigenetic with a vigorous upregulation of ligand production and autocrine stimulation [57]. ErbB2 was initially invoked as a proto-oncogene when a chemically-induced tumor model system was found due to an activating mutation in the transmembrane domain of this receptor [76]. While this mutation has yet to be defined in human carcinomas, increased levels of erbB2 have been noted in many human carcinomas, breast cancer in particular [56]. The precise role of erbB3 and erbB4 in prostate

cancer is still very unclear. ErbB3 at least must function via transactivation and aggregation with other members of the family, as it lacks intrinsic kinase activity.

The primary role of these receptors in cancer progression in general is highlighted by the fact that two new therapies specifically target EGFR (Iressa by AstraZeneca and Tarceva by OSI Pharmaceuticals) and erbB2 (Herceptin by Genentech). These are the first biologicals to target signaling receptors and the first, along with Gleevec (anti-abl), to inhibit tyrosine kinase function. While these therapies have been approved for other carcinomas, they are being explored in prostate carcinomas due to the upregulation of signaling through these receptors in this cancer [77, 78].

1.4.2 Signaling from the EGF receptor

EGFR is a prototypical receptor tyrosine kinase and activates a myriad of signaling cascades upon ligand occupancy [74] (Figure 4). Not unexpectedly, many activated signaling proteins participate in multiple cellular responses; e.g. ERK(1/2) MAPKinases when active are involved in cell proliferation as well as motility. In this vein, the precise locale of the active signaling protein dictates its participation in a specific cellular response. ERK, translocates to the nucleus where it modifies transcription of certain genes needed in cell growth, whereas membrane translocated ERK (seen upon EGFR activation) can mediate detachment of the cell membrane via calpain activation [79]. Similarly, PKCδ is required for EGFR induced transcellular contractility [34] but it also known as a pro-apoptotic signal in diverse cell types when activated [80]. As seen with ERK, nuclear versus cytoplasmic PKCδ exhibit completely diverse functions. Thus, the ultimate fate of a cell depends not just on the activation of signaling receptors but also on the locale of its final effector molecules. It appears that the cell chooses among these sometimes mutually exclusive responses depending on the current cell proteome, other signaling pathway extant, and the temporospatial aspects of EGFR and secondary effector signaling.

Ligand occupancy drives activated EGFR into a conformational change which leads to receptor dimerization and auto (cross) phosphorylation of multiple tyrosine residues within the cytoplasmic domains. Binding of multiple adapting molecules, including Grb2, Shc and Sos, to these phosphotyrosine residues activates the well characterized Ras - Raf - Mek signaling

pathway that finally activates ERK/MAP Kinase pathway required for cell proliferation and gene expression [81, 82]. As mentioned above, transient ERK activation is employed at the cell membrane for cell detachment [79]. Another signaling pathway activated by EGFR involves phospholipase C-γ (PLCγ) activation that enables motility via reorganizing cellular cytoskeletal architecture [21]. The PLCγ signaling pathway is motility specific and is not required for proliferation [83]. Upon activation, PLCγ hydrolyses membrane bound PIP2 leading to liberation of actin modifying proteins that impinge on actin and promote its polymerization [84, 85]. EGFR also activates the signal transducer and activator of transcription (STAT) group of transcription factors, small GTPases including rac, rho, and cdc42, Phosphatidylinositol 3-kinase (PI3K), certain Protein kinase C isoforms (PKCδ), and to a lesser extent phospholipase D and tyrosine kinase Src [74]. While the STAT group of transcription factors are predominantly linked to growth factor mediated cell proliferation and survival, unpublished reports from our laboratory have highlighted their role in EGF induced migration as well.

The phosphatidylinositol 3-Kinase (PI3K) signaling pathway is likely as contributory as Ras - ERK pathway to cell proliferation and survival [86] and occurs downstream of both growth factor receptor signaling as well as integrin activation. PI3K's most evident action is to phosphorylate the 3' position of the inositol ring, with preference for the membrane phospholipid PI(4,5)P2 to produce phosphatidylinositol 3,4,5- triphosphate (PIP3). The counterregulatory tumor suppressor PTEN antagonizes PI3K actions by removing this modification [87]. The production of PIP3 leads to activation of Akt/Protein Kinase B via intermediary kinase(s) PDK1 (and possibly PDK2). Akt signaling strongly promotes cell survival and proliferation [88].

PI3K activity has also been linked to increased migration, matrix metalloproteinase production and stromal invasion of certain tumors [89, 90], with some of this occuring through activation of Akt1 [91, 92]. Overexpression of Akt2 / Protein kinase Bβ, increases migration and invasion of breast and ovarian cancer cells via upregulation of β1 integrins [93]. Integrins also activate PI3K, with the PI3Kinase pathway being required for αvβ3 mediated migration of highly invasive PC3 cells [94].

1.4.3 Ligand-receptor interactions in EGFR signaling

Upon binding the known, high affinity ligands, EGF receptors undergo internalization and subsequent degradation within the endosomes [74, 95]. This serves primarily as an attenuation mechanism [95]. However, internalized EGF-EGFR complexes still activate certain signaling pathways that are crucial in cell proliferation and gene expression [96, 97]. Once internalized, EGFR occupancy determines receptor and ligand fate (Figure 3). EGF remains tightly bound to the receptor driving both ligand and receptor to degradation. TGF-α, which does not signal from the endosomes since the endosomal acidic pH leads to dissociation of TGF-EGFR complex, results in EGFR recycling, while ligand sorts with the fluid phase in that 2/3 is shunted to the lysosome. Interestingly, when ligand is replenished in an autocrine fashion, TGF-α produces more prolonged EGFR signaling as the receptor is 'spared' compared to EGF autocrine production [98]. Thus, the specific ligand present during autocrine stimulation of prostate carcinomas, predominantly being TGF-α, dictates EGFR signaling in a temporal and spatial manner.

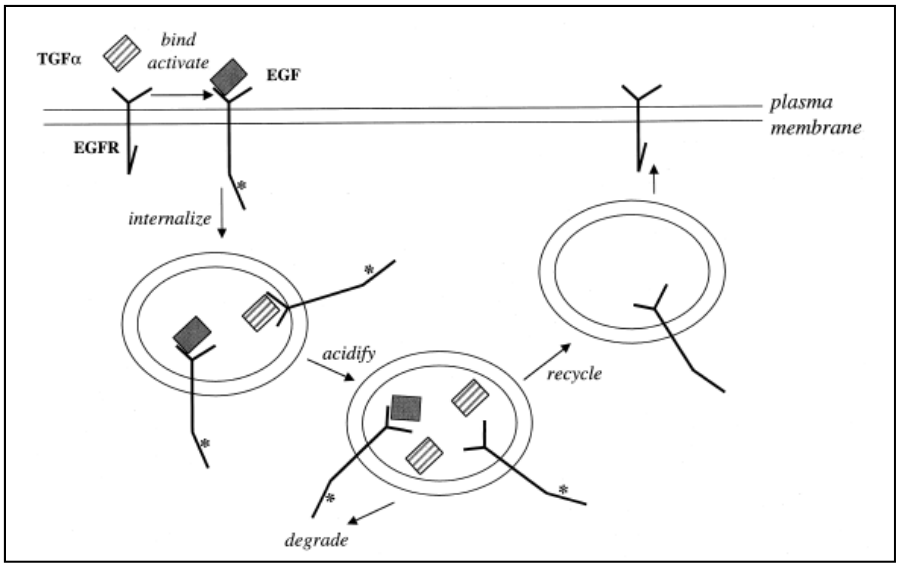


Figure 3. Ligand induced trafficking and disposition of EGF receptor.
EGF-EGFR complexes are tightly bound and undergo degradation whereas TGF-α dissociates in the acidic environment within the lysosomes and the EGFR is recycled to the membrane. *Adapted with permission from Wells A. EGF receptor. 1999, Int Jour. of Biochem and Cell Bio; 31, 637-643.*

A second mode of EGFR activation that also limits EGFR degradative attenuation involves transient binding by ultra-low affinity ligands [99, 100]. Some of the EGF-like repeats in the ECM components tenascin-C, laminin, and decorin bind to EGFR but in a manner that is sufficiently transient so as not to drive internalization [99-101]. This restricts signaling from EGFR to the plasma membrane, and under this restriction motility appears preferential to proliferation at limiting levels of ligand (unpublished observations). These low affinity EGFR ligands can be liberated or unmasked from the extracellular matrix via the action of matrix metalloproteinases during conditions of organogenesis, matrix remodeling or tumor invasion [100]. The fact that many of the ECM proteins, tenascin-C in particular, are upregulated (or dysregulated) in a variety of pathophysiological states strengthens the observation that cell-ECM interactions or "sensing of the stroma" by the cell is a vital element during cell migration [102].

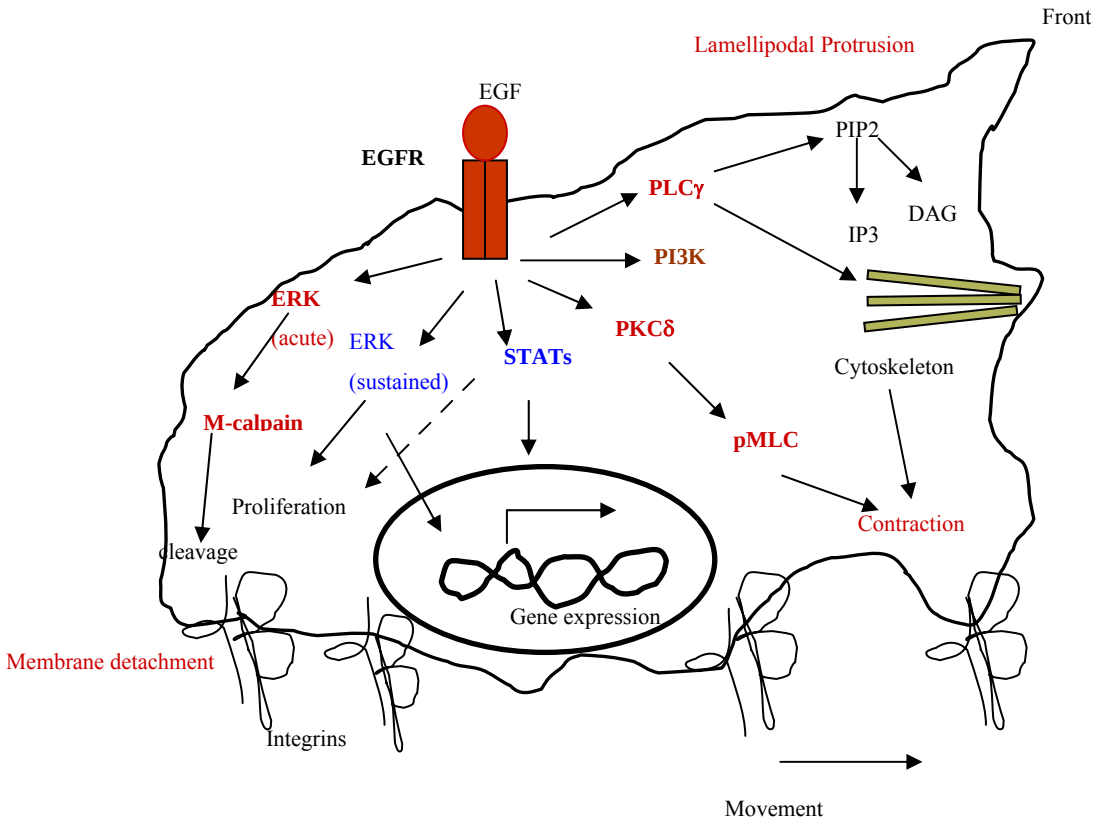


Figure 4. EGF receptor signaling.
Signaling through EGFR activates a myriad of signaling cascades that are involved in diverse cellular processes including cell migration. Some of the crucial effectors and the key biophysical processes they control are shown. In red are signaling proteins that are relatively 'motility specific' whereas blue represent proteins involved in proliferation as well as motility (e.g. STATs, ERK). *Adapted with permission from Cell motility in cancer invasion and metastasis; 2005, Edited by Alan Wells; Elsevier publications.*

1.5 ADHESION RECEPTORS IN CELL MOTILITY

Migration of cells is mediated by integrin adhesion receptors, which comprise a family of homologous, transmembrane heterodimers built specifically from a combination of one of the twelve alpha and one of the nine beta subunits [103]. These receptors recognize the cues, both soluble and matrix embedded, from by the cell's immediate surrounding environment, the extracellular matrix [104]. Integrins physically connect the extracellular matrix proteins like fibronectin and laminin to cellular cytoskeleton and provide tractional forces both to and from the substratum during migration [105]. In addition, signaling through these receptors is pivotal for motility over a range of adhesive conditions [29, 30]. Integrin engagement with extracellular matrix ligands is required for RTK mediated activation of downstream signaling cascades [27]. Clustering of integrins at points of substratum attachment recruits several cytoskeletal proteins including actin, talin, vinculin, paxillin, etc. as well as several non-receptor tyrosine kinases like FAK and src to the site of focal adhesions [106-110]. Evidently, focal adhesions are not merely sites of attachment but also signal actively and are the principal mediators of signal transduction events from the external environment to the intracellular mileu [111, 112]. Prevention of cell adhesion to the substratum via blocking specific integrin receptors has evidently shown to abrogate cell migration [113]. In addition, integrin receptor signaling is crucial in mediating cell survival in a variety of conditions, with apoptosis observed immediately with loss of cell adhesion in a variety of untransformed cells [103, 104, 114, 115]. Malignant cells overcome this problem by altering the integrin receptor subtypes that provide the necessary 'adhesion' signals despite cell detachment during tumor extravasation and dissemination [116, 117]. This has been observed with α6β4 integrin subtype that binds laminin and transmits survival signals via activation of Rac [117].

Integrin receptors provide the interface with the surfaces. These adhesion receptors actively signal during this process, enabling or preventing the receptor tyrosine kinase (RTK)-mediated responses, or even, themselves, driving the motility [118]. Extensive evidence suggests that unregulated growth and migration of tumor cells is due in part to an alterations of integrin expression [114] accompanied with a loss of cell-cell adhesion molecules [119-121], particularly E-cadherin with this finding ubiquitious to almost all carcinomas[122-125]. Cadherins mediate

cell-cell cohesion in normal cells but is downregulated in majority of cancers leading to a loss of cell-cell contacts [126]. In prostate cancer cells, EGFR is preferentially localized only along the tight baso-lateral surfaces. A loss of cell-cell cohesion exposes the EGF receptors on the baso-lateral surfaces and enables significant autocrine signaling through ligand (TGF-α) – receptor binding [127]. Autocrine signaling further downregulates E-cadherin expression with this cycle finally continuing to tumor invasion.

Epigenetic alterations within the cellular proteome enable tumor invasion through the stroma by enabling the tumor cells to perceive the cues from the extracellular matrix. In addition, altered integrin expression remodels the ECM to a tumor cell's own advantage by selective signaling via 'beneficial' integrin receptors [103]. Tumor cells specifically acquire invasive properties by selective elimination of certain integrin receptor subtypes/subunits like β4 [128, 129] while retaining others that favor tumor stromal invasion, particularly the α6β1 class[115, 128]. In addition, changes in integrin receptor expression alter the intracellular signaling through various signal transduction pathways including the well characterized PI3K and ERK pathways [114, 130] enabling tumor cell survival in a variety of 'abnormal' conditions seen during metastatic colonization. It must be noted that while the intracellular signaling cascades activated might share the same molecular members as those derived from RTK, the temporospatial aspects of signaling are quantitatively and likely qualitatively different, and thus drive distinct cell behaviors. Thus, it is sufficient to say, that alterations within integrin receptor expression, like selective deletion of some subtypes with upregulation of others, enable tumor cell motility.

1.6 MATHEMATICAL MODELING APPROACHES TO CELL MOTILITY

1.6.1 Motivation

Cellular responses are guided by different environmental cues that vary during different pathophysiological states. A cell can make intelligent decisions by responding to specific

environmental signals since it is equipped with a highly effective, albeit complex, signaling network comprising thousands of signaling proteins [131]. Such a connected proteomic network enables the cell to choose among certain mutually exclusive tasks; e.g. a cell can choose proliferation over motility or apoptosis given the right environmental cues and extant of other signaling pathways. Most human diseases are a result of dysregulation of one or more signaling pathways and hence identifying such 'altered' signaling nodes can not only furnish our understanding about disease pathogenesis but also guide future therapeutic goals [8, 132]. Cell motility is one such complex biophysical response that is required for biological homeostasis and is dictated by a variety of extracellular signals including growth factors (EGF,VEGF, PDGF) and their signaling receptors, cytokines (interleukins like IL-6), chemokines (IP-9, IP-10, PF4), the extracellular matrix components (fibronectin, collagen) and other networks [65]. Thus, multiple stimuli converge on redundant signaling pathways through specific receptors which enables the cell to 'process' such information flow and make decisions accordingly. Thus, to target cell motility for therapeutic purposes, it is crucial to understand where individual proteins function as connecting nodes within such a signaling network.

Cell migration is a complex cellular response guided by spatio-temporal activation of a myriad of activated proteins. Our understanding of this cellular response is limited to semi-quantitative patterns at best. Traditional experiments have generated gigantic number of data sets with hundreds of measured variables. Some of these players act upstream (like PLCγ) directly activated by growth factor receptors while others are more downstream effectors like actin and myosin. The complexity of this 'motility specific' proteomic network is not limited to these and includes other members like rho, rac, cdc42, PI3K, etc. Each of these molecules is needed for motility since their inhibition retards migration of target cells [34, 133]. However, it is yet unclear how these proteins interact together in such a network system. This is because rarely have all the signal transduction cascades been studied from a 'systems biology' perspective. Moreover, studying all these proteins together is not always feasible using traditional laboratory based experimental protocols (like western blotting, kinase assays, ELISA) which are time consuming and can only be applied to a few molecules at a time. Thus, to assimilate all the information content, computational / mathematical modeling is needed that enable studying proteomic networks created from such measurements in a holistic manner. Such a sophisticated analysis can interpret the quantitative relationships of 'signals' with 'responses' [134].

1.6.2 Choosing a modeling approach

The choice of a model depends upon multiple experimental parameters some which include the availability of data sets, the amount of variables within the data, the number and reproducibility of measurements, the applicability of the measured data to the *in vivo* milieu and the most important of all, the biological question(s) that need to be answered using the model [134, 135]. Thus, a wide spectrum of computational techniques is available depending upon the answers that are sought [135] (Figure 5). Many mathematical models have been indispensable for explaining complex biological phenomena; e.g. Janes et al, have used partial least-squares modeling to predict cellular responses from measurements of upstream signaling network in cytokine induced apoptosis [136]. This model identifies two groups of intracellular signals that are activated using concomitant cytokine (apoptotic) and growth factor (survival) stimulation: stress-apoptosis signal group and a survival signal group and explains how a cell chooses to undergo apoptosis versus survival depending upon the status of intracellular activated signals. Woolf et al have utilized Bayesian network analysis to link proteomic signaling activities with embryonic stem cell differentiation responses to extracellular cues [137]. This model non-intuitively predicted the association of ERK activation with ES cell differentiation along with Raf phosphorylation with proliferation of differentiated ES cells. Thus, these novel modeling approaches, when applied to biology, have the power to predict complex cell behaviors based on intracellular signaling events.

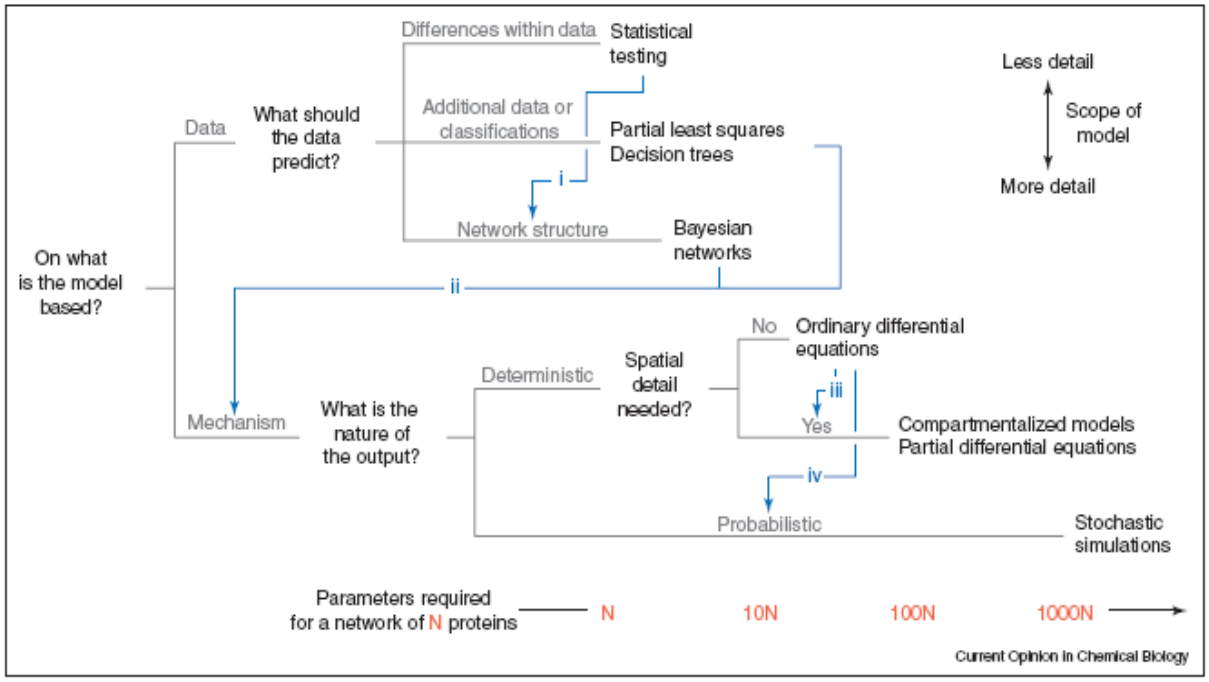


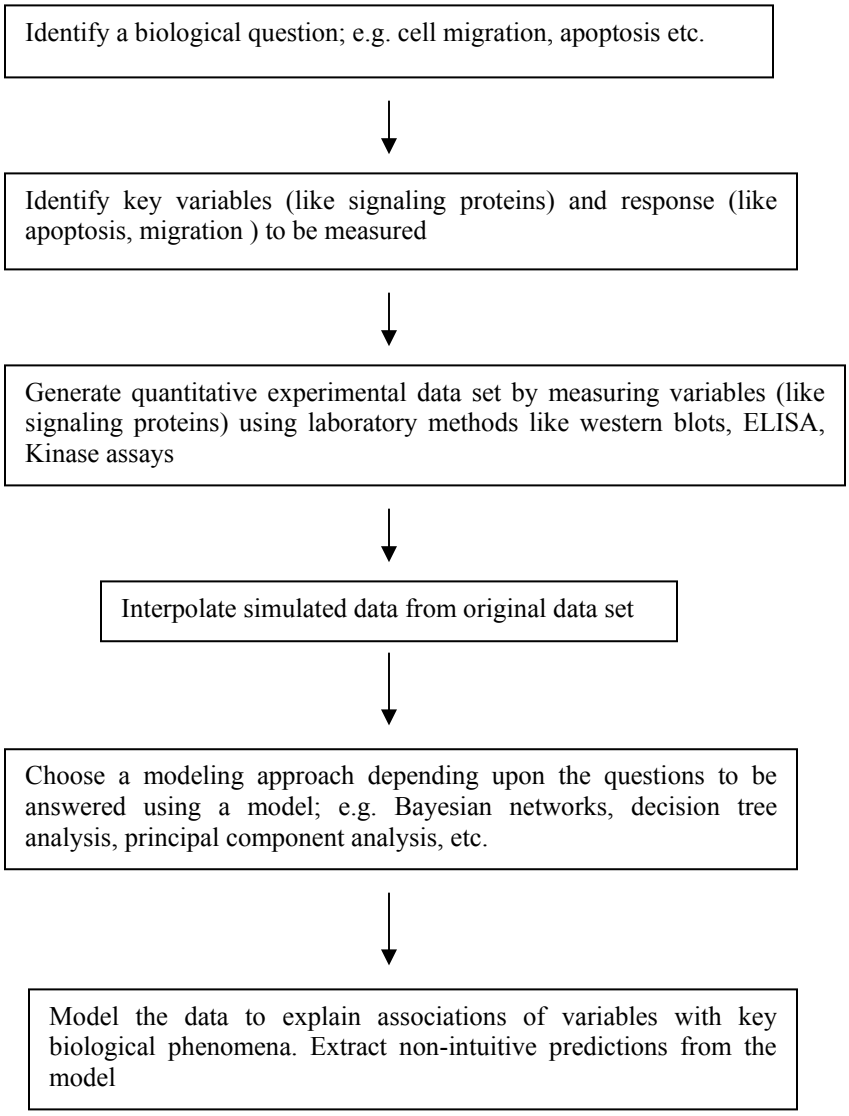
Figure 5. Decision tree for selecting different modeling approaches.
Models with differing levels of detail are organized vertically, and parameter requirements are arranged horizontally on a logarithmic scale. Blue arrows illustrate potential interconnections between models: (i) statistical tests leading to new insights into network structure; (ii) data-driven models that reveal new network mechanisms; and ordinary differential equations that test the need for (iii) spatial detail or (iv) stochasticity. *Reproduced with permission from Janes et al. A biological approach to computational models of proteomic networs. Current opinion in Chemical Biology 2006, 10:73-80.*

Modeling is based upon the data set chosen for a given set of experimental conditions. This begs some important questions that need to be addressed like which variables (such as signaling proteins) need to be measured under a given condition to generate such a data set? Is the data set sufficient to draw counter-intuitive predictions and conclusions from the model? During cell proliferation for example, multiple kinases and transcription factors are actively signaling leading to gene expression. Since measurements of all different molecules is time consuming and cumbersome to undertake, the 'most significant' molecules that have shown to impact proliferation under given set of conditions are selected empirically. One drawback of this

technique is that it relies on 'judgement' and can exclude some critical signaling switches that may have a bigger impact on the measured response. As a result, model construction usually suffers from two hurdles. Models with too little detail tend to be constrained but erroneously oversimplified whereas models that incorporate too many details lack credibility since the additional confounding parameters can reduce the 'predictive power' of the model. In addition, predictions from such detailed models are difficult to test experimentally due to their complex experimental designs. Thus, an ideal model should contain enough parameters and measured variables that capture an important biological trend from experiments and which can be then utilized to predict and test non-intuitive biological conclusions using accepted laboratory methods.

1.6.3 Algorithm for model construction

Despite significant advances in biotechnology, the complexity of human biological situations often renders the experimental data to be limited in its ability to draw significant conclusions. Thus, an interpolation (or expansion) of the measurements is a must to create a biologically relevant data set. Polynomial modeling offers one such avenue for data expansion and has been widely utilized [11]. While some noise cannot be excluded from experimental measurements due to technical errors, consistent and technically reproducible data is the key for constructing a scientifically valid model that simulates a biological phenomenon. The model can also suffer from inaccurate or inadequate predictions due to a lack of proper variables chosen for measurements, which most often are decided from empirical judgement. Many probabilistic models like Bayesian networks assume the data to be uncertain and hence can comply with a variable amount of noise within the data set [137]. Usually, models are simplified so that they extract only the most significant variables that are linked to the biological responses. Thus for model construction, it is pivotal to chose variables like activated proteins that are most relevant to the biological phenomenon; e.g. measuring activation status of ERK, PLCγ , PKCδ, MLC, calpain, etc. for motility. A brief algorithm of model construction is shown below.


Identify a biological question; e.g. cell migration, apoptosis etc.
Identify key variables (like signaling proteins) and response (like apoptosis, migration) to be measured
Generate quantitative experimental data set by measuring variables (like signaling proteins) using laboratory methods like western blots, ELISA, Kinase assays
Interpolate simulated data from original data set
Choose a modeling approach depending upon the questions to be answered using a model; e.g. Bayesian networks, decision tree analysis, principal component analysis, etc.
Model the data to explain associations of variables with key biological phenomena. Extract non-intuitive predictions from the model

1.7 SUMMARY AND INTEGRATION

The genesis of diseases and their progression results from dysregulation of various critical intracellular signaling nodes within diverse human tissues. Such altered signaling leads to dysregulated cell functions, motility, being one of them. Clearly, the complexity of such a biophysical event requires holistic approaches to elucidate contributions by various intracellular signaling switches since these are amenable to pharmacological interventions. Cancer progression for example, results, atleast in part, from epigenetic perturbations within multiple signaling cascades that lead to an aggressive cellular phenotype. One characteristic of such phenotype is its increased motility that is central to invasion. Since most biological cellular responses are accomplished via epigenetic modifications (phosphorylation, translocation, cleavage) of existing signaling proteins without a significant change in their stoichiometry, our understanding can only be expanded by accurate high through - put proteomic measurements that can be incorporated in computational models for simulation. Such models can then non-intuitively identify signaling nodes, which if targeted together, can most significantly influence a biological response.

Our studies, as described below, outline a novel mathematical modeling approach to understanding such complex proteomic networks that govern cell motility. Further, as an application of such a modeling approach, we show that after identifying crucial signaling nodes, targeting them using conventional methods, can limit invasion of prostate cancer cells.

2.0 MODELING OF SIGNAL RESPONSE CASCADES USING DECISION TREE ANALYSIS

Sampsa Hautaniemi[1,2], Sourabh Kharait[3], Akihiro Iwabu[3], Alan Wells[3], Douglas A. Lauffenburger[1]

[1]Biological Engineering Division
Massachusetts Institute of Technology
Cambridge, MA 02139

[2]Institute of Signal Processing
Tampere University of Technology
Tampere, Finland

[3]Department of Pathology
University of Pittsburgh Medical Center
Pittsburgh, PA 15213

Published in *Journal of Bioinformatics*, 2005. Vol 21, no.9, 2027-2035. Used with permission, format adapted for dissertation.

2.1 ABSTRACT

Motivation: Signal transduction cascades governing cell functional responses to stimulatory cues play crucial roles in cell regulatory systems and represent promising therapeutic targets for complex human diseases. However, mathematical analysis of how cell responses are governed by signaling activities is challenging due to their multivariate and nonlinear nature. Diverse computational methods are potentially available, but most are ineffective for protein-level data that is limited in extent and replication.

Results: We apply a decision tree approach to analyze the relationship of cell functional response to signaling activity across a spectrum of stimulatory cues. As a specific example case, we studied 5 intracellular signals influencing fibroblast migration under 8 conditions: 4 substratum fibronectin levels and presence vs. absence of epidermal growth factor. We propose techniques for preprocessing and extending the experimental measurement set via interpolative modeling in order to gain statistical reliability. For this specific case study, our approach achieves 70 % overall classification accuracy, and the decision tree model reveals insights concerning the combined roles of the various signaling activities in governing cell migration speed. We conclude that decision tree methodology may facilitate elucidation of signal-response cascade relationships and generate experimentally testable predictions, which can be used as directions for future experiments.

2.2 INTRODUCTION

Physiological cell behavioral functions, such as proliferation, death, differentiation, and migration, are governed to a large degree by networks of signaling proteins whose activities are influenced by a variety of extracellular cues: environmental agents such as chemical ligands, mechanical forces, radiation, toxins, pathogens, and so forth. Dysregulation of these networks is often associated with inappropriate cell and tissue behavior, so that signaling cascades are considered to be promising therapeutic targets for complex pathologies such as diabetes, cancer, and inflammatory diseases [181].

Quantitative experimental measurement of cell signaling protein properties - i.e., their levels, states (phosphorylation, cleavage, etc.), activities, locations - is more challenging, as opposed to gene - level measurements, to undertake in highly multivariate fashion. Consequently, while measurement of mRNA expression for hundreds and thousands of genes across a spectrum of conditions has become commonplace, analogous measurement of protein properties as listed above remains limited to the order of tens at best. A critical consequence of this situation is that many of the informatics methods by which computational analyses of genomic data are now being typically pursued are not readily applicable to proteomic (if that term can be used properly for coverage of only about tens of proteins) data. This is the problem that our effort here is directed toward addressing: finding appropriate computational techniques to elucidate useful models of the relationships between protein signals and cell functional responses to extracellular cues given the quantitative data across diverse conditions.

As a motivating case study, we consider cell migration, which is a central biological process in several pathological states such as tumor invasion as well as physiological ones such as wound healing [138]. Migration can be strongly influenced by both soluble environmental cues (e.g. growth factors and cytokines) and insoluble substratum cues (e.g. extracellular matrix proteins). In our specific experimental problem, we are studying the migration of tissue fibroblasts in response to four levels of surface fibronectin (Fn) concentrations in the absence or presence of epidermal growth factor (EGF), thus offering eight cue conditions. Fn is a ligand for integrin adhesion receptor-mediated signaling pathways and it has been shown to significantly effect migration of fibroblasts as well as other cell types including many tumor cells [58]. EGF also exhibits a strong influence on migration of normal tissue cells, including fibroblasts, and

various types of cancer cells, via signaling pathways mediated by EGF receptor (EGFR) [31]. Indeed, the EGFR system has been associated with the development and progression of a large number of tumors and is one of the most prominent pathways for therapeutic targets in human cancers. Furthermore, integrin and EGF pathways have been identified to crosstalk during cell migration, so it is highly relevant to study them together. While a very large number (in the dozens, easily) of signaling proteins downstream of integrins and EGFR potentially involved in regulation of migration can be identified, our experimental measurements focus here on the following five which have been shown to be among the key molecular switches in the motility signaling cascades: EGFR itself, extracellular-regulated kinase (ERK), myosin light chains (MLC), protein kinase Cδ (PKCδ), and phospholipase Cγ (PLCγ). These signaling proteins play significant roles in driving major biophysical processes, such as lamellipod protrusion, cell / substratum attachment and detachment, and cell contractile force generation and transmission, which underlie the net cell migration behavior [6]. Our experimental measurements are accomplished by quantitative immunoblotting, a standard but laborious procedure that typically limits the number of proteins and conditions which can be examined for any given situation under normal (at least academic laboratory) circumstances.

Many data-driven modeling approaches aim at finding correlations or cause-effect relations between genes or proteins. The resulting model is usually validated by comparing selected parts of the modeled relations with literature or with additional biological experiments without considering how good the model is for predicting outcomes of biological processes. In contrast, we are seeking to achieve two objectives in our analysis of signal transduction cascades. The first objective is to build a model from which the most relevant signaling proteins in regard to response can be identified. The second objective is to assess prediction accuracy of the model. The algorithmic methodology we propose to accomplish these goals is decision tree modeling. Often, as in our present case, the experimental data are noisy and the amount of observations is inadequate for dependency modeling or prediction. Therefore, before analyzing the data with decision trees, the data should be preprocessed and, if the amount of the data is insufficient for robust analysis, interpolative simulation of additional, internally-consistent data points might be considered as a computational aid.

The order of this study is as follows. First, we discuss an analysis of variance (ANOVA) based quality control approach, a minimum description length (MDL) based polynomial fitting

method to simulate data points and prediction with decision trees. Second, we apply these approaches to a case study, where we aim at classifying cell migration speed using phosphorylation levels of five signaling proteins.

2.3 APPROACH

In this section, a strategy to analyze a signal transduction cascade in regard to a cellular outcome is presented. Depending on how the signaling protein activation levels are measured, the resulting data set is often times noisy. Therefore, data quality control and normalization are imperative along the course of signal transduction cascade analysis. In Section 2.3.1 we present a quality control protocol for replicate measurements. In Section 2.3.2 we create additional data points via interpolating polynomial models that are chosen according to the Minimum Description Length (MDL) principle, and undertake validation efforts in Section 2.3.3. A procedure for constructing decision tree models is described in Section 2.3.4, and the experimental methods including data preprocessing are summarized in Sections 2.3.5 and 2.3.6.

2.3.1 Quality Control

A topic that has been somewhat neglected in several systems biology studies is data quality control. The objective of the quality control step is to identify samples that are aberrant due to non-biological reasons (e.g. technical or measurement errors). If such outliers are not identified, they may confuse the analysis method and result in spurious conclusions. On the other hand, a very stringent quality control criterion and discarding outliers without careful consideration can cause loss of valuable information. Therefore, measures taken after identifying an outlier sample should be dependent on the reasons why the sample was aberrant.

Here we present a statistical quality control algorithm for data sets consisting of multidimensional samples with replicates. Let vector $p_i \in R^{nx1}$ contain measurements for jth sample. We assume here that n is the same across all the samples but the algorithm below allows missing values. In our case study there are four fibronectin levels for each EGF level and since the EGF levels are dealt with separately until decision tree analysis and therefore $n = 4$ for each p_i. Further, let r_i denote the number of replicates for jth sample. Now, outlier replicate samples can be found using the following analysis of variance (ANOVA) based algorithm:

For jth sample

Test H_0: $\mu_1 = \mu_2 = \ldots = \mu_r$ with a one-way ANOVA and perform a multiple comparison for the ANOVA results using the Tukey-Kramer with significance level α.

If any of the replicate samples is aberrant, flag it according to the following rules**:**

R1 If a sample is statistically different from two or more samples, flag the sample.

R2 If there are several samples that could be flagged with R1, or two samples are statistically different, flag the sample whose deletion gives the minimum standard deviation for the means of the remaining samples.

Repeat (a)--(b) until H_0 is not rejected.

Repeat until all samples are processed.

The crux of the above algorithm is the ANOVA with the Tukey-Kramer multiple comparison test [185]. In general, the following assumptions are needed for the ANOVA:

- Samples are independent.
- Variances are constant across the samples.
- Observations are approximately Gaussian distributed.

As the quality control algorithm is applied to identify outliers among replicates and the replicates are usually measured with the same or similar kind of apparatus, it is reasonable to assume that variances are approximately the same. Further, except in cases of failures to clean or calibrate the measurement apparatus after use, samples should be independent. The ANOVA is not very sensitive to violations of the normality assumption, so the normality assumption is not a major one. Moreover, often times several independent sources affect the measurements and due to central limit theorem the data tend to be approximately normally distributed. The assumptions behind the ANOVA are usually fulfilled in biomedical research, so ANOVA based quality control algorithm could be applicable to many experimental setups.

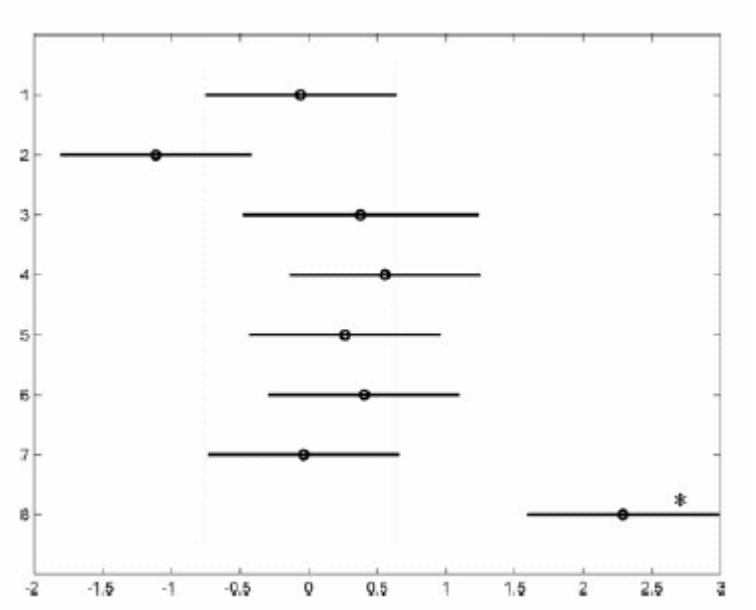


Figure 6. An example of quality control plot for Myosin light chain (MLC), EGF = 0.
A circle denotes a mean and a line corresponds to a comparision interval. Replicate 8 (marked with a star) is aberrant from all the other replicates and is discarded. Replicates 1 and 8 correspond to bands A and B in figure 2 respectively.

2.3.2 Parametric model for the data

An insufficient number of data points relative to the number of variables and interaction processes may impede or prevent identification of dependencies among the variables. A solution to this problem is to create a parametric mathematical model based on the data at hand, which is then used to interpolatively simulate additional data points so that dependencies between variables can be modeled and used in prediction. It is imperative to emphasize that the objective of this approach is to merely generate multiple realizations of pseudo-measurements that are internally consistent with the statistical distribution of the actual measurements, rather than creating new information in a more extrapolative manner.

If a preprocessed data set consists of several replicates, it may be worthwhile to replace replicate observations with a single value that is the most plausible value given the data. This value is referred to as a point estimate. Traditionally, the point estimator is chosen to be

(arithmetic) sample average because it is the best linear unbiased estimator for Gaussian distributed data and error estimates are straightforward to calculate. However, the breakdown point for the sample average estimator is $1/n$, where n denotes the number of data points, meaning that even one outlier drastically affects the point estimate. This is highly undesirable, and therefore we use the median, which has breakdown point of 1/2, as a point estimator.

One drawback with the median is that deriving error estimates analytically may be difficult. This drawback can be overcome with bootstrapping [183]: First, create B bootstrapping samples and compute median value for each bootstrapping sample. Error estimate for a point estimate is standard deviation of the bootstrapped medians.

With few exceptions, a trend for a biological process can be linear, biphasic or plateauing at one end. Several of these trends can be captured using polynomial models that have several benefits:

- Reliable polynomial modeling can be done with a relatively small sample size while still capturing highly nonlinear trends.
- Polynomial modeling is not confounded by few missing values.
- Discontinuous trends can be modeled with piece-wise polynomials.
- Polynomial fitting procedures, such as least-squares and maximum likelihood methods, are included in practically every statistical modeling software.
- Simulation of the polynomial model is straightforward and fast.

The polynomial equations for the parametric model data simulation are described in detail in Appendix A.

2.3.3 Validation of the parametric models

After a parametric polynomial model is constructed with the NML procedure, it is useful to check how good the model is for the original measurements. Since we assume the data to be approximately Gaussian, the goodness of the model can be checked by considering a Gaussian distribution whose mean is the simulated value and the standard deviation is obtained via Equation 7 (Appendix A). If each point estimate is located close to the mean and, for example,

not above or below 2.5 % of right and left tails, the model can be considered as statistically feasible.

It may also be useful to perform statistical tests such as the Z-test to test whether the point estimate (or original measurements) could originate from the model. If several point estimates belong to the extreme ends of the distribution, doubts may be cast over the validity of the model.

2.3.4 Finding dependencies between variables with the decision tree analysis

The majority of the studies in the field of systems biology aim at finding dependencies between variables. These models are, however, rarely used to predict the outcomes of cellular processes. In this section we provide means to achieve both of these objectives with decision trees [182]. Decision trees have several virtues that are useful for biomedical research:

- Decision trees can be effectively applied to any data structure, in particular to discrete, continuous or mixed data.
- Decision trees are capable of resulting in good prediction accuracies for highly nonlinear prediction problems.
- Prediction rules are easy to interpret.
- Decision trees perform a stepwise variable selection and reduce complexity.
- Decision trees are very robust against outliers.

The basic idea behind the decision trees is to first identify prediction rules from the data and then illustrate them as a binary tree where each terminal node (leaf) corresponds to a class and the other nodes represent measured variables. An example of a rule is "IF the phosphorylation level of ERK is high AND the phosphorylation level of MLC is high THEN cells migrate at medium speed." This rule can be readily seen in Figure 11. The rules are constructed by recursively splitting the data into smaller and smaller regions so that after each split the new data subset is "purer" than the old data subset [182]. A pure decision tree predicts all the classes in the training set correctly. In real world applications a pure (or close to pure) decision tree is very

large and almost surely suffers from overfitting. Thus, a decision tree is usually constructed in two phases. The first phase, tree growing, is done until splitting does not significantly improve the measure of purity. The second phase, tree pruning, is done in order to avoid overfitting. Here, we use the cost-complexity pruning approach [182] because we are able to create a separate pruning data set. Briefly, the tree pruning phase starts with a very large (overfitted) decision tree. The cost-complexity pruning method selectively produces a sequence of subtrees until only the root node is included in the subtree. In the cost-complexity pruning approach the sequence of subtrees is achieved by minimizing the sum of misclassification cost and the complexity of the tree. For profound discussion on the tree growing and the cost-complexity pruning methods we refer to [182].

In general, decision trees suffer from two drawbacks: Masking and instability [182]. Masking may occur if the relation between class (i.e. migration speed) and measured variables (i.e. signaling proteins) is very complex. In this case a variable may be partially duplicated by another variable and if two variables result in almost equally pure subsets, the level of noise may govern which variable is used in the splitting. If this happens in the early phase of tree growing, two decision trees may look dissimilar potentially hindering the interpretation of the results. In addition, masked variables may not show in the decision tree, which may again hinder understanding the results. These drawbacks are further discussed in Section 2.4.2.

2.3.5 Signaling protein experiments

We utilized NR6 mouse fibroblasts for our studies. These cells are derived from the 3T3 lineage and are devoid of endogenous EGF receptor (EGFR). We have overexpressed human EGFR in these cells, hence refered to as NR6 wild type (NR6 WT), and they provide and excellent model system to study EGFR mediated signaling events as well as cellular biophysical processes like migration [139]. Equal number of NR6 WT cells were plated on fibronectin coated surfaces and allowed to grow in alpha modified eagle's medium containing 7.5 % fetal bovine serum (FBS) for 24 hours, by which time cells reached about 90 % confluence. Fibronectin coating concentrations of the surfaces were 0.1, 0.3, 1 and 3 μg/ml $Fn \in \{0.1,\ 0.3,\ 1,\ 3\}$. Subsequently, cells were quiesced in a medium containing 0.5 % dialyzed (with minimum growth factors) FBS for another 24 hours, to remove the effect of exogenous growth factors

present in the serum. Cells were either lysed in the quiescent medium without any exogenous human EGF or stimulated with 10 nM (saturating concentration) of human EGF for five minutes. In the subsequent discussion, 0 nM EGF and 10 nM EGF conditions are denoted with *EGF* = 0 and *EGF* = 1, respectively. After stimulation, cells were washed once with ice cold PBS, and then lysed in lysis buffer containing 50 mM HEPES, pH 7.4, 150 mM NaCl, 1 % Triton X-100, 1 mM Na Vanadate and 10 % glycerol supplemented with protease inhibitors including 1μg/ml Leupeptin, 1 μg/ml Aprotinin and 1 mM Phenylmethylsulfonylfluoride (PMSF). Cell lysates were quantitifed using Biorad protein assay. Equal amount of total proteins were mixed with the loading buffer containing 4 % SDS (w/v), 0.1 M Tris-HCl, pH 6.8, 20 % glycerol, 0.2 % Bromophenol blue and 5 % β-mercaptoethanol, boiled for 5 minutes and then loaded on either 7.5 % (for analysis of pPKCδ, pERK, pEGFR, pPLCγ) or 15 % (for pMLC) SDS polyacrylamide gels. Cell lysates were resolved by electrophoresis and subsequently transferred onto nitrocellulose membranes, after which, membranes were immunoblotted with specific antibodies to detect the specific proteins or their activated phospho-protein forms.

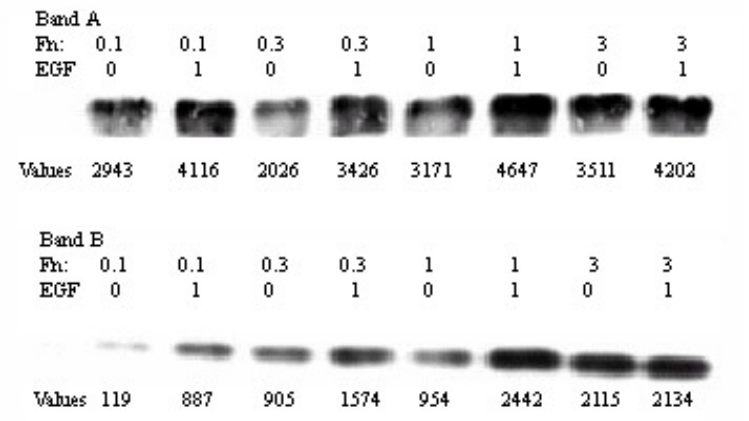


Figure 7. Two immunoblots for myosin light chain (MLC).
All Fibronectin (FN) and EGF conditions with non-normalized values are shown. FN values of the substrates were 0.1, 0.3, 1, 3 μg/ml and EGF was used as 10 nM.

Protein	EGF = 0	EGF = 1	Total
EGFR	7	7	8
ERK	5	8	8
MLC	6	5	8
PKCδ	7	6	8
PLCγ	5	5	5

Table 1. The number of replicates for signaling proteins before and after quality control.

Protein	$\hat{\beta}_{\mathrm{EGF}=0}$	$\hat{\sigma}_{\mathrm{EGF}=0}$	$\hat{\beta}_{\mathrm{EGF}=1}$	$\hat{\sigma}_{\mathrm{EGF}=1}$
EGFR	$0.24x - 0.02$	0.10	1.9	0.15
ERK	0.51	0.25	4.1	0.20
MLC	0.20	0.04	$0.08x + 0.48$	0.04
PKCδ	0.06	0.06	0.32	0.04
PLCγ	$0.36x - 0.86$	2.6	$0.22x + 2.38$	0.38

Table 2. Polynomial estimates and standard deviation estimates for signaling proteins under presence and absence of EGF using NML criterion.
β is polynomial estimate and σ is standard deviation estimate.

2.3.6 Data preprocessing

Immunoblots were quantified with the NIH image analysis densitometry software. The software generates an area plot for each protein band, the density of which represents the amount of the protein in each lane. In the signaling protein experiments, the quantitative values generated represented the activated status of a protein since the proteins detected were in their activated or phosphorylated state. Examples of two immunoblotting bands are given in Figure 7.

For the NML and decision tree analysis, the band densities were normalized by the value of the first lane ($Fn = 0.1$ and $EGF = 0$) for each immunoblot (between-band normalization). After this normalization, results become comparable between immunoblots since the experimental conditions in each of the experiment were kept constant. For quality control, the bands were within-band normalized: all protein conditions in a band without exogenous EGF were normalized by the value with $EGF = 0$ and $Fn = 0.1$, while all protein conditions in a band with exogenous EGF were normalized by the value with $EGF = 1$ and $Fn = 0.1$. The within-band normalization ensures that proteins under the same EGF condition within a band are comparable. Prior normalization all basal values below 250 were converted to 250 in order to prevent division by a small value that is likely due to noise. After normalization, all the values were $\log_2$-transformed.

Normalization was followed by the ANOVA based quality control approach (Section 2.3.1) with $\alpha = 0.05$. An example of a quality control plot for MLC is given in Figure 6. Replicate 8 (marked with a star) is aberrant from the seven other replicates and is discarded. Also replicate 2 is discarded due to rule R1 given in Section 2.3.1. Replicates 1 and 8 correspond to bands A and B in Figure 1, respectively. The numbers of the replicates before and after the quality control are given in Table 1.

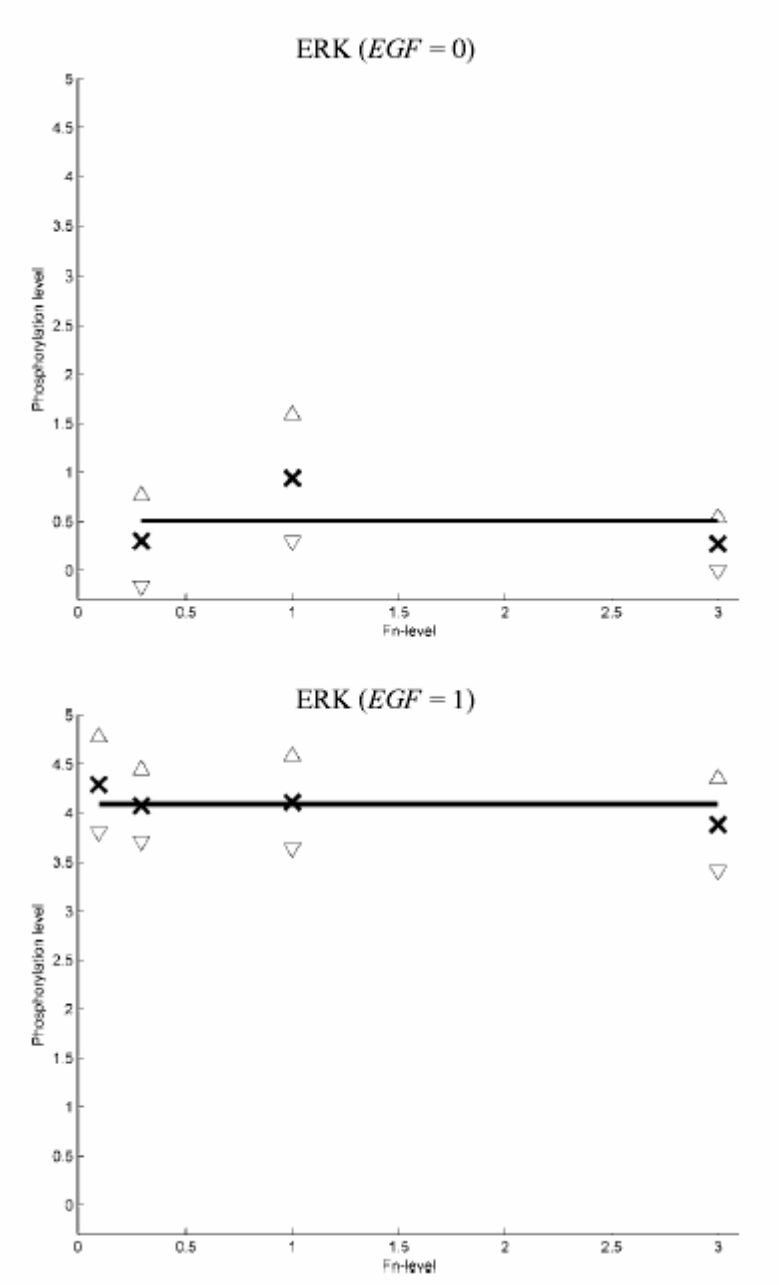


Figure 8. Point estimates, cross; upper triangle and lower inverted triangle, error estimates and fitted polynomial for ERK across Fibronectin levels.

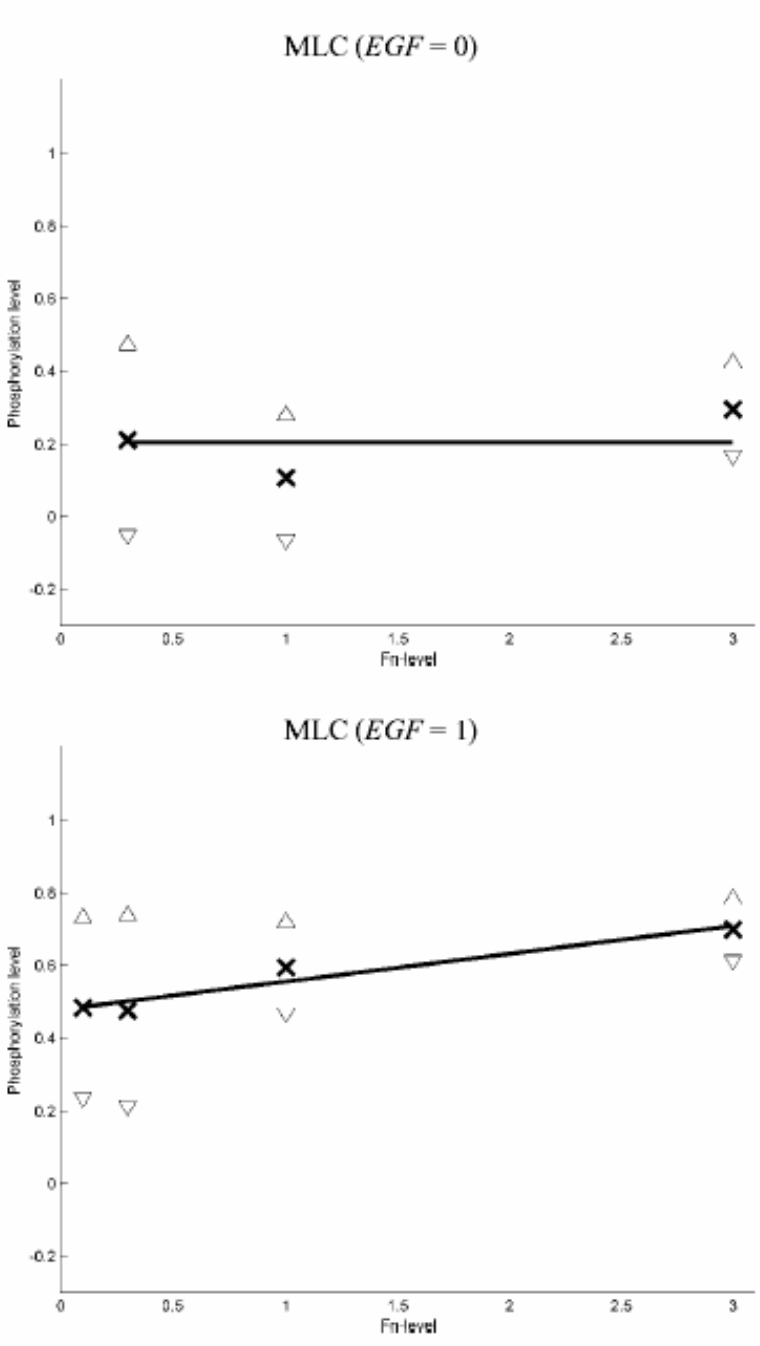


Figure 9. Point estimates, cross; upper triangle and lower inverted triangle, error estimates and fitted polynomial for MLC across Fibronectin levels.

2.4 RESULTS

Cell migration is a crucial cellular function that contributes, for example, to wound healing, normal immune responses, as well as drives progression of diseases such as tumor invasion and metastasis [138]. In general, migration consists of a complex assembly of five biophysical processes: polarization, protrusion, adhesion, contractility and retraction. While the effects of biophysical processes to migration speed are somewhat well-known, the effects of signaling proteins that govern these processes and their dependencies are less so. In this section we explore how five signaling proteins (EGFR, ERK, MLC, PKCδ and PLCγ) affect cell migration speed under combinations of two extracellular cues using the methods discussed in Section 2.3.

The cues used here are four different surface fibronectin concentrations with or without additional stimulation with EGF. Earlier studies [31] have shown that if migration speed is measured as a function of fibronectin levels, presence or absence of EGF has a dramatic impact on migration speed: If EGF is present, migration speed is biphasic, while in the absence of EGF, cells migrate at a constant speed. The data in [31] consist of four fibronectin-levels for 0nM and 25nM EGF, resulting in eight measurements. Accordingly, we measured the phosphorylation levels of the five signaling proteins using the same condition for $EGF = 0$ (0nM EGF) as in [31]. For condition $EGF = 1$, we used 10nM EGF for the signaling proteins, while it was 25nM in [31]. Since both 10nM and 25nM EGF are identical in motility and both are saturating, the data for signaling protein and cell migration speed are comparable.

2.4.1 Parametric model for the signaling proteins and migration speed

Having four observations per one *EGF*-level is not enough for reliable identification of dependencies between the signaling proteins and migration speed. Therefore, we applied the procedure given in Section 2.3.2 to generate more data using simulation. Before applying the NML approach, we computed median phosphorylation level for each protein for all Fn-levels using the data from replicate experiments. Each median was accompanied with an error estimate that was computed with bootstrapping ($B = 5{,}000$).

Due to the small number of the data points we restricted the maximal polynomial degree in the NML approach to two, i.e. $\Omega = \{1,2,3\}$. Further, the polynomial models were constructed

separately for the values under $EGF = 0$ and $EGF = 1$ conditions. Polynomial orders for the signaling proteins using the NML criterion are given in Table 2. An example of the polynomial models and associated point and error estimates for ERK and MLC is given in Figures 8 and 9.

Standard deviations and point estimates for migration speed are given in [31]. We computed pooled standard deviation with Equation. 7, where we made a conservative approximation that $r_i = 70$ since the estimates were based on 70-100 cells. The pooled standard deviations were 3.0 and 3.4 for $EGF = 1$ and $EGF = 0$. Polynomial fitting for migration speed was done in log-log space based on the model validation procedure depicted in Section 2.3.3. For $EGF = 1$ the polynomial order was two ($-0.49x^2+0.07x+5.8$), while for $EGF = 0$ it was zero. As our objective was to predict slow, at medium speed or fast migrating cells, the values were further discretized into three categories (slow, medium speed, fast) using the Lloyds algorithm [186], where the training data were obtained from the noiseless polynomial model. The model, discrete categories, simulated data and the original measurements for migration speed are illustrated in Figure 10.

2.4.2 Decision tree for migration speed

Using the polynomial models we simulated observations between $Fn = 0.1$ and $Fn = 3$ using $\Delta = 0.0001$ resulting in 58,002 observations per variable. The protein phosphorylation values were discretized with the Lloyds algorithm so that the number of the discrete categories equaled to the number of parameters in the polynomial models for each protein: EGFR = {0,1,2}, ERK = {0,1}, MLC = {0,1,2}, PKDδ = {0,1} and PLCγ = {0,1,2,3}. With this discretization approach proteins that are affected more by the extracellular stimuli, and thereby can be considered more informative, are described with more discretization categories than proteins with lower polynomial degrees. Discrete categories reflect relative phosphorylation levels. For example, given that there are two and three discrete categories for ERK and MLC, respectively, ERK = 1 denotes that ERK is highly phosphorylated, while MLC = 1 means that the phosphorylation level of MLC is medium.

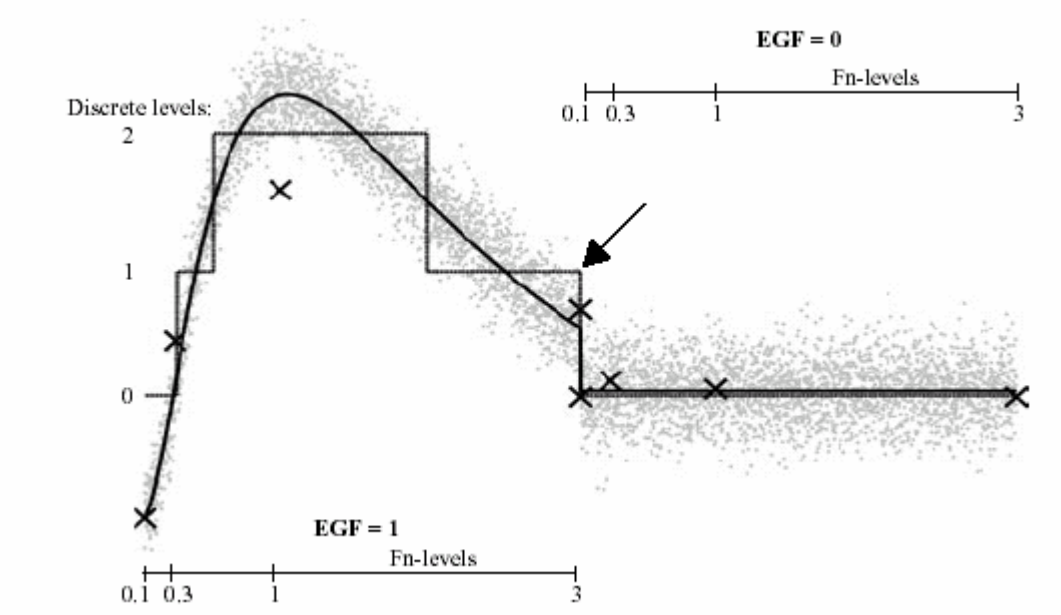


Figure 10. Polynomial fit (solid line), discretization categories (dotted line), original observations (cross) and simulated, noisy data (dots) for migration speed.
The change from EGF = 1 and EGF = 0 is marked with an arrow.

In order to overcome the instability problem with the decision trees we first constructed 10,000 decision trees without pruning. The parameters for growing the decision trees were as follows. Splitting criterion was Gini-index [182], prior probability for minimum number of observations for impure nodes to split was set to be five. We defined the misclassification costs to be such that for misclassifying a slow (medium) speed to medium speed (fast) cost is one, but if slow speed is misclassified to fast, cost is two. This resulted in the following cost matrix

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix},$$

where each row and column corresponds to a migration speed category.

After the tree growing phase, all 10,000 trees were pruned with the cost-complexity pruning method [182]. For the validation step we created 1,000 data sets. Due to computational

reasons Δ was set to 0.001, so each validation and pruning data set consisted of 5,802 observations. The following criterion was used to choose the best tree model:

$$D = \frac{1}{T}\sum_{i}^{T}\left|y_i - \hat{y}_i\right|, \qquad (7)$$

where y_i denotes true classes for ith test data set, y_i denotes predicted classes and T is the number of test data sets (here 1,000). After the pruning and validation, there were 23 separate decision tree models and the best decision tree was the one that minimized Equation. 7. Although the best decision tree is chosen using Equation. 7, we report also mean classification accuracy (CA), which is the mean of the classification accuracies across 1,000 test cases. Classification accuracy corresponds to the number of correct classifications divided by all cases. The best decision tree (CA = 70 %) is given in Figure 11. Round nodes correspond to the signaling proteins and square nodes to the migration speed classes. Classification rules and their relative importance can be seen easily from Figure 11. For example, IF ERK = 1 AND MLC = 1, THEN cells migrate fast, and 62 % of the measurements for the fast migration class (in the training set) can be explained with this rule.

If the signaling proteins were not discretized, the best decision tree consists of only MLC and PLCγ (results not shown) and CA was slightly below 70 %. Now, based on this decision tree graph, it could be argued that cell migration speed is dependent only on MLC and PLCγ, and ERK is irrelevant when predicting cell migration speed. However, earlier studies have shown that ERK is one of the key signals governing migration speed [79, 140] so its absence in the decision tree model was unexpected. When we searched for explanations for the exclusion of ERK it turned out that ERK was masked by MLC. This can be seen by comparing data for ERK and MLC in Figures 8 and 9. When *EGF* = 0, both ERK and MLC are constant with approximate the same level of phosphorylation. However, when *EGF* = 1, phosphorylation levels for ERK are again almost constant but very high, whereas MLC activity is increasing linearly. Thus, the decision tree growing algorithm considered ERK useless for migration speed classification given MLC, which was undesirable. While some guidelines for detecting masked variables are given in [182], these are not helpful in getting the masked variable into the model.

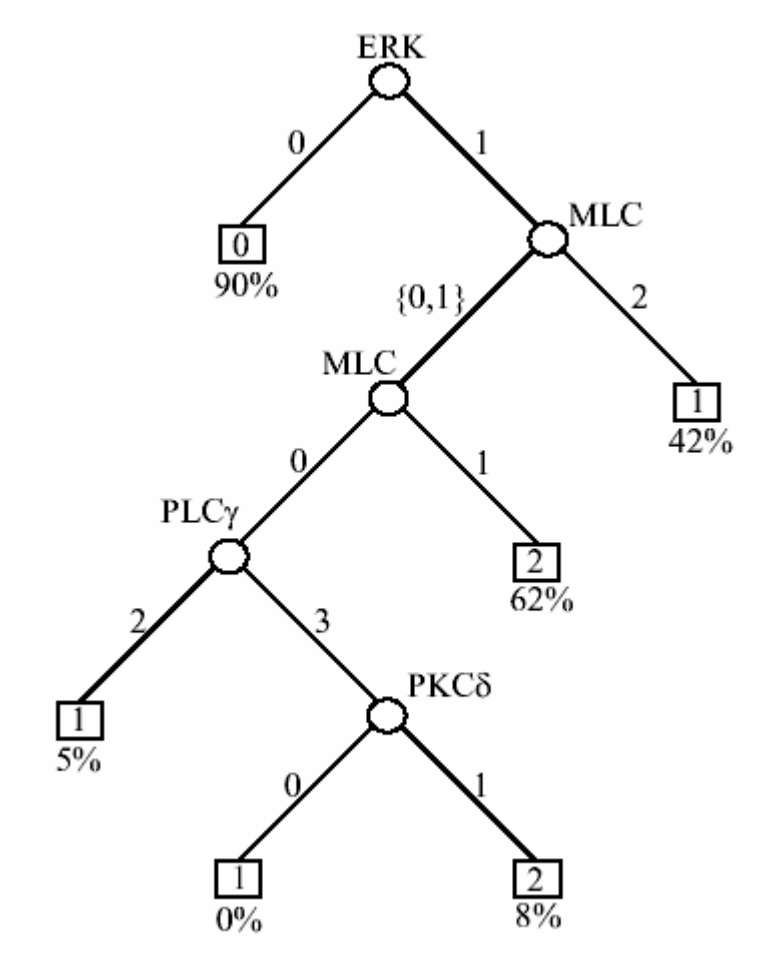


Figure 11. The best decision tree for classifying migration speed using signaling proteins.
Round nodes denote variables whereas square nodes denote migration speed categories. Integers attached to the arc correspond to the split of the parent nodes. Under each migration speed category the fraction of cases explained by that classification rule is given. For example, if ERK = 0, the migration speed category is 0 and 90 % of the observations (in the training set) for the migration speed category 0 can be explained by this rule.

This case study provides a proof-of-principle that the proposed data-driven modeling approach is applicable to biomedical research. Accordingly, detailed discussion on biological implications of the results is out of the scope of this study and will be elaborated elsewhere

(manuscript in preparation). Briefly, ERK and MLC regulate the adhesion / contraction ratio [34, 140], which is one of the most important biophysical processes during the migration cycle [6]. Thus, it is not surprising that these two proteins together are fairly good predictors for migration speed. This immediately suggests that in further studies, ERK and MLC should be studied together rather than individually.

2.5 DISCUSSION

Analysis of signal transduction cascades is an important application in several biomedical research. In this study we have presented a data-driven modeling approach to perform such analysis. In our case study we have applied the modeling approach to model and predict whether cells are moving slowly, at medium speed or fast, using a set of intracellular signaling proteins under various levels of fibronectin and EGF cues. The resulting decision tree graph indicates that the phosphorylation level of ERK alone tells whether cells are migrating slowly. In order to obtain higher classification accuracy for the cells migrating at medium speed or fast, also MLC, PLCγ and PKCδ are needed. These results highlight the central idea of systems biology, i.e. complex biological processes cannot be analyzed by perturbing only one component at a time but there is a need to study several components simultaneously. However, usually it is not known what these components are. Our results indicate that the decision tree analysis can be used to suggest what components should be studied together.

Based on Table 2 it can be argued that the information content of the signaling protein data set is low. Degrees of the polynomial functions, however, do not tell the whole truth. For example, ERK has zero order polynomial in the presence and absence of EGF, but the absolute difference between these constants is large. That is, ERK acts like a switch triggered by the EGF status and clearly brings in information to the analysis. Further, after inducing EGF, cell migration measurements were performed after eight hours in order to observe maximal migratory response for the cell type used in this study [31]. Therefore, it was expected that in five minutes the changes in the signaling proteins phosphorylation levels may not yet be visible at the migration speed level. From another standpoint, the conclusion that a data set is not rich in its information content may also be valuable information, and the NML modeling with decision tree analysis provide means to assess this issue. One of our future directions is to measure signaling protein activities in regard to cell migration speed at different time points, and the methods given here can be used to approximate time it takes for extracellular stimuli to have an effect on the signaling proteins. These results may be used when estimating rate constants for temporal mathematical modeling.

Quality control is an unavoidable part of biomedical research; studies not performing quality control implicitly state a 100 % confidence in the measurements. In this study we applied

a statistical quality control approach for replicate measurements with 95 % confidence level. We also tested 80 % confidence level and no quality control and the resulting decision trees resulted in 62 % and 67 % classification accuracies, respectively (results not shown). Detailed discussion on quality control issues and choices of the confidence levels is beyond the scope of this study, but it would make an interesting topic for further study. In essence, one of our future directions is to first identify quality features, learn a classifier with them, and use the trained classifier to assess quality control as described in [141].

The overall obtained classification accuracy for simulated data, 70 %, for three migration speed categories is quite good given that original signaling protein measurements were done at one time point of only five minutes after stimulation, whereas cell migration was measured at eight hours. Furthermore, cell migration speed data and signaling protein measurements originate from different studies: The cell migration data set was done in 1999 while signaling protein data set was done in 2004. As a consequence, there are some differences between the experimental setups causing noise to the analysis. Classification accuracy can also be used as a yard-stick for sufficiency of the measured data set in regard to modeling a biological process. If classification accuracy is poor, it could be an indication that the data set does not comprise enough variables or information in order to model the biological process in question. In our case study, the classification accuracies for medium and high migration speeds were fair. The most likely reason for this is that our measurements cover only a limited portion of the signaling network components critically involved in governing migration. Merely as one relevant facet of this highly multi-variate system, for instance, there is accumulating evidence that virtually all of the key MAP kinases influence cell motility in diverse ways [142]. This shortcoming can, of course, be addressed by enlarging the scope of the measured signaling component space to the extent cost-effective. Accordingly, the decision tree results can be helpful to determine what components should be measured in future experiments, and whether there is a need to measure additional components. On the other hand, our results demonstrate that decision trees are applicable to studies where several key components are not observed.

To our knowledge this is first study where cell migration speed is quantitatively predicted using phosphorylation levels of signaling proteins. Several other modeling approaches such as Bayesian networks [137,187], neural networks or support vector machines [184] have their own benefits and drawbacks. The two latter methods are very good classifier approaches in various

applications but they suffer from a major drawback; dependencies between the variables and their relevancies are practically impossible to obtain from the model. In contrast, Bayesian networks have been mainly used to obtain dependencies between variables but it is not self-evident that a Bayesian network that aims at describing dependencies between the variables performs well when predicting cellular outcomes. Moreover, the variables used in learning a Bayesian network are usually required to be discrete or Gaussian distributed, which may be an implausible requirement.

The decision tree based modeling is not supported by a unique and solid mathematical background. Thus, it is imperative to report parameter settings in detail so that the results can be reproduced. Furthermore, decision trees require a relatively large training data set, which may not be feasible to obtain. This requirement, however, is not unique to decision trees but is present with the other classification and modeling approaches as well. Here we have expanded the data set via interpolating polynomial functions whose order was determined with the MDL principle. Parameters for polynomial functions are straightforward to estimate and several well-established methods exist for this purpose. When polynomial functions do not yeild satisfactory results, the alternative might be Monte Carlo based techniques. However, as Monte Carlo methods are notorious for being computationally demanding, we argue that the polynomial models should be applied before trying more complex methods.

When additional data are simulated, it is important to choose the extracellular conditions so that they span over a large range because it is safer to interpolate than extrapolate. Another requirement is that there should be enough data points so that nonlinear trends can be captured. The methods presented in this study do not pose upper limits for the extracellular cues but in its current form the decision tree analysis can be applied to only one biological process at a time. One of our future directions is to develop a multidimensional decision tree that is capable of predicting several cellular outcomes simultaneously. A multidimensional decision tree would enable, for example, identification of signaling proteins that are associated with high cell migration speed and avoidance of apoptosis.

2.6 CONCLUSIONS

We have presented a decision tree-based modeling approach for analysis of complex and multidimensional signal transduction cascades. Our case study demonstrates that decision trees can provide several insights to signal transduction cascades. We conclude that decision tree methodology may facilitate elucidation of signal-response cascade relationships and generate experimentally testable predictions, which can be used as directions for future experiments.

2.7 ACKNOWLEDGEMENTS

We thank Dr. Fei Hua for constructive suggestions regarding the manuscript. This work was supported by the NIGMS Cell Migration Consortium, NCI grant CA88865 to DAL, and the Academy of Finland and Emil Aaltonen Foundation.

3.0 PREDICTING BIOPHYSICAL RESPONSES OF CELL MOTILITY USING DECISION TREE ANALYSIS OF INTRACELLULAR SIGNALING CASCADES

Sourabh Kharait[1*], Sampsa Hautaniemi[2,3*], Akihiro Iwabu[1], Shan Wu[2], Douglas A. Lauffenburger[2], Alan Wells[1]

[1]Department of Pathology
University of Pittsburgh Medical Center
Pittsburgh, PA 15213

[2]Biological Engineering Division
Massachusetts Institute of Technology
Cambridge, MA 02139

[3]Institute of Signal Processing
Tampere University of Technology
Tampere, Finland

* Equal contributors

3.1 INTRODUCTION

Physiological cell behaviors are, to a large extent dependent upon various extracellular cues like chemical ligands, micro-organisms, toxins, radiation, and so forth. Such stimuli exhibit distinct cell responses by selectively mediating signaling from signal transduction cascades. Signaling events in turn ensue due to the spatial and temporal fluctuations in the activation status of numerous proteins that act like switches within larger proteomic networks [8]. Such proteins are activated epigenetically by phosphorylation of various residues, cleavage by proteases, enzymatic activity and subcellular translocation making such protein measurements challenging. However, to modulate cell behavior in therapy, a thorough understanding of these biochemical switches in needed so they can be effectively targeted.

Migration of a variety of cells is crucial for various homeostatic biological responses during wound healing and inflammation [1]. Such biological property of cell motility is dysregulated in cancers leading to cancer progression and metastasis [58]. Thus targeting motility, can be employed in limiting mortality and morbidity of a variety of human diseases including cancer [143]. But such targeting suffers from limitations since variety of cancers utilize alternative pathways within the protein networks to promote their progression. While each of such target disruption has shown to be effective in abrogating motility and invasiveness of tumor cells *in vitro*, this cannot be readily applied to *in vivo* situation due to our limited understanding of how the intracellular signaling networks work. Thus, a clear delineation of the interplay of key proteins mediating cellular properties is crucial to future efforts aimed at drug discovery and individualized treatment [144, 145].

Targeting cell motility is a challenge given its complexity. One approach towards understanding motility is to break it down into discrete and individual biophysical components [6, 138]. The principal required processes include acquisition of cell directionality with a front and a rear end with lamellipodal protrusion at the front and detachment of cell membrane at the rear (Figure 1). The polarized cell then contracts using the ubiquitous actin-myosin contractile machinery to produce the force needed for locomotion. Thus productive migration ensues due to the repetitive cycling of these complex biophysical events in a temporally organized manner. It is thus evident, that such a complex event is exhibited by a coordinated signal propogration and amplification / attenuation within existing intracellular proteomic networks. Our goal is to define

key signaling switches governing cell migration that can be targeted for modifying this cellular behavior.

Research efforts in the field of cell biology have generated enormous 'raw-data' that needs further stratification using sophisticated methods. Computational modeling can compile and classify gigantic proteomic data sets produced by experimental laboratory techniques and extract vital information not readily apparent by conventional analytical techniques. In addition, mathematical models can expand data sets to proportions that can be used to make non-intuitive predictions related to biological responses [11]. We have previously described [11] a novel approach, namely Decision tree analysis (chapter 2), in studying cell migratory events based on measurements of key intracellular signaling proteins. This study was inspired by previous observations from Maheshwari et al that elucidated the biophysical components of fibroblast migration across a range of different extracellular cues. Different cellular biophysical processes including cell speed were measured across 8 different experimental conditions (4 different surface FN levels and presence or absence of EGF). The observations indicated that cells move fastest with EGF stimulus when the surface fibronectin concentration (or cell-substratum adhesiveness) is in the intermediate range whereas minimal motility was observed at the two extreme conditions. However, substratum fibronectin concentration (and also the extracellular matrix) alters motility not just by altering surface adhesiveness but also by actively signaling through the integrins towards intracellular downstream cascades [29]. Thus in this study we aimed at elucidating the 'quantitative contributions' of different signaling proteins in dictating cellular motility across different extracellular cues (FN and EGF). In this study as well, we measured, using quantitative western blotting, the activation status of five key signaling proteins under the same 8 experimental conditions. Decision trees were created by compiling this data with the measurements of cell speed from Maheshwari et al. The decision trees were applied to map the hierarchial interplay of signaling proteins in governing crucial cellular biophysical events (including cell speed). Our model places MLC mediated cell transcellular contractility as the most vital element in motility. Further, the model also predicts that differences in its quantitative inhibition (total versus subtotal) by targeting MLC can have drastically divergent cellular effects. We have tested these non-intuitive predictions from the model using experimental methods thereby validating the model. Such studies have profound implications in

therapy in identifying crucial signaling nodes that should be quantitatively disrupted for the most significant biological response.

3.2 APPROACH

Detailed experimental protocols are provided in chapter 2.0 under specific sections. To avoid unnecessary repetition, only a brief overview of the approach and experimental methods is presented in this chapter.

The study is conducted as follows: We first generate quantitative immunoblotting data under the specific conditions of four different fibronectin levels of the substratum and presence or absence of EGF (FN of 0.1, 0.3, 1, 3 μg/ml and – or + of 10 nM EGF). We have previously measured key biophysical processes across the same experimental conditions. EGF of 10 nM is saturating of EGF receptor and was added for 5 minutes (acute), 1 hour (intermediate) and 16 hour (long term) to elucidate signaling events during a wide range of time. The immunoblots were quantitated using NIH image to generate quantitative data followed by normalization using the first experimental condition i.e. FN of 0.1 μg/ml and no EGF (see section 2.3.5 / 6). We utilize ANOVA based quality control to exclude any data point if it classified as an 'outlier' depending upon the rest of the samples within that data set (section 2.3.1). We then expand the data using polynomial modeling and generate approximately 50,000 replicates of each signaling protein within the 8 different experimental conditions (4 FN levels and – or + EGF) (section 2.3.2). Similarly, we use polynomial modeling to generate additional replicates for cell speed based on the initial measurements from Maheswhari et al [31] within the 8 experimental conditions (four different FN levels and – or + EGF). We then generate decision trees by integrating this quantitative data of biochemistry (western blots) and biophysics (cell speed across fibronectin) that provides a hierarchial map of key intracellular switches in governing cell speed (section 2.3.4 and 2.4.2).

3.3 RESULTS

3.3.1 Quantitative immunoblotting

Our aim is to elucidate the relative contributions of different signaling proteins in mediating migration across different extracellular conditions. The biophysical data included cell speed, membrane protrusion activity, cell spread area, surface adhesion, and membrane retraction and has been previously measured by Maheshwari et al [31]. We utilized 10 nM of EGF for our studies in NR6WT cells whereas Maheshwari et al have utilized 25 nM. Both of these concentrations are saturating for the EGF receptor numbers in these cells and thus can be assumed to be similar in their cellular effects. In addition, EGF was added to the cells for periods of 5 minutes, 1 hour and 16 hours to capture the entire (temporal) activation spectrum of signaling proteins.

Addition of EGF activated EGFR within minutes and this signal was transmitted downstream to all other signaling cascades measured (Figure 12). Interestingly, EGFR activation profile mirrored that of ERK within early time periods of EGF stimulation (5 minutes). ERK activation was robust compared to quiesced (control) cells and was minimal at 1 hour (of EGF stimulus) due to signal attenuation with minimal change over different surface fibronectin concentrations (Figure 13). Thus ERK functioned like a 'switch' turned on dependent mainly on EGFR signaling. EGFR signaling also activated PLCγ and PKCδ linearly across increasing surface FN levels with resultant MLC activation through PKCδ. MLC activation begins within a few minutes of EGF stimulation and reaches a plateau at about 2 hours and can be appreciated upto 24 hours after EGF stimulus. Interestingly, MLC activity was inversely biphasic with lowest levels at intermediate FN concentration (0.3 and 1 μg/ml) captured after 1 hour of EGF stimulation (Figure 13). Thus, using these experimental conditions, we captured important quantitative and temporal trends of molecular activation. One representative immunoblot for 5 minutes and 1 hour of EGF stimulation is shown (Figures 12 and 13). Quantitation and normalization of this data has been previously described in detail (sections 2.3.5 / 6).

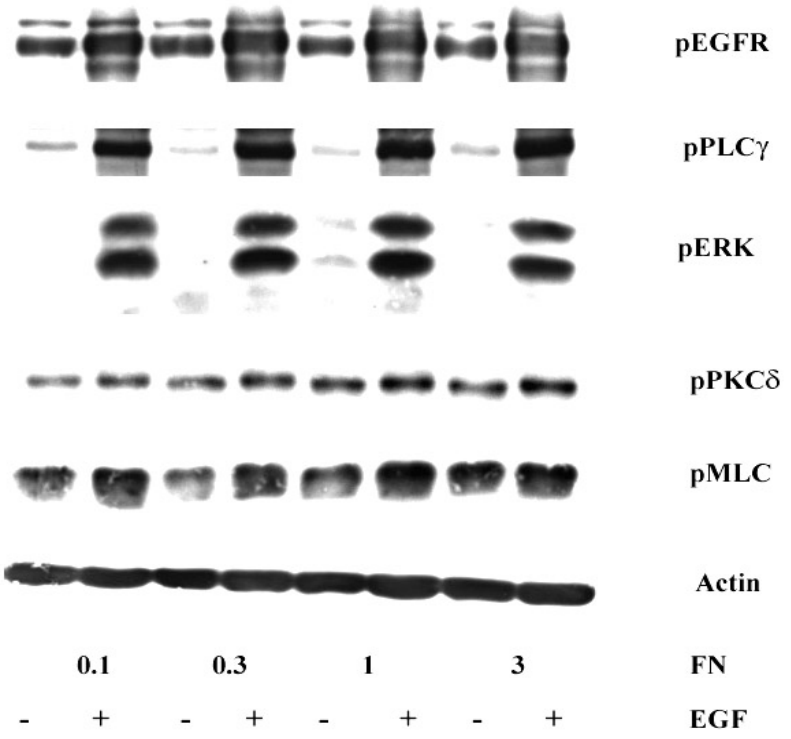


Figure 12. Western blotting data for EGF of 5 minutes across different fibronectin concentration of surfaces. Tissue culture plates were coated with different fibronectin (FN) concentrations. NR6WT cells were grown on these surfaces for 24 hours in complete growth medium and quiesced for another 24 hours in medium containing 0.5 % dialyzed FBS. EGF was added for a period of 5 minutes, cells washed once with PBS and lysed. Cell lysates were resolved using SDS-PAGE and immunoblotted using specific antibodies for various phosphorylated proteins.

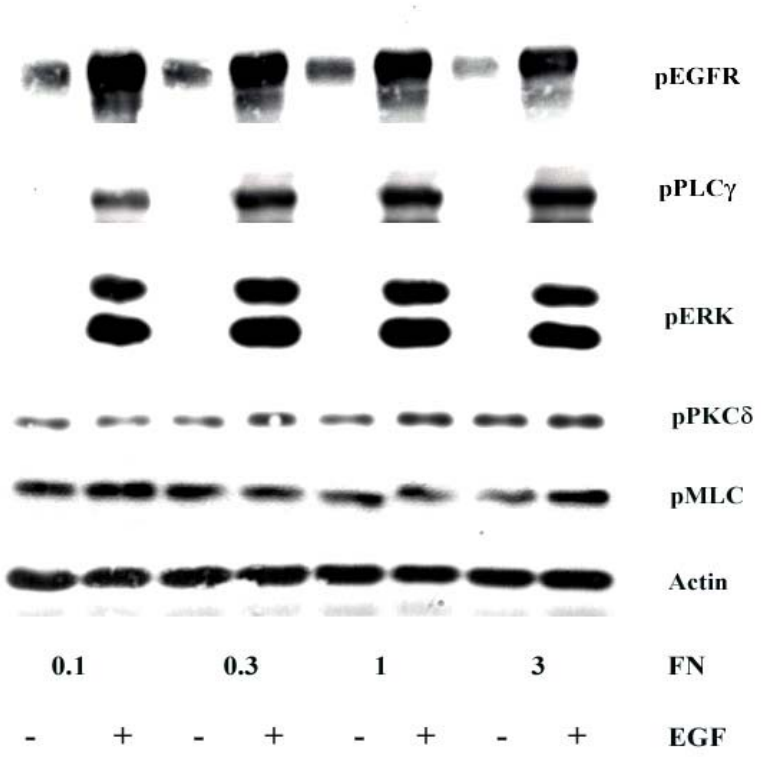


Figure 13. Western blotting data for EGF treatment of 1 hour across different fibronectin concentration of surfaces. Tissue culture plates were coated with different fibronectin (FN) concentrations. NR6WT cells were grown on these surfaces for 24 hours in complete growth medium and quiesced for another 24 hours in medium containing 0.5 % dialyzed FBS. EGF was added for a period of 1 hour, cells washed once with PBS and lysed. Cell lysates were resolved using SDS-PAGE and immunoblotted using specific antibodies for various phosphorylated proteins.

3.3.2 Data expansion using polynomial interpolation

An insufficient number of data points can limit effective conclusions that can be drawn about dependencies among variables (signaling proteins). Thus, the measured experimental data has to be expanded using mathematical models to interpolate new data points. These simulated data points are multiple realizations of pseudomeasurements that are internally consistent with the statistical distribution of the original measured data points [11]. The details of creating polynomial equations and their application to data interpolation have been previously explained in section 2.3.2 and Appendix A. The polynomial interpolations for cell speed, based on the original data from Maheshwari et al, is shown in figure 14. Similarly, polynomial data for phospho-MLC is shown in figure 15. The data points are discretized in three levels depending upon the activation status as 0= low, 1= medium, 2= high. This is true both for signaling proteins as well as cell speed measurements. Crosses denote the actual measurements, either cell speed or signaling protein data, whereas the red lines denote the distribution of 'simulated' data points. It is evident that pMLC is inversely biphasic across fibronectin coating concentration of surfaces with lowest levels found at intermediate FN levels and highest levels found either below FN of 0.5224 μg/ml or more than 2.6 μg/ml. This also corresponds to highest cell speed observed within this FN concentration range. The interpolated data from all other signaling proteins (along with MLC) was employed for constructing decision trees.

Once the polynomial model is constructed, it is necessary to check how good the new 'simulated' data set (and hence the polynomial model) is for original measurements. The new data is assumed to be in Gaussian distribution with its mean as the simulated value and standard deviation estimated from equation 7 (appendix A). If each of the data points lies close to the mean and does not fall below or above the 2.5 % of the left or right tails, the model can be considered statistically feasible. Examples of polynomial modeling for signaling protein measurements at 5 minutes of EGF stimulation (the variables) is shown in figure(s) 8 and 9. A similar example of polynomial modeling and data simulation for cell speed (the response) is shown in figure 10.

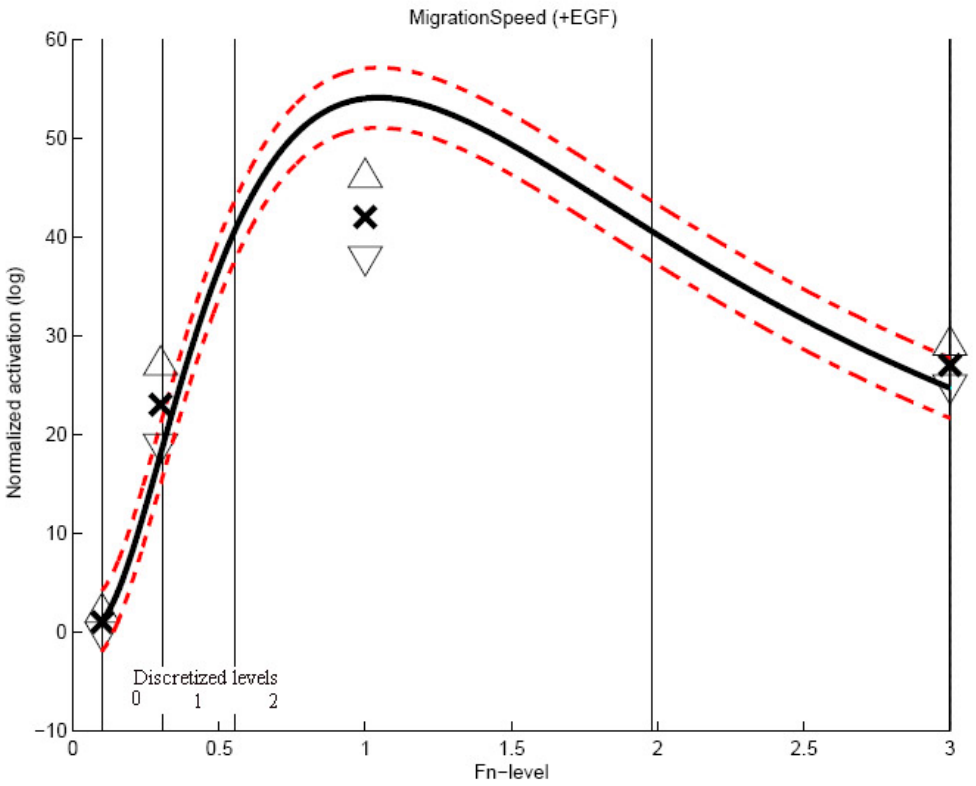


Figure 14. Polynomial interpolation data for cell speed under EGF stimulation.
Crosses are actual measurements, upper and lower triangles are error estimates red line denotes the range of distribution of the interpolated values using fitted polynomial across a range of fibronectin coating concentrations. Cell speed is biphasic and discretized as low (0), medium (1) and high (2) across FN; i.e. High speed is seen between FN of approximately 0.52 to 2 μg/ml.

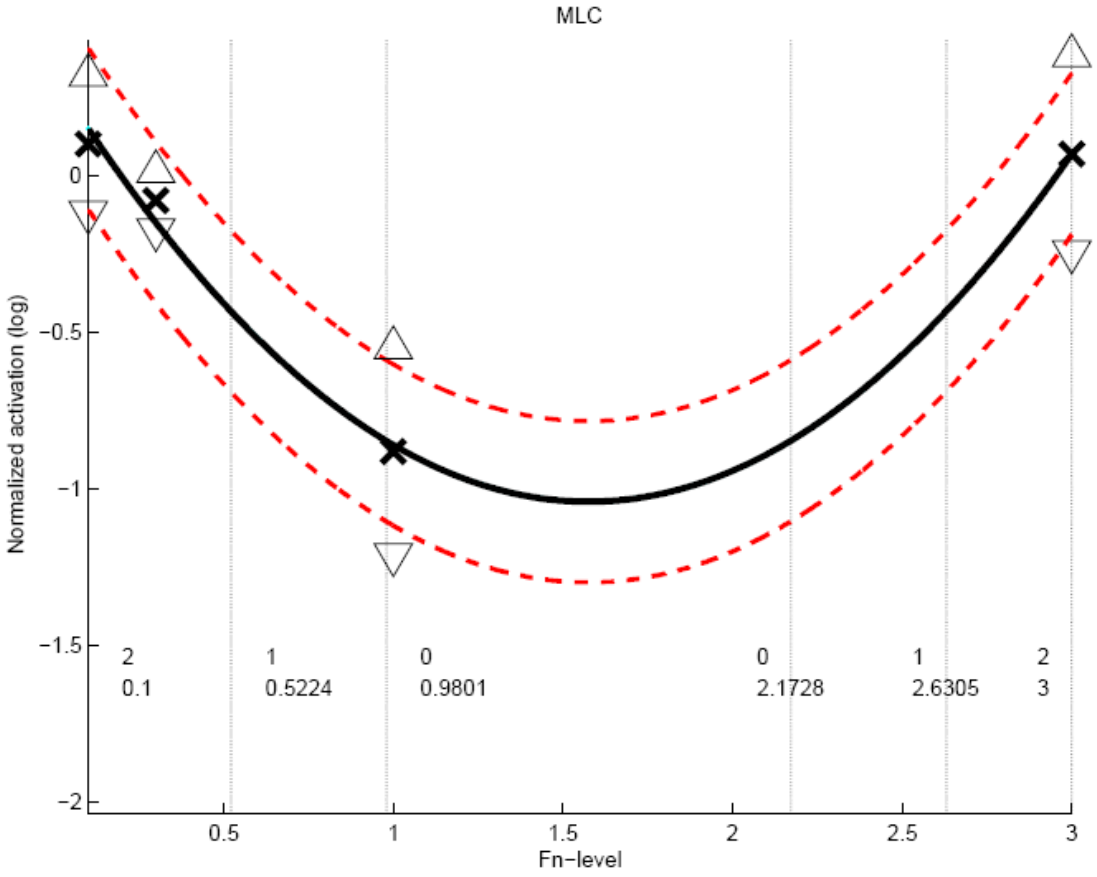


Figure 15. Fitted Polynomial interpolation data for activated MLC from the 1 hour EGF treatment data set. Crosses are actual measurements, upper and lower triangles are error estimates red line denotes the range of distribution of the interpolated values using fitted polynomial across a range of fibronectin coating concentrations. pMLC is discretized as low (0), medium (1) or highly activated (2) along FN concentrations. Thus, high pMLC is seen with EGF treatment if FN is less than 0.5224 and more than 2.6305 μg/ml.

3.3.3 Decision tree analysis of signaling proteins predicts a critical role of MLC based cell contractility in mediating cell migratory responses

Rarely have complex cellular behaviors been studied from a 'systems biology' perspective. A complex and well orchestrated cellular response like cell migration can only manifest from optimal quantitative activation of tens and hundreds of signaling proteins. Thus, an important question that is to be asked when altering cell motility in therapy is 'what are the relative quantitative contributions of each of the hundreds of signaling proteins towards such a biological response?' Or specifically 'how much of ERK and MLC need to be activated for cells to exhibit maximum motility?' By answering some of these questions, we can begin identify such crucial switches needed to be abrogated or amplified depending upon the desired cell response. This is the problem that we have approached using decision tree analysis.

The construction of decision trees has been covered in detail is sections 2.3.4 and 2.4.2. We conducted decision trees using the above five key signaling proteins activated by different time periods of EGF treatment. The decision trees obtained from three different EGF treatments (5 minute, 1 hour and 16 hour measurements) yielded different classification efficiencies of observations from the training set. The 5 minute decision tree predicted 70 % of observations from the training data set (figure 12), where as the 1 hour decision tree had a predictive power of 76 % (figure 13). The 16 hour decision tree could only predict 54 % of the observations from the training data set and hence was eliminated due to the questions of its applicability in making biologically valid predictions.

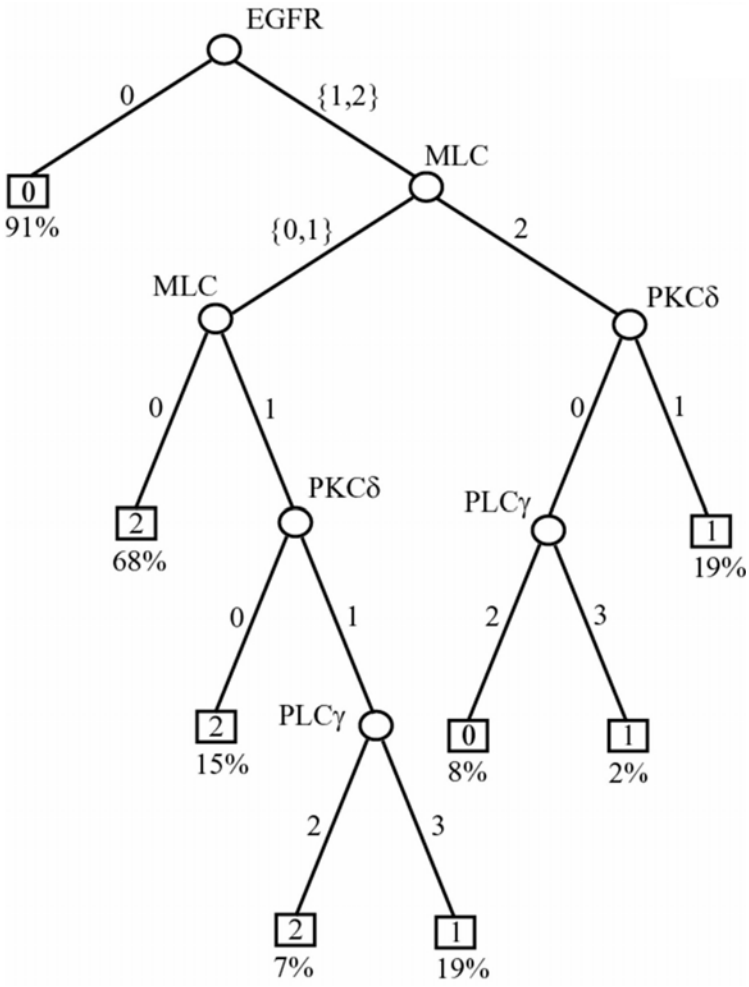


Figure 16. Decision tree of signaling proteins for cell speed using 1 hour of EGF stimulus data.
Round nodes denote variables whereas square nodes denote migration speed categories. Integers attached to the arc correspond to the split of the parent nodes. Under each migration speed category the fraction of cases explained by that classification rule is given. For example, if EGFR is high = 1 or 2 and MLC is low or 0, the migration speed category is 2 (high) and 68 % of the observations (in the training set) for the migration speed category 2 can be explained by this rule. Compare with 5 min decision tree in figure 11.

3.3.3.1 Cell contractility mediated by myosin light chain filament activation is central to motility

Contractile force production is enabled through the actin-myosin coupling upon activation of regulatory myosin light chains [34, 42]. That cell contractility is required for cell movement, was described by Iwabu et al [34]. Interventions that totally disrupted PKCδ mediated MLC

activation abrogated cell motility by decreasing cell contractility. Each of the molecules selected for modeling (like PLC, ERK) governed one or more individual biophysical events during motility (described in section 1.2). As such, our decision tree analysis was useful in predicting which molecules, and therefore which of the biophysical processes they controlled, were hierarchically important in governing motility. Interestingly, after EGFR activation, MLC mediated contractiliy was the most crucial ingredient. The decision tree discretized MLC activation into low (0), medium (1) and high (2). According to the predictions, the cells move with highest speed, when after EGFR activation, MLC phosphorylation is low (Figure 16) and 68 % of cells that move with high speed can be explained with this rule alone. Thus, lowering MLC mediated contractility can actually increase cell speed whereas total MLC inhibition can abrogate cell motility. While the effects of total MLC inhibition on cell motility have been intuitive and published by Iwabu et al, the biphasic nature of cell migration (speed) upon subtotal inhibition is non-intuitive and novel. Thus, our decision tree model predicted MLC to be an important regulatory node governing cell motility and predicted that subtotal versus total abrogation of this node can have drastically opposite cell responses.

3.3.3.2 Subtotal inhibition of MLC activation increases cell speed

Our model predicted that subtotal lowering of MLC activation increases cell speed. This was particularly relevant to the conditions where surface FN concentrations were below 0.522 μg/ml or greater than 2.6 μg/ml as derived from figure 14 and 15. Both of these conditions are associated with high phosphorylated MLC levels with minimal cell speed and there is apparent dsyregulation in the balance between the substratum adhesion strength versus contractility; i.e. despite high MLC activation in both conditions, there is too little substratum adhesion at 0.1 μg/ml while it is in excess at 3 μg/ml [31]. Thus at 0.1 μg/ml, contractility supersedes adhesion strength whereas this phenomenon is reversed at the condition of 3 μg/ml of surface fibronectin.

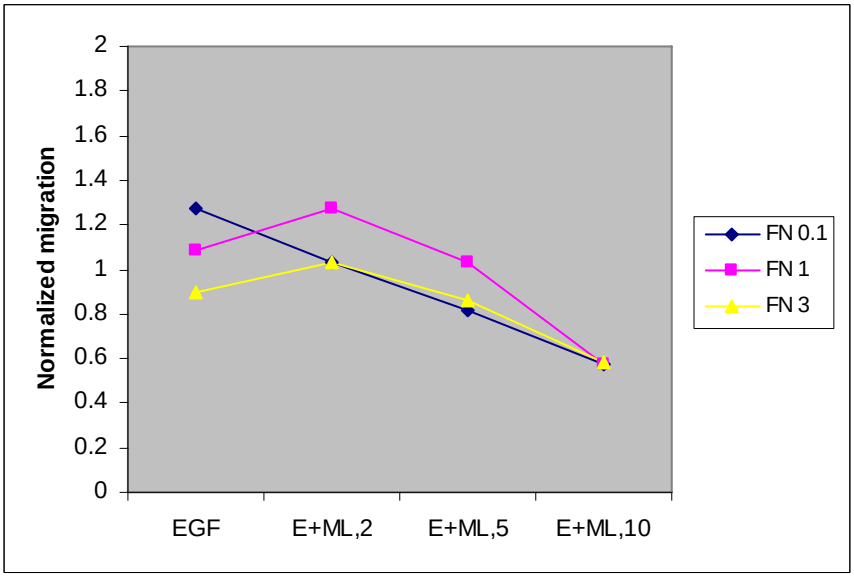
To test the model predictions under such conditions, we employed a well described and validated MLCK inhibitor ML-7 to measure fibroblast migration speed under the same extracellular conditions (4 FN concentrations - / + EGF). Such a downstream inhibitor was chosen (over PKCδ inhibitor Rottlerin) because it is MLCKinase specific and hence the resultant cellular responses can be attributed directly and specifically to MLC inhibition since PKCδ is

involved in diverse cellular responses other than motility [146]. This is likely reflected in the Decision tree analysis wherein MLC lies hierarchically above PKCδ.

We initially measured cell migration on fibronectin using the 'scratch assay' under a range of ML-7 concentrations within the culture medium containing saturating levels of EGF. In parallel, Western blot analysis of activated MLC (with EGF treatment) showed a linear decrease in phosphorylated MLC levels with increasing ML-7 concentration (not shown). Under the same conditions, lower ML-7 concentration (2-3 μM) increased fibroblast migration relative to EGF alone (Figure 17). Specifically, within fibronectin concentration of 1 and 3 μg/ml, subtotal inhibition of MLC increased cell migration; i.e. migration distance was greater in conditions that had EGF and ML-7 within the medium relative to EGF alone (Figure 17 A). In the condition with FN of 0.1 μg/ml, any further lowering of MLC activation did not increase cell migration but rather reduced it. Total inhibition of MLC using ML-7 (at a range of 10-15 μM) completely abrogated motility. These initial experiments were in accordance with the predictions from our decision tree model.

To apply these predictions specifically to individual biophysical events during motility, we measured fibroblast migration speed using single cell tracking under exactly same experimental conditions. Speed was measured as the distance traveled by an individual cell over a given period of time (10 hours) [31]. We found that lower ML-7 concentrations increased cell migration distance as well as speed relative to EGF alone (experiments in progress). Again, motility was totally abrogated with complete MLCKinase inhibition. Thus, the biphasic response of cell speed was shown to be dependent upon MLC mediated contractility. Our decision tree model predicted this phenomenon whereas our experimental conditions validated the model predictions. These observations are preliminary and the experiments are still in progress to compile the quantitative data.

A)



B)

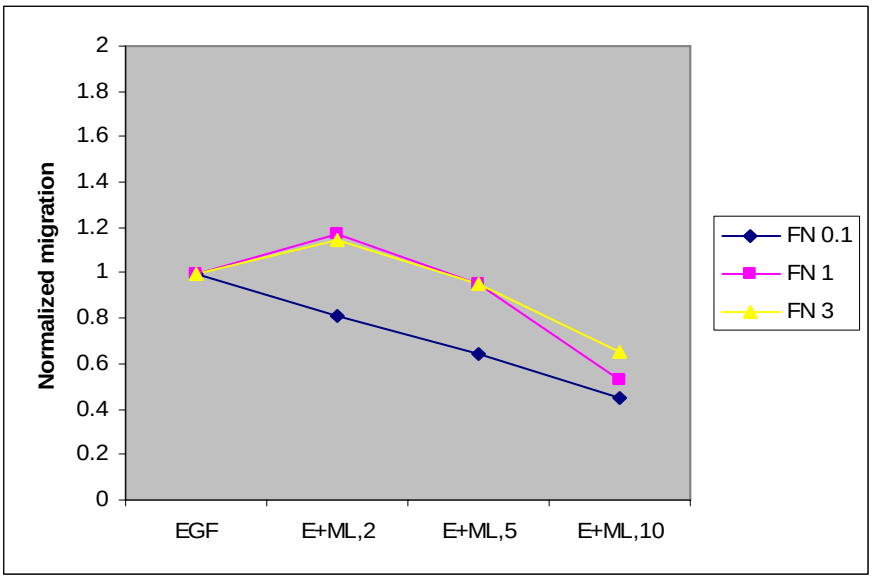


Figure 17. Subtotal inhibition of myosin light kinase increases cell migration.
NR6WT fibroblasts were grown on surfaces coated with three different fibronectin concentrations (0.1, 1, 3µg/ml) and quiesced in serum restricted conditions for 24 hours. The cells were scraped using a pipet tip and the migration of cells into the denuded area was assessed over a period of 24 hours. EGF (E above) and ML-7 (ML above) was added throughout the experimental conditions. Lower doses of ML-7 (2-5 µM) were shown to increase cell migration as compared to EGF treatment alone. A. Migration values normalized to the first condition, FN of 0.1µg/ml without EGF. B. Migration values normalized to EGF treatment within each FN condition.

3.3.3.3 Subtotal inhibition of myosin light chain activity increases migration of cancer cells

To assess if our predictions could be employed to different cells under a range of experimental condtions, we utilized MDA-MB-231 breast cancer cell lines and assess their migratory response across a range of MLCKinase (and hence MLC) inhibition. These cells overexpress EGF receptor and actively exhibit autocrine stimulatory loops that drive their migration and invasiveness [147]. Motility was assessed using a transwell Boyden chamber with EGF as a chemotactic cue in the bottom wells and ML-7 within the seeding medium. In accordance with the findings in fibroblasts, migration of MDA-MB-231 cells was substantially higher when the medium contained low to medium concentration (3-10 μM) of ML-7 in addition to EGF as compared to EGF alone (Figure 18). The term 'low' or 'medium' in regards to ML-7 concentration is obtained by titration of activated MLC levels and varies with cell types; i.e. for NR6WT cells, 10 μM of ML-7 is high whereas the same is 'medium' for MDA-MB-231 cells. In other words, the amount of MLC downregulation that is achieved by 10 μM in NR6WT cells is approximately similar to that achieved by 15 μM in MDA-MB-231 cells. Transmigration was completely blocked when ML-7 concentration completely abrogated MLC activity (>30 μM).

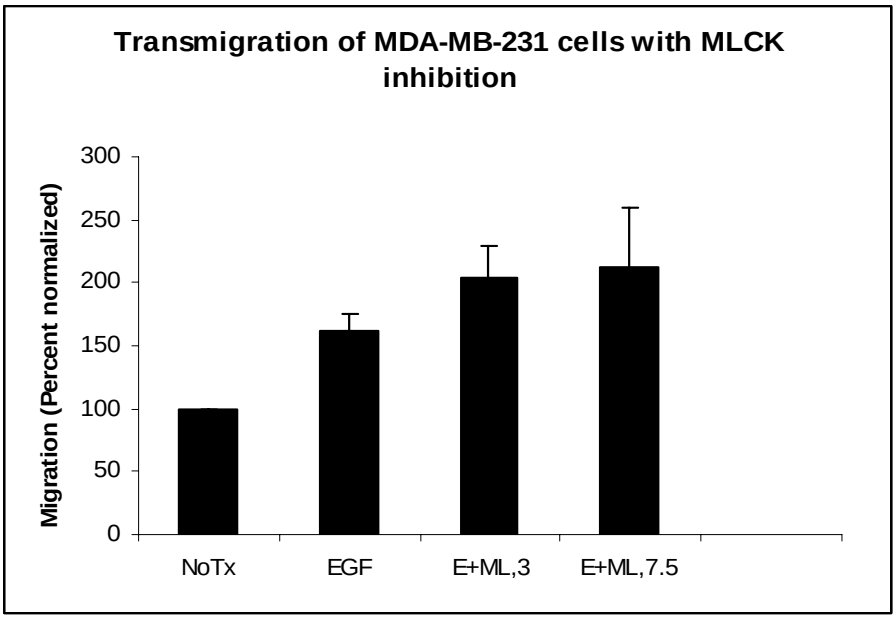


Figure 18. Subtotal inhibition of MLC increases migration of MDA-MB-231breast cancer cells.

20,000 cells were subjected to transmigration across a boyden chamber in serum restricted conditions for 24 hours. EGF(E) and increasing concentration of MLCK inhibitor, ML-7, (ML above) were added to the medium where indicated. The migrated cells were fixed in formaldehyde, stained and manually counted.

3.4 DISCUSSION AND CONCLUSIONS

Most disease states like cancer, are a result of perturbations within multiple signal transduction pathways rather than single genetic mutations commonly underlying diseases like cystic fibrosis. These signaling pathways comprise of nodes act as signal amplifiers, transmitters or distributors to different signaling proteins within the network. Thus, numerous and usually, proteins with redundant activity profiles govern such complex cellular phenomena. Evidently, altering cell behaviors is difficult without a thorough understanding of how these signaling switches work synergistically. While enormous data sets are available for biological conditions, such data sets have not been integrated to provide information about the interlinked and branched signaling networks. Therefore, targeted therapies often fail because cells utilize parallel and alternative pathways to mediate the necessary biological functions. Identification and modulation of key signaling nexi from such complex networks can alter cell behaviors and yield favorable responses [148, 149].

We utilized decision tree analysis to identify the crucial effectors of cell motility depending upon a set of extracellular cues. Fibronectin was selected since NR6WT fibroblast express alpha5beta1 integrin receptors that are actively involved in cell signaling during motility. Also, these being adhesion receptors provide a counter-balance against the motogenic EGF receptor that is overexpressed in these cell lines. Such adhesion versus motility balance is present *in vivo* environments, where motility of cells is dictated by the cellular ecology, cell-substratum and cell-cell adhesion profile, extracellular matrix components along with a spectrum of soluble and matrix-embedded extracellular stimuli [1, 99]. Our model was based on the quantitative measurements of five signaling proteins that are activated downstream of the EGFR and are known to mediate key biophysical events of motility. Arguably, such model could suffer from predictive power due to the possible exclusion of other key signaling proteins (like FAK, calpain, etc). However, our model predicted 76 % observations within the training data set accurately. Future experiments are aimed at incorporating other key signaling proteins within this

foundational decision tree model. The model will then be similarly tested for predicting cellular biophysical events with any discrepancies be refined and the model accordingly 'fine-tuned' to maximize its classification efficiency.

Our decision tree model clearly identified MLC mediated contractility as the most crucial biophysical event during EGF induced motility. This does not mean however, that disrupting other cellular events, like PLCγ based lamellipodal protrusion, will not abrogate motility. The utility of decision trees is to predict the switches that upon disruption, can produce the 'most significant' response. It also identifies molecules that need to be inhibited together to alter motility. In our model that was based on 5 minute and 1 hour EGF stimulation data set (Figures 12 and 13), contribution by ERK was masked by similar activation profile observed with EGFR. This does not mean that ERK is not vital in motility since disrupting ERK reduces migration [79] but rather means that ERK activation profile was captured by measuring EGFR activation. Further, the model predicted in accordance with Glading et al [79] that motility requires functional ERK activation since 90% of cells that migrated could be explained to operate using this rule alone (Figure – 11, 5 minute decision tree). Furthermore, even the 5 minute data set predicted 70 % observations within the training set correctly although maximum motility is observed atleast 4-8 hours after EGF addition. This was due the fact that 5 minute data measurements captured activation trends of important molecules like ERK that are indispensable for cell migration but are usually attenuated at 1-2 hours after EGF stimulus when motility has started becoming a stable biophysical response. Such transient activation is sufficient to elicit motility since ERK transmits the signal downstream towards the final effectors of motility before attenuation. Additionally, the model indicates that ERK functions like an 'on-off' switch during motility: if ERK (and/or the EGFR) is active, the cells will move depending upon the profile of other signaling proteins but if ERK is inactive, the motility is practically negligible since 90 % of cells with minimal motility could be predicted by this rule (Figure 11). This also points to a new proposition: targeting MLC and ERK together to retard cell migration.

Our model, non-intuitively predicted that lowering MLC activation, but not totally abrogating it, can paradoxically increase cell speed. These predictions held true, atleast in part, in our population based 'scratch assay' that assessed cell migration under different concentrations of MLCKinase inhibitor, ML-7. Of interest, subtotal inhibition of MLCKinase under higher fibronectin concentration of substratum increased cell speed (Figure 17) whereas under lower

substratum adhesive conditions (0.1 μg/ml), further reduced it. Motility is a function of optimum balance between cell-substratum adhesion versus cell contractility that enables cells to break some cell-substratum adhesions but form newer ones as the cell moves [30]. This is evident at intermediate FN concentration of surfaces in our experiments [31]. The adhesion-contractility balance is impaired at the two extreme conditions where too little adhesion precludes a cell from generating sufficient adhesions for locomotion; hence further lowering of contractility even by subtotal inhibition of MLC further reduces motility (figure 15). On the other hand, too much surface adhesiveness (FN of 3 μg/ml) maintains a cell in a 'stuck' situation due its inability to detach. This is because higher surface fibronectin promotes excessive integrin receptor engagement evenly on the surface rather than keeping it selective at focal adhesions. Cell-substratum adhesiveness is a result of ligand concentration, receptor number or ligand-receptor affinity, with maximum motility (and cell speed) occurring at intermediate level of cell-substratum adhesion strength [30]. Thus a higher FN concentration results in a cell stuck to the surface with a high intrinsic contractile force. In such situations, any decrease in contractility can be predicted to increase cell motility by reinstating the adhesion versus contractility balance and enabling cell detachment, breakage of focal adhesions with formation of new ones. This was indeed confirmed by our initial experiments (Figure 17) using 'scratch assay'and observed in single cell tracking experiments.

Using the popular 'scratch assay', we found that cell motility was highest when cells migrated on 0.1 μg/ml of FN under EGF influence (figure 17 A). This is in contrast to the data from Maheshwari et al where motility was least at the two extreme fibronectin concentrations ; i.e. at 0.1 and 3 μg/ml. We attribute this discrepancy to the nature of the 'population' based scratch assay we used to assess total cell migratory distance versus single cell tracking that Maheshwari et al employed to measure cell speed. In addition, while performing the 'scratch' on the confluent monolayer, it is possible to scrape the layer of fibronectin coated for the precise experimental situation. Thus, 'scratch assay' can only be utilized to test preliminary predictions from the model. Our initial experiments to assess cell speed using single cell tracking under the above experimental conditions have confirmed findings from Maheshwari et al. In addition we have unequivocally found that subtotal reduction of MLC activation can indeed increase speed of cell migration. These experiments are in progress.

Changes in the extracellular matrix components alter cell signaling during many patho-physiological states. Fibronectin has been shown to be deposited along the wound beds during wound contraction, and enables the recruitment of fibroblasts into the denuded area [150]. Once there, fibroblasts contract and approximate the wound edges enabling closure of the wound edges {Allen, 2002 #343}. During these states, the lack of motogenic signal like EGF drives signaling through the integrins and transmits contractile force back to the matrix rather than used for motility [105]. In our experiments, a range of these condition were reproduced (although simplified) by coating the surface with linear ranges of fibronectin concentration (0.1 to 3 μg/ml). Thus, lowering MLC activation, under such conditions, can enable cell motility by 'redirecting' such contractile force for locomotion as well as lowering cell-adhesiveness enabling cellular detachment from the substratum [151]. Thus, MLC based contractility is identified as the fulcrum which if tipped in one direction can promote motility over adhesion.

These findings have profound implications for therapy. Identifying key nodes enables quantitative manipulations using pharmacologic methods for specifically desired cellular responses. It also points to the importance of how these signaling proteins are regulated stoichiometrically. Our predictions held true even when applied to breast cancer cells, where lower doses of ML-7 promoted cell migration. While a complete abrogation of MLC can be beneficial in limiting tumor cell motility, partial inhibition using lower pharmacological doses can increase tumor cell motility and invasion leading to devastating consequences. This further points to the importance of applying newer modeling approaches to fully characterize the role of signaling cascades in mediating cellular behaviors. Such understanding will enable precise therapeutic targeting of key signaling nodes and open the door to individualized 'patient-tailored therapy' [148].

4.0 ABROGATION OF PROTEIN KINASE Cδ MEDIATED TRANSCELLULAR CONTRACTILITY LIMITS MIGRATION AND INVASIVENESS OF PROSTATE CANCER CELLS

Sourabh Kharait[1] , Rajiv Dhir[1], Douglas Lauffenburger[2] and Alan Wells[1#]

[1]Department of Pathology, University of Pittsburgh School of Medicine and Pittsburgh VAMC, Pittsburgh, PA-15261

[2]Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA-02139.

[#]Corresponding author:

Alan Wells, MD DMSc
S713 Scaife Hall,
Department of Pathology
University of Pittsburgh
3550 Terrace street
Pittsburgh, PA -15261.
Ph: 412-647-7813
Fax: 412-647-8567
wellsa@upmc.edu

Accepted for publication, March 2006. *Biochemical and Biophysical Research Communications.* Format and contents are modified for dissertation.

4.1 ABSTRACT

Tumor progression to the invasive phenotype occurs secondary to upregulated signaling from growth factor receptors including that for the epidermal growth factor (EGF) family of ligands. The resultant aberrant signaling drives the key cellular responses of proliferation, migration and invasion. Recently, Protein Kinase Cδ (PKCδ) was shown to be involved in EGFR-mediated fibroblast contractility and motility. We hypothesized that PKCδ-mediated transcellular contractility is required for migration of prostate tumor cells and can be explored as a possible target for therapeutic intervention. Two invasive prostate tumor cell lines, DU145 cells overexpressing wildtype human EGFR (referred to as DU145WT) and PC3 cells, were studied. PKCδ is overexpressed in these prostate carcinoma cells relative to normal prostate epithelial cells, and is activated downstream of EGFR leading to cell motility via modulation of myosin light chain (MLC) activity. EGFR-mediated activation of PKCδ and its key target substrates was reduced by pharmacological (Rottlerin) or molecular (siRNA) abrogation in both the DU145WT and PC3 cell lines, and these interventions significantly decreased EGF-induced migration and invasion of both cell lines *in vitro*. As an initial exam of the human situation, PKCδ protein was greater in two human prostate cancer tissue lysates as compared to normal donor prostate. Immunohistochemical analysis of active (phosphorylated) PKCδ in prostate tissue sections showed robust PKCδ activity in prostate cancer tissue as compared to almost undetectable levels in normal donor prostate tissue. In prostate carcinomas, activated PKCδ was observed both at the cytoplasmic and nuclear locale. Thus, we conclude that PKCδ is a critical governor of prostate cancer cell migration and that its abrogation can limit prostate tumor invasiveness.

Key Words: Myosin light chain, cell contractility, prostate carcinoma, migration, invasion.

4.2 INTRODUCTION

Prostate cancer is the most commonly diagnosed cancer in men. The morbidity and mortality associated with prostate cancer results from both invasiveness through the capsule into surrounding tissues and formation of distant metastases [61]. Tumor progression is achieved by genetic as well as epigenetic changes including amplification and upregulated signaling through cell surface receptors, particularly the epidermal growth factor receptor (EGFR) [3, 58]. EGFR is overexpressed in a variety of solid tumors including those of the breast, prostate, brain (glioblastoma multiforme), bladder, and lung [52-57], with this increased signaling correlating with tumor progression to invasion and metastasis [58]. As tumor cells also secrete the autocrine activating ligands such as TGF-α [59], these autocrine activating loops provide a rationale for targeting this signaling axis [152, 153]. While such agents have reached accepted clinical use, their efficacy has been limited by systemic toxicity and the ability of other receptors to supply the tumor progression signals. Thus, selectively targeting key intracellular molecular switches for the required cell behaviors would be one avenue to augment our armamentarium aimed at preventing tumor invasion and metastasis.

Increased cell motility facilitates, at least in part, tumor progression [154]. In many tumors, autocrine signaling through upregulated growth factors impels this [58]. Targeting EGFR-induced motility, however, needs a thorough understanding of this highly orchestrated process. One well-characterized approach is to break down the overall cellular migratory response into discrete biophysical processes. Principally, migration of a polarized cell results from protrusion of lamellipods at the front end, attachment at or near the leading edge, transcellular contractility, and finally detachment of the cell membrane at its rear end to enable productive locomotion [6]. In recent years, key molecular switches have been identified for each of these steps. Previous experiments have shown that targeting specific signaling pathways that control discrete cellular biophysical events needed for migration can be effective in abrogating motility and invasion of prostate tumor cells[14-16, 155]. However, while each step is required for motility, it is not known how they interact to regulate motility. Recently, we have used a modeling approach to define the information content hierarchy of these switches, and found that

transcellular contractility is central to productive locomotion [11]. As the actomyosin contraction machinery is central to many cellular processes required during homeostasis, we sought the key steps between it and EGFR, and to the recently identified protein kinase Cδ (PKCδ) switch [34].

The protein kinase C (PKC) enzyme family comprises at least 11 members classified into three groups: the classical, the novel and the atypical, depending upon their requirement for calcium and diacylglycerol (DAG) for activation. PKCδ, almost ubiquitously expressed, belongs to the novel group, and is independent of calcium for its activity, but it still can be activated by DAG, phospholipids and phorbol esters [146, 156]. PKCδ mediates diverse cellular functions in normal as well as cancer cells and is in majority of human tissues [146, 156]. In addition to its documented role in cancer cell survival and acquisition of chemotherapeutic resistance [146, 157], PKCδ mediates migration of a variety of cells, including carcinoma cells [34, 158]. Studies have shown PKCδ to be involved in sustained migration of EGFR-overexpressing breast cancer cells [158]. Histopathologic studies on breast cancer tissue specimens have also shown marked overexpression of PKCδ in aggressive tumors [159]. Another study in prostate tumor specimens correlated multiple alterations in expression patterns of PKC isoenzymes with different grades of tumor invasiveness, though they did not highlight a role for PKCδ [160]. Interventions that down regulated PKCδ in invasive breast cancer cells, substantially reduced their ability to invade *in vivo*, predominantly by reducing their motility [161]. Taken together, its role in modulating cellular migratory properties combined with the observation that it is overexpressed in certain tumors, point towards a possible role of PKCδ in promoting tumor progression. However, how this key effector is regulated and where it fits within tumor progression networks is not known.

As PKCδ signaling contributes to the transcellular contractility required during growth factor induced motility [34], we asked if PKCδ abrogation could reduce invasiveness of prostate tumor cells as such invasiveness has been shown to occur downstream of growth factor receptor signaling, that being primarily from EGFR in prostate carcinomas [58]. Using two different invasive prostate cancer cell lines, we showed that inhibiting PKCδ pharmacologically and molecularly reduced migration and invasion of prostate cancer cells downstream of EGFR signaling. Additionally, activated, as well as total PKCδ levels are higher in prostate carcinomas as compared to normal prostate tissue as assessed by immunohistochemistry and immunoblot

analysis. Thus we propose that PKCδ is a crucial mediator of invasiveness of prostate carcinoma cells and a potential therapeutic target in limiting prostate tumor invasion and metastasis. Such molecule-specific inhibition of the morbid and mortal prostate cancer invasion and metastasis could represent an important new approach because the slow growth of prostate carcinomas limits the efficacy of cell cycle-directed agents.

4.3 MATERIALS AND METHODS

4.3.1 Cell lines

DU145 (also referred to as DU145 parental or DU145P) cells represent an invasive prostate cancer cell line. We have overexpressed EGF receptor (EGFR) in these cells, thereby refered to as DU145WT [60, 162]. These cells exhibit increased motility and invasiveness as compared to the parental DU145 as previously shown [60, 162]. DU145 cells were grown in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum, and 1% supplement of each of the following: MEM non-essential amino acids, sodium pyruvate, penicillin/streptomycin and L-glutamine (all from GIBCO, Gaithersburg, MD). The DU145WT cells were grown in selection medium (for EGFR) containing a final concentration of 350 μg/ml of G418. Cells were quiesced in serum free medium for a period of 24 hours for studies with EGF stimulation.

To confirm our findings pertaining to invasive prostate tumor cell lines, we tested a second invasive prostate tumor cell line, PC3, that is derived from a metastatic focus in the bone [163]. Cells were grown in F12K medium (GIBCO) containing 10% FBS and 1% of supplement as described above. Cells were quiesced in medium containing 0.5% dialysed FBS for a period of 24 hours for studies with EGF stimulation.

Normal prostate epithelial cells were obtained from Cambrex Inc. (Walkersville, MD) and grown as per the manufacturer's instructions.

4.3.2 siRNA constructs and transfections

Two specific siRNA duplexes to PKCδ (GenBank^{TM} accession number NM_006254) were designed and purified by Integrated DNA Technologies (IDT) (Coralville, IA). The first (PKCδ1) siRNA1 sequence was 5'– GGUCCUGGGCAAAGGCAGCTT – 3' whereas the second (PKCδ2) siRNA2 sequence was 5'- GGACAUCCUGGAGAAGCUCTT - 3'. The siRNA

sequence to green fluorescent protein (GFPsiRNA) was 5'- GACCCGCGCCGAGGUGAAGTT – 3' and was used as a negative control. For a second control, we constructed a mutant siRNA to PKCδ by changing C to G at 4^{th} and G to C at the 16^{th} nucleotide positions within the siRNA2 sequence to PKCδ as described above. All siRNA transfections were performed using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) as per the user instructions. Cells were plated in growth medium without any antibiotics in a 6 well plate and were allowed to reach ~ 80 % confluence on the day of transfections. Briefly, 4 μl of 20 μM siRNA (80 picomoles) was diluted in 200 μl of OPTI-MEM. 4 μl of Lipofectamine 2000 was diluted in 200 μl of OPTI-MEM and incubated at room temperature (RT) for 5 minutes. The diluted siRNA was combined with the diluted Lipofectamine and the sample incubated for another 20 minutes at RT. 400 μl of siRNA-Lipofectamine complexes were then added to the growth medium and the cells incubated for 24 hours at 37^{o} C in a tissue culture incubator. Medium was then changed to complete growth medium and cells were incubated for another 24 hours. Cells were lysed and lysates resolved using SDS-PAGE to analyze the knockdown of PKCδ protein levels using immunoblotting.

4.3.3 Reagents and antibodies

Polyclonal antibodies to PKCδ (catalog number sc-937), PKCα and phosphorylated MARCKS were purchased from Santa Cruz Biotechnology (SantaCruz, CA). Antibodies against phosphorylated (thr505) PKCδ (catalog number 9374), phosphorylated PKCα/βII, phosphorylated (serine) PKC substrates, phosphorylated (ser19/thr18) MLC (catalog number 3674) and phosphorylated ERK/MAPKinase were purchased from Cell Signaling Technology (Beverly, MA). Monoclonal antibody to MLC was obtained from Sigma (St. Louis, MO). Rottlerin, a specific PKCδ inhibitor and PD153035, a specific EGF receptor tyrosine kinase inhibitor, was purchased from Calbiochem (San Diego, CA). *In vitro* invasion of cells was assessed using Matrigel coated Invasion chambers obtained from BD Biosciences (Discovery Cellware) (Bedford, MA).

4.3.4 Immunoblotting

Upon indicated treatment conditions, cells were scraped and lysed in preheated (SDS) sample buffer as described above and boiled for 5 minutes. The samples were then cooled at room temperature, resolved by SDS-polyacrylamide gel electrophoresis (gels ranging from 7.5 to 15 %) and transferred onto PVDF membranes. The membranes were blocked for 1 hour using 5 % non-fat milk and incubated with respective primary antibodies for another 2 hours at room temperature. The membranes were washed using Tris buffered saline containing Tween 20 for 30 minutes and incubated with secondary antibodies linked to horse radish peroxidase for 1 hour at room temperature. The proteins were detected using enhanced chemiluminescence method using standard protocols (Pierce, Rockford, IL).

4.3.5 *In vitro* two-dimensional motility

Cell migration was measured as the distance traveled by the cells into an acellular area [83]. Cells were seeded in 6 - well tissue culture plates for a period of 24 hours in growth medium. Cells were quiesced for another 24 hours in serum free medium at which time cells formed a confluent monolayer. A denuded area was created by scraping with a pipet tip, washed three times with phosphate buffered saline (PBS) to remove dead cells, and kept under serum free conditions throughout the experiment. EGF at 10 nM (and inhibitors or diluent as indicated) was added to the serum free medium. Cells were then photographed using an inverted microscope immediately following scraping (0 hour condition) and 24 hours later (24 hour condition) in exactly same three different areas. The photographs were merged and analyzed using Adobe photoshop program to determine the average distance traveled by the cells in 24 hours. All experiments were performed in triplicate and repeated at least three times.

4.3.6 *In vitro* transmigration assay

To assess the migratory potential of transfected cells, we used the modified Boyden (transwell) chambers available from BD Biosciences (Bedford, MA). Transmigration was assessed by the ability of cells to migrate through a porous (8 micron) PET membrane towards a chemotactic cue. Inserts were placed in corresponding wells of a 24 well plate. The bottom wells contained complete growth medium with 10 % FBS. Equal number (2.5×10^4) of cells were plated in the cell culture inserts in serum free medium containing 1 % BSA. After, 24 hours the medium was changed to serum free medium and the cells were allowed to transmigrate through the porous membrane filter for another 24 hours. The un-migrated cells at the top of the membrane filter were removed by a cotton swab while the cells at the bottom of the filter were fixed in 2 % formaldehyde, stained using Diff Quik stain (Allegiance, McGraw Park, IL) and manually counted in three different fields. All experiments were performed in triplicate and repeated at least three times.

4.3.7 *In vitro* Matrigel invasion assay

Invasive potential of cells was determined by their ability to invade through a multi-cell thick layer (~100 μM) of extracellular matrix (Matrigel). The invasion assay has been previously described [14]. Briefly 25,000 cells were plated in each of the invasion chambers on top of the matrigel layer. The bottom chambers contained complete growth medium with 10% FBS. Cells were initially plated in serum free medium containing 1% BSA for the first 24 hours. The medium was changed to serum free medium for the next 24 hours of the experiment. Where indicated, EGF, 10nM was added both to the top insert of the chamber containing cells as well as the bottom chamber. After 48 hours, the chambers were removed from the bottom well plates and the non-invasive cells on top of the membrane filter were removed with a cotton swab. Invaded cells at the bottom of the chambers were fixed in 2% formalin for 30 minutes and stained with Diff Quik staining kit according to standard staining protocols. The cells within each membrane filter were manually counted under a light microscope in three different fields, and each experiment was similarly performed in triplicate and repeated at least three times.

4.3.8 Cell proliferation assay

To assess the effects of Rottlerin and/or siRNA transfections on cell growth, we directly counted cells using a Coulter counter. Briefly, cells were grown in complete growth medium in 12 well plates for 24 hours. The growth medium was replaced by serum free medium containing 10 μM of Rottlerin and cells incubated for another 24 hours. Before counting, the cells were washed twice with PBS to remove the dead cells, trypsinized and resuspended in 1 ml of complete medium. The cells were then centrifuged at 4 degrees C at 1000 rpm for 5 minutes and the cell pellet was resuspended in 1 ml of fresh growth medium. The cells were then counted using a standardized coulter counter method. Counts from Rottlerin treated cells were compared to DMSO treated cells which served as vehicle control. Similar protocol was used to assess proliferation of cells transfected with specific siRNA. In the siRNA experiments, Day 1 represented the day of transfection where as Day 3 and Day 5 represented 48 and 96 hours post-transfection. At these time points, cells were trypsinized and counted using the automated coulter counter. Each experiment was performed in triplicate and three times.

4.3.9 Prostate tissue lysates

Frozen prostate tissue was obtained from the University of Pittsburgh Health Sciences tissue bank according to the institutional review board (IRB) guidelines under the supervision of a genitourinary pathologist. Two samples of prostate cancer tissue with a gleason score of 7, and two samples of normal donor prostate tissue without any evidence of prostate disease, matched for age (55-65 years), were assessed for the expression level of PKCδ. Briefly, 100 mg of prostate tissue was homogenized using a hand held homogenizer on ice in 1 ml of a buffer containing 50mM HEPES, pH 7.4, 150 mM of NaCl, 1mM Na - Vanadate, 1 % Triton-X and 10 % glycerol. Protease inhibitors (10 μg/ml of leupeptin, 100 μg/ml of aprotinin and 1 mM of AEBSF) were added to the buffer just before cell lysis. The lysates were incubated on ice for 20 minutes, then centrifuged at 4^0 C for 20 minutes at 11,000 rpm. The supernatent was collected in a fresh eppendorf tube, and the lysates frozen at - 20^0 C until used. For immunblotting, the

lysates were thawed on ice, and the protein concentration in each sample estimated using BioRad Protein assay. The samples containing equal amount of protein were then mixed with equal volume of sodium dodecyl sulfate (or SDS) - sample buffer containing 0.1 M Tris-HCl, pH 6.8, 4 % SDS, 20 % glycerol, 0.2% bromophenol blue and 5 % β-mercaptoethanol (reducing agent) and immunoblotting carried out using the protocol described below.

4.3.10 Immunohistochemistry of prostate tissue specimens

Immunohistochemical staining to detect phosphorylated (activated) PKCδ (pPKCδ) was performed on paraffin embedded prostate tissue sections using a Dako Auto-Immunostainer. Five tissue sections of prostate carcinoma with Gleason scores of 7 or 8 on diagnosis and 5 normal donor prostate sections were analyzed using an antibody to pPKCδ. Briefly, the paraffin sections were initially deparaffinized and heat induced epitope retrieval was performed using a Biocare decloaking chamber. The sections were then quenched with 3% hydrogen peroxide, blocked with 10% normal goat serum and incubated with primary antibody (phospho-PKCδ, threonine 505) for 45 minutes followed by the rabbit envision + polymer. DAB was the chromogen used for the localization of the antigen. Counterstaining was performed with hematoxylin. The slides were then dehydrated and coverslips were applied. Photographs were taken using 40 X objectives and images processed using SPOT software. Reading was performed by a genito-urinary pathologist (R. D.).

4.4 RESULTS

4.4.1 PKCδ is overexpressed and active in prostate carcinoma cells

To investigate the role of PKCδ in prostate cancer, we first quantified the levels of PKCδ in prostate cancer cell lines. We compared the total PKCδ levels in normal prostate epithelial cells (PrE) to those in carcinoma cell lines (PC3, DU145WT and DU145 Parental) using immunoblot analysis of cell lysates. Our data indicate that PKCδ protein levels are significantly higher in these prostate cancer cell lines as compared to normal epithelial cells (Figure 19A).

To determine if PKCδ is activated by EGFR in prostate cancer cell lines, we stimulated DU145WT cells with 10 nM of EGF for 10 minutes and probed for active (phosphorylated) PKCδ protein using a specific phospho-(Threonine 505) PKCδ antibody (Figure 19B). Phosphorylation of threonine 505 residue can contribute to activation as well as stability of PKCδ since this residue lies within the activation loop domain of the enzyme [164]. EGF induced a boost in PKCδ activity that was inhibited by the specific EGFR tyrosine kinase inhibitor PD153035.

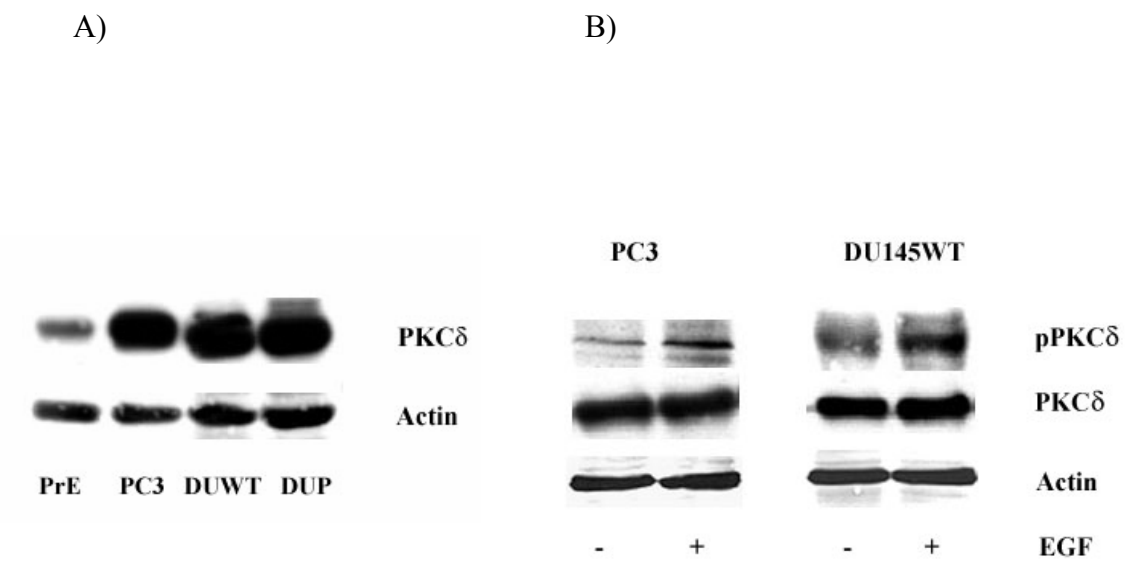


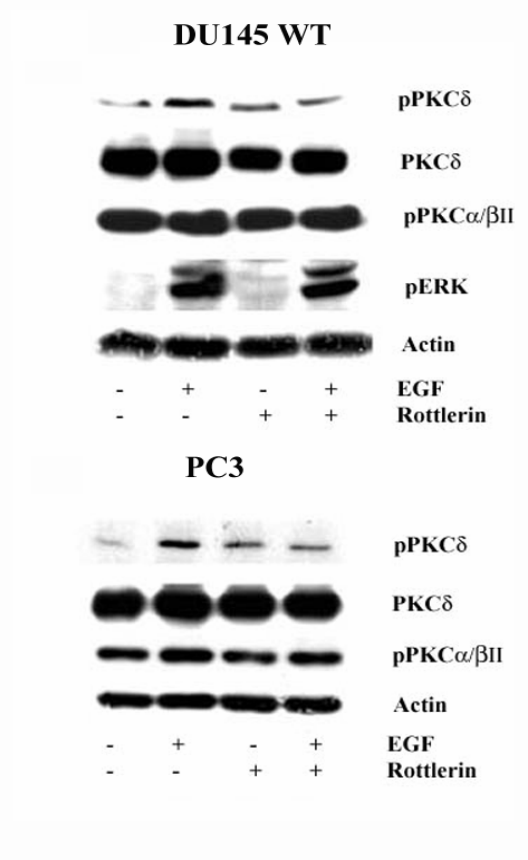
Figure 19. Expression and activation of PKCδ in prostate cancer cells.
A. Total cell lysates were prepared from different prostate cancer cell lines (PC3, DU145WT and DU145 Parentals) and analyzed using SDS - PAGE for expression levels of total PKCδ using western blotting. Normal prostate epithelial cells (PrE) were used as control to compare the PKCδ expression level. A representative immunoblot of three independent experiments is shown. B. PC3 and DU145WT cells were grown in complete growth media and quiesced in serum free medium for a period of 24 hours when they were 80% confluent. 10nM of EGF was added to the medium for a period of 30 minutes. The cells were lysed and samples resolved using SDS-PAGE. Activated status of PKCδ was probed using a specific phosphorylated (Thr505) PKCδ antibody.

4.4.2 EGF induced motility and invasion of prostate carcinoma cells is PKCδ-dependent

At doses up to 10 μM, Rottlerin has minimal effects on other PKC isoenzymes, thus selectively inhibiting PKCδ activity [158, 161, 165]. In PC3 and DU145WT cells, 7.5 μM Rottlerin significantly abrogated EGF induced phosphorylation of PKCδ (Figure 20A). As expected, the phosphorylation of other PKC isoenzymes (PKCα, βI/II) was unaffected. Also, phosphorylated ERK MAPKinase levels within the cells were unaffected at this dosage confirming that the decrease in PKCδ activity was not due to pan-inhibitory effects of Rottlerin. Additionally, Rottlerin decreased activated (phosphorylated) status of key PKC target sites including MARCKS and phospho- (serine) PKC substrates (Figure 20B). Our experiments reinforce previously published observations that Rottlerin can be used in prostate cancer cells to abrogate PKCδ activity [158].

Because pharmacological agents are typically not devoid of non-specific effects that cannot be completely assessed, we downregulated PKCδ protein levels using genetic (RNAi) interventions. siRNA to PKCδ decreased PKCδ levels by >70% as analyzed by immunoblotting of cell lysates (Figure 21). PKCα levels within the transfected cells were unaffected under the same conditions confirming the specificity of siRNA against the PKCδ isoenzyme.

A)



B)

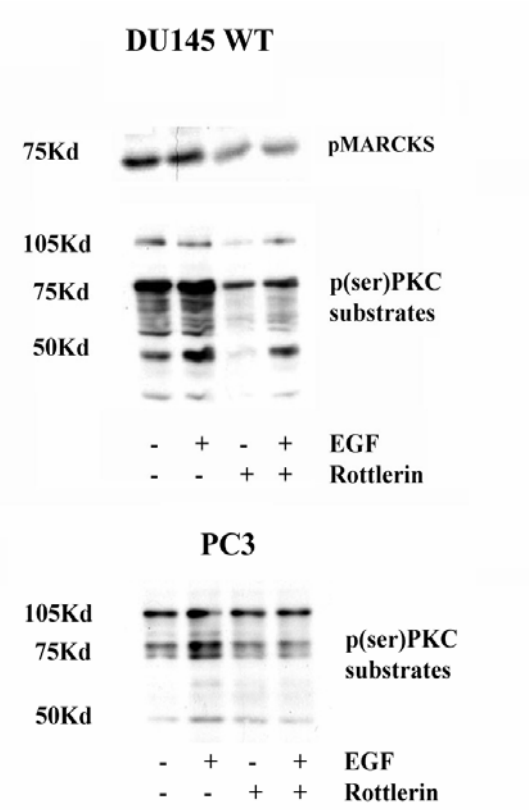
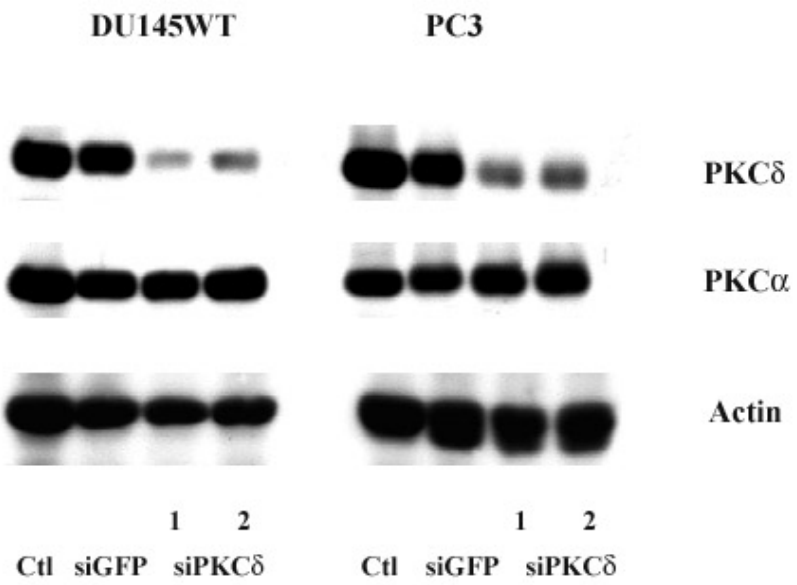


Figure 20. Abrogation of PKCδ activity using Rottlerin in prostate carcinoma cells.
A. DU145WT (top panel) and PC3 cells (bottom panel) were grown in complete growth media, quiesced for 24 hours in serum free media and stimulated with 10 nM of EGF for 30 minutes. Where indicated, 7.5 μM of Rottlerin was added to the serum free media, 30 minutes prior to EGF stimulation. Representative blot series of three different experiments are shown. B. Rottlerin decreases activation status of selective PKC substrates. Samples were prepared as mentioned above, resolved by SDS-PAGE and immunoblotted for activated MARCKS. Similarly, we probed for activated status of key PKC (serine) substrates in both cell lines. The experiment was performed three times.

A)



B)

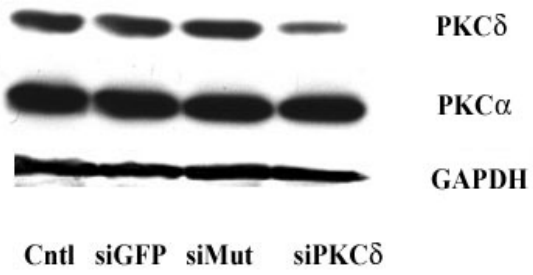


Figure 21. Downregulation of PKCδ using siRNA.
A. Specific or Mock (GFP) siRNA was transfected in prostate carcinoma cells as described in materials and methods. Immunoblot analysis of cell lysates 48 hours after transfection shows downregulation of PKCδ protein by more than 70% as determined by densitometry. A representative immunoblot of three independent experiments is shown. B. For a second control, cells were transfected with a mutated PKCδ siRNA sequence (siMut). Mutation in the siRNA sequence did not downregulate PKCδ levels as seen in the immunoblot confirming the specificity of the siRNA sequence.

Inhibition of PKCδ activity with Rottlerin decreased the EGF-stimulated migration of prostate cancer cell lines into the acellular area (Figure 22). Interestingly and in contrast to PC3 cells, Rottlerin did not reduce basal migration of DU145WT cells, consistent with the observation that basal (or unstimulated) phosphorylated levels of PKCδ were unaffected by Rottlerin at 7.5 μM concentration. These observations point towards a limitation of pharmacological agents in achieving complete inhibition of an enzyme at non-toxic doses. In accordance with this, cells that were transfected with PKCδ siRNA showed substantially lower transmigration through the 8 micron pores of PET membrane (Figure 23). Here, both basal as well as EGF stimulated transmigration was significantly reduced as compared to cells transfected with GFP siRNA that served as controls.

Motility is a required process for invasion of tumor cells through the surrounding stroma. Whether this reduced motility of PKCδ depleted cells led to reduced invasiveness, we assessed the invasive potential of both cell lines across Matrigel under the same conditions. EGF induced invasiveness of PC3 and DU145WT cells was significantly reduced by Rottlerin (Figure 24). PKCδ specific siRNA transfected cells were significantly less invasive than mock (GFP siRNA) transfected cells (Figure 25). Here, too, basal invasiveness was reduced in the cells lacking functional PKCδ. Thus, the decrease in invasiveness of prostate carcinoma cells as noted above correlated with their decrease in motility.

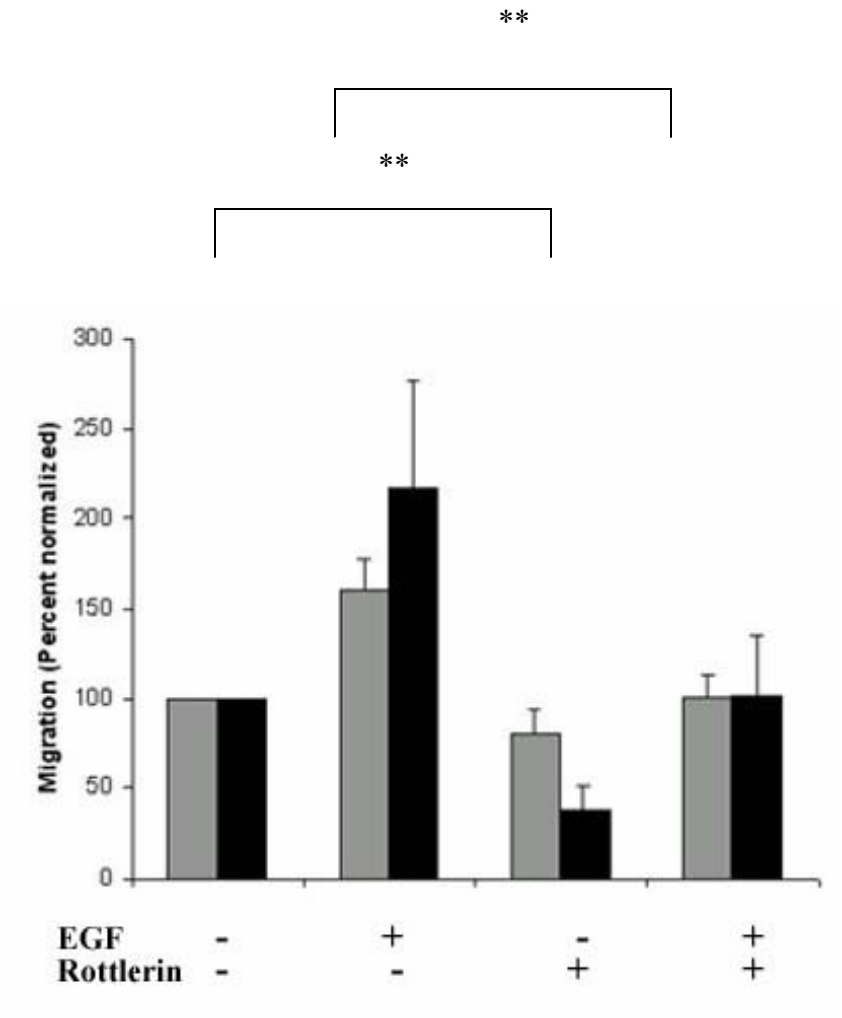


Figure 22. Abrogation of PKCδ activity using Rottlerin decreases migration of prostate carcinoma cell lines.
Rottlerin decreases migration of DU145WT (grey bars) and PC3 (black solid bars) cells. Cells were grown to a confluent monolayer in complete growth media, quiesced for 24 hours and scraped with a pipette tip to create a wound. The migration of cells into the acellular area was studied by incubating them in serum free medium containing 10 nM of EGF and 7.5 μM of Rottlerin for 24 hours. Each experiment of three was performed in triplicate; average ± S.E.M. are shown; (** = $p < 0.05$ comparing Rottlerin to diluent alone, or EGF plus Rottlerin to EGF alone).

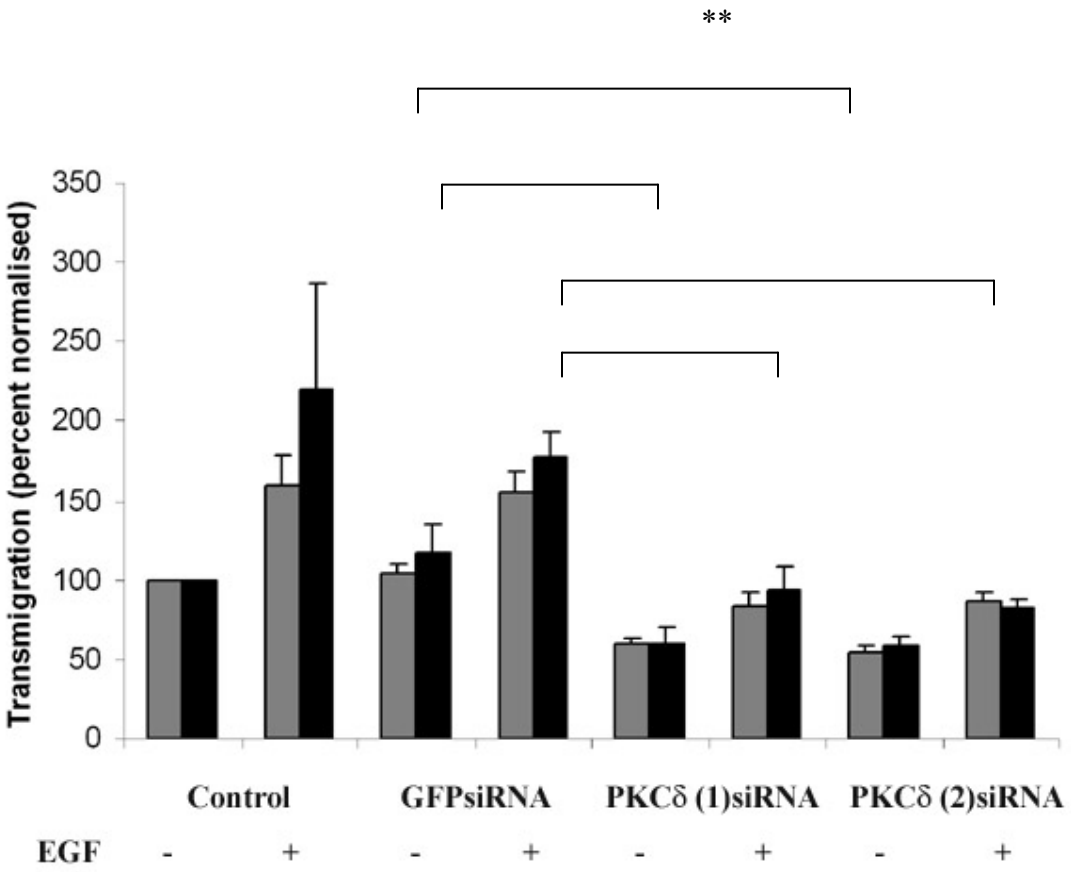


Figure 23. PKCδ depleted cells exhibit lesser transmigration across a transwell chamber.
PKCδ in DU145WT and PC3 cells was depleted using 2 siRNA sequences and 25,000 cells were loaded into the upper part of each transwell chamber. The data are average ± S.E.M. of three independent experiments performed in triplicate; (** = $p < 0.05$ as compared across the bars).

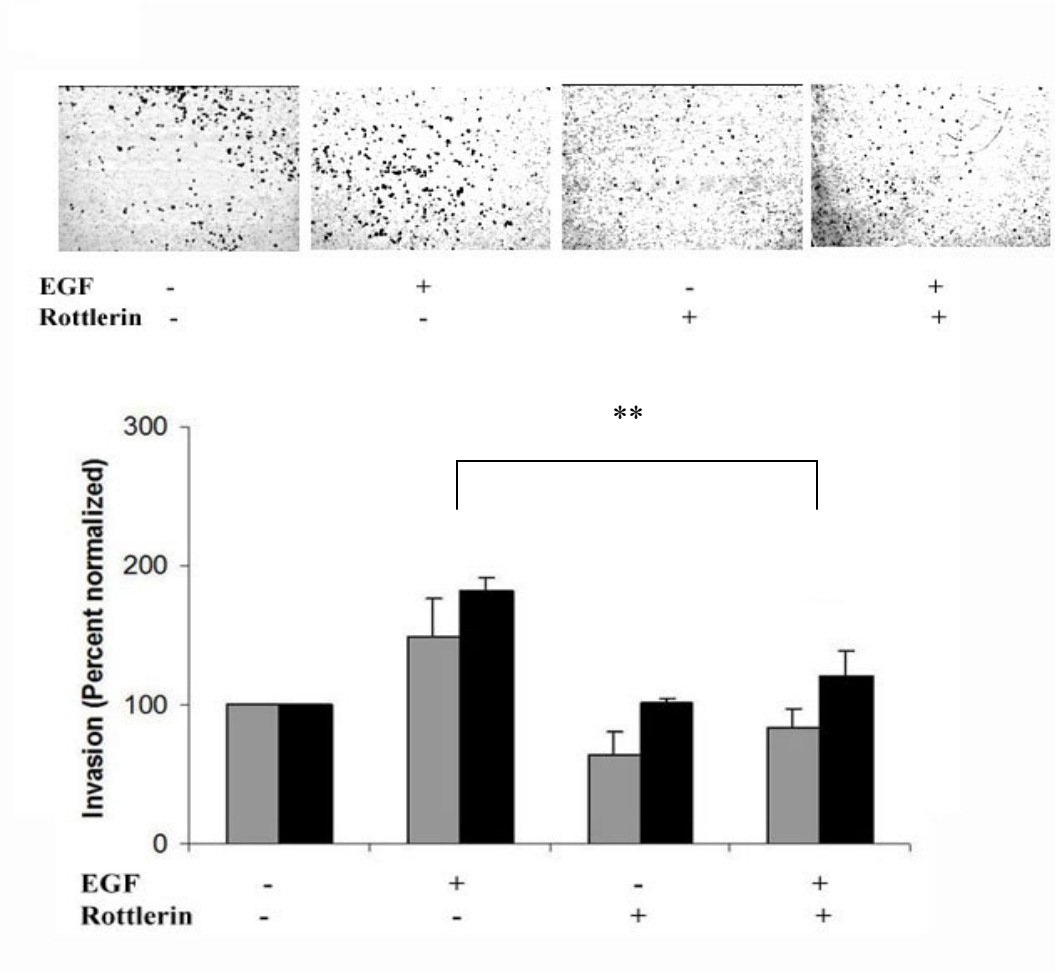


Figure 24. PKCδ abrogation using Rottlerin decreases invasion of prostate carcinoma cells.
25,000 DU145WT (grey bars) or PC3 cells (black solid bars) were loaded onto the upper well of each chamber and challenged with serum free medium containing 10 nM of EGF and 7.5 μM of Rottlerin. Data represent average ± S.E.M. of 3 different experiments each in triplicate; (** = $p < 0.05$ comparing EGF to diluent alone, or EGF plus Rottlerin to EGF alone). Representative pictures (top) of the membrane filters were taken at 10X magnification using a phase contrast microscope.

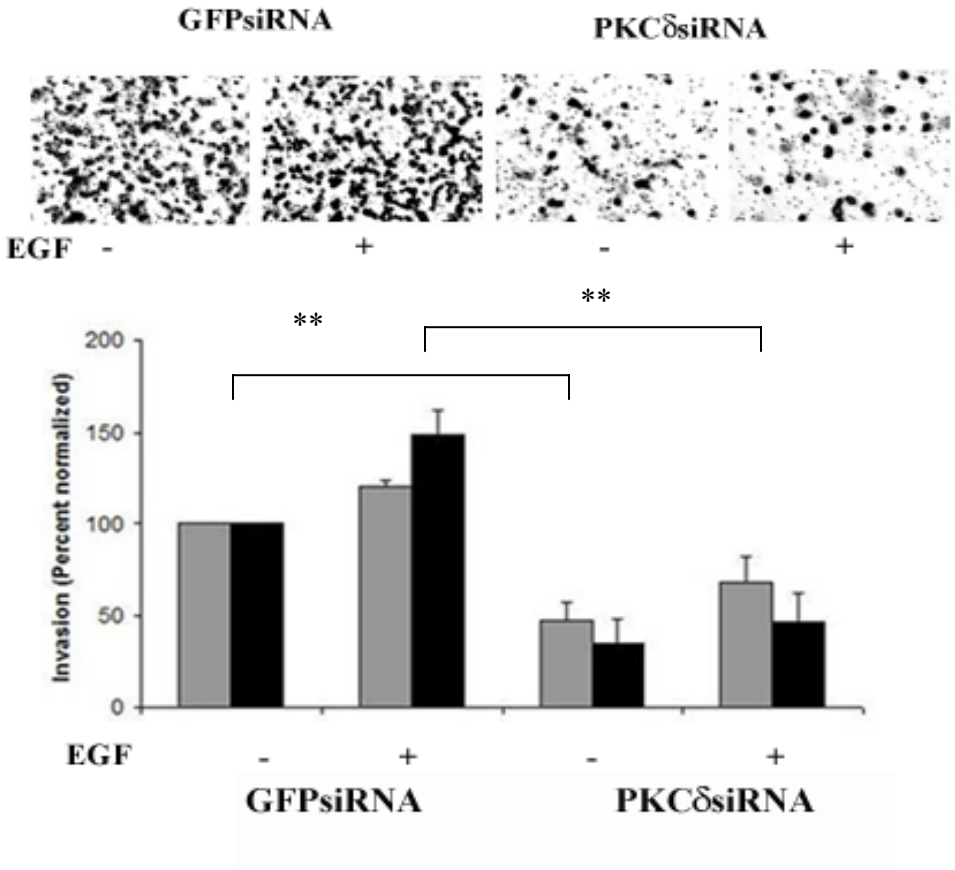
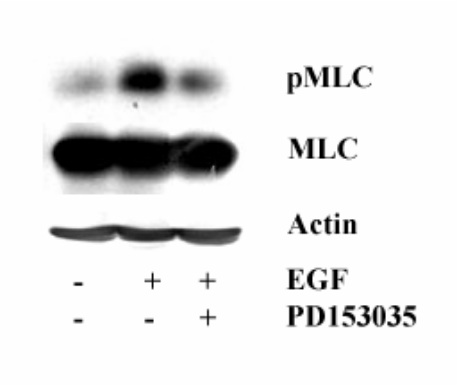


Figure 25. PKCδ downregulation using RNAi reduces invasion of prostate cancer cells.
Cells were subjected to the Matrigel invasion assay for 48 hours. The data represent the average ± S.E.M. of three independent experiments performed in triplicate; (** = $p < 0.05$ as compared across the bars . Representative pictures (top) of the membrane filters were taken at 10X magnification using a phase contrast microscope (twice as many cells were loaded for these pictures compared to those in Figure 3 to accentuate the decrease from basal noted for the siRNA).

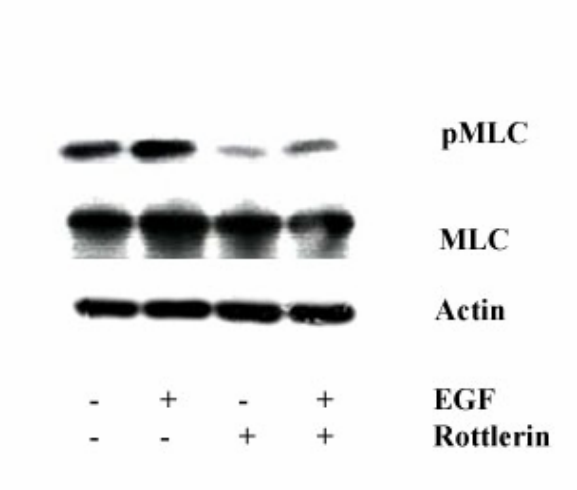
4.4.3 PKCδ modulates activation status of myosin light chain in PC3 cells

PKCδ modulates cellular contractility of fibroblasts [34], and reports have implicated a role for a PKC intermediary in EGF mediated myosin II activation and apparent subcellular localization in prostate carcinoma cells [166]. Limiting activation of the regulatory myosin light chains can inhibit motility via reducing cell contractility. We found that EGF activates myosin light chain in PC3 cells (Figure 26A). We assessed if PKCδ was an intermediary in mediating EGF induced cellular contractility in prostate cancer cells. Indeed, Rottlerin limited basal as well as EGF induced phosphorylation of MLC in PC3 cells (Figure 26B). Activation of MLC is evident within 15 minutes of addition of saturating levels of EGF and peaks at around 3-4 hours when motility becomes a stable biophysical response. We assessed MLC activation at 30 minutes and 3 hours of EGF treatment in both cell lines under specific siRNA transfected conditions. PKCδ siRNA transfected PC3 cells showed significantly lesser activation of MLC in response to EGF (Figure 26C). As our model of migration-related pathways from EGFR utilizes MLC-mediated contraction, an initial experiment demonstrated that blocking MLC activation pharmacologically limited invasion (data not shown). This finding is confirmatory of others well established in the literature [167, 168], though these earlier communications did not place MLC actions in the context of growth factor-mediated invasiveness.

A)



B)



C)

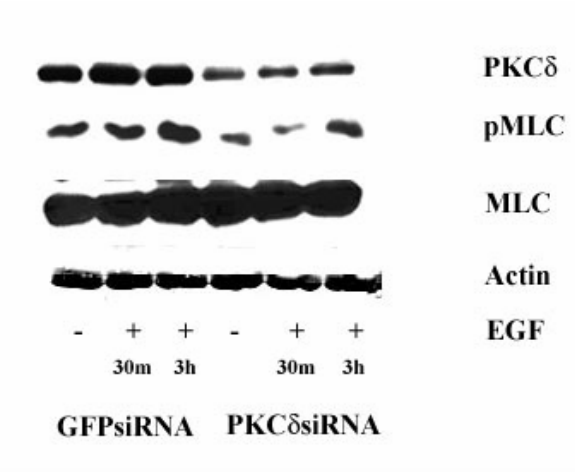
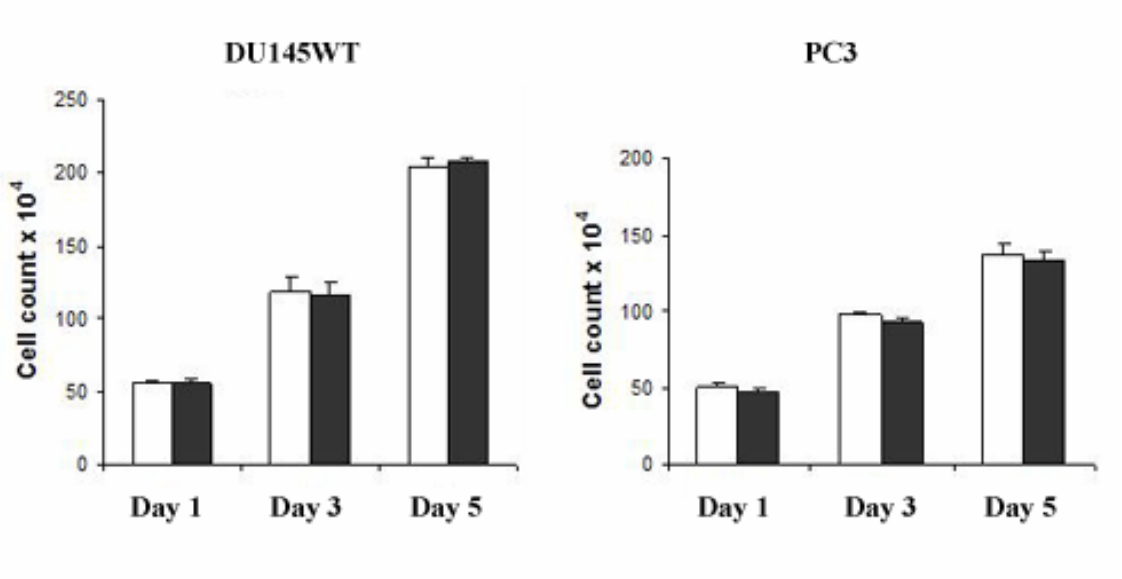


Figure 26. PKCδ modulates myosin light chain (MLC) phosphorylation in PC3 prostate carcinoma cells.
A. PC3 cells were grown in complete medium, quiesced for 24 hours and stimulated with 10 nM of EGF for 3 hours. Immunoblotting for specific proteins was carried out using SDS-PAGE. PD153035 at 1μM was added 30 minutes before EGF. B. Addition of Rottlerin decreases phosphorylation of myosin light chains in PC3 cells. C. PC3 cells were transfected with specific siRNAs as previously described. 48 hours after transfection, cells were quiesced for another 24 hours and stimulated with 10 nM of EGF for 30 minutes (short term) and 3 hours as indicated. In all sections, shown are one of three similar immunoblot series.

4.4.4 Depletion of PKCδ using siRNA does not affect proliferation of prostate cancer cell lines

To rule out the possibility that the decrease in invasion was a result of decreased cell viability, we assessed proliferation of transfected cells on the 3rd day (48 hours) and 5th day (96 hours) after siRNA transfection by direct live cell counting (Figure 27A). The PKCδ-depleted cells showed no difference in proliferation as compared to mock-transfected cells. Similarly, Rottlerin even at a higher concentration (10 μM), showed no significant toxicity as assessed by serial cell counts at 24 hours (Figure 27B). These time points correspond to the assay times in the invasiveness measurements above and are excess to those for motility and transmigration.

A)



B)

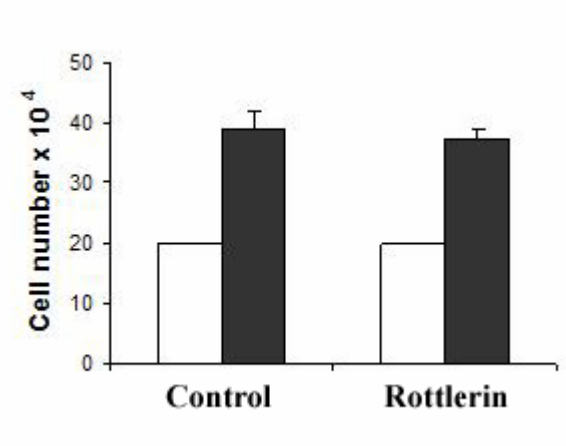
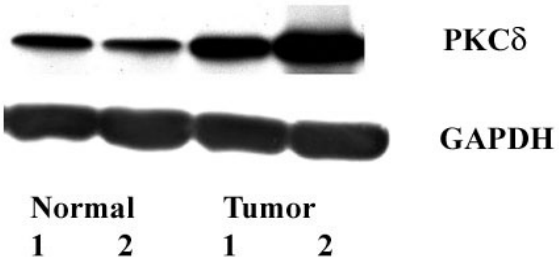


Figure 27. PKCδ signal abrogation does not affect proliferation of prostate carcinoma cells.
A. Downregulation of PKCδ using siRNA does not affect cell proliferation. Prostate cancer cells were transfected with 80 picomoles of specific siRNAs (GFPsiRNA – open bars; PKCδsiRNA-black solid bars) as indicated on Day 1. Cell proliferation was assessed by live cell counting using an automated coulter counter method. 48 hours post-transfection siRNA transfected cells were resuspended in 1 ml of growth medium. Cells were then counted using a Coulter counter. Similarly, cells were counted 96 hours post-transfection or on Day 5. B. Rottlerin does not affect cell proliferation/viability at 10 μM concentration. To assess the toxicity of Rottlerin on prostate cancer cells, we incubated equal number of both PC3 and DU145WT cells with 10 μM Rottlerin in serum free medium for 24 hours and assessed live cell count using the coulter counter method as mentioned above. Open bars represent day 1 and black solid bars represent cell count on day 2. In both parts, shown are mean ± S.E.M. of three experiments each in triplicate.

4.4.5 PKCδ is overexpressed and activated in human prostate cancer

As PKCδ is linked to properties of carcinoma progression, we determined whether this was also found in *de novo*-occurring human prostate carcinomas. We found that prostate cancer tissue overexpressed PKCδ protein relative to normal donor prostate in two specimens of each (Figure 28A). Based on this, we analyzed five prostate adenocarcinoma tissue sections for activated PKCδ using IHC against phosphorylated PKCδ. Immunohistochemical analysis showed that all five prostate cancer tissue sections presented pPKCδ staining predominantly in the nucleus while relatively lower staining was visible in the cytoplasm. Significantly lower pPKCδ staining was noted in normal donor prostates (Figure 28B). While all the prostate carcinomas were of moderately severe nature (Gleason scores of 7 or 8), this positive staining also included foci of high grade prostatic intraepithelial neoplasia (PIN). Initial staining with a prostate tissue tumor micro-array demonstrates similar distinction of CaP and PIN from normal donor prostate, strongly suggesting that differences were not due to staining issues. While a more-extensive and indepth study is underway to evaluate pPKCδ as a marker for prostate cancer and/or its progression, these initial findings demonstrate a clear distinction in PKCδ and pPKCδ levels in human prostate cancer.

A)



B)

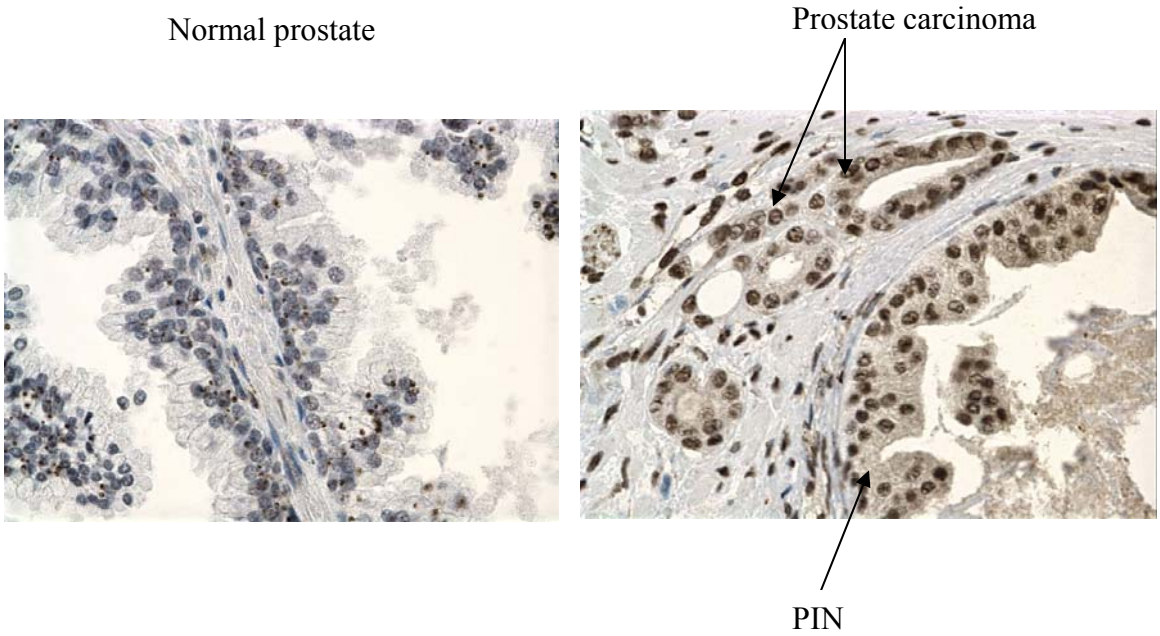


Figure 28. PKCδ is overexpressed and activated in prostate cancer specimens.
A. Tissue lysates were prepared from two normal donor prostate and two prostate cancer tissue samples as described. 10 μg of protein was loaded in each lane and total PKCδ levels were probed; GAPDH served as a loading control. A representative immunoblot of three independent experiments in shown. B. Representative photomicrographs from one of the five normal prostate tissues and one of five prostate adenocarcinomas are shown. The tumor tissue section also contains foci of prostatic epithelial neoplasia (PIN). These specimens were evaluated for activated PKCδ using immunohistochemistry (IHC) against pPKCδ. Normal prostate tissue shows no staining with phosphorylated-PKCδ antibody as compared to a strong staining in all five prostate carcinomas including prostatic epithelial neoplasia (PIN).

Condition	Mean	Range	s.e.m.
CaP	1.80	1 -- 3	0.37
PIN	2.20	1 -- 3	0.37
Nl Donor	1.00	1 -- 1	0.00

Table 3. Quantitation of the prostate cancer (CaP), PIN, and normal donor prostate staining.
The nuclear staining was scored on a scale of 0 to 3, with 0 being no epithelial cell staining, 1 being <10% of nuclei staining, 2 being >10% but <25%, and 3 being >25%. Both CaP and PIN are different from normal donor at $P < 0.05$ but similar to each each other, using the Student ttest.

4.5 DISCUSSION

Growth factor - or cytokine - induced cell motility is a crucial determinant of progression and dissemination of many tumors, particularly prostate cancer [169]. Previous studies have shown that targeting one or more biophysical events underlying cell motility can be an effective modality in limiting cancer spread in experimental systems [14, 15]. One outstanding question is which key molecular switch to target in an attempt to limit tumor progression [138]. Mathematical modeling of cellular signaling cascades has recently highlighted the central role of myosin light chain and hence, transcellular contractility, in cell motility [11]. As such, we determined whether abrogating PKCδ, a molecular switch for growth factor receptor - mediated contractility and thereby migration, limits the motility and invasiveness of prostate cancer cell lines.

PKCδ activity was reduced both pharmacologically and molecularly. This dual approach was used to avoid the unanticipated side effects and unknown targets of any singular intervention. While the interventions did not alter the level or phosphorylation of the other classical PKC isoforms expressed in these prostate carcinoma cells – namely PKCα and PKCβ (DU145 cells also present message for PKCε (data not shown)), other unknown targets cannot be excluded. To address the question of the specificity of molecular interventions, we constructed a mutated siRNA for PKCδ by changing G to C at two nucleotide positions within the PKCδsiRNA sequence. This mutant siRNA did not lower PKCδ (nor PKCα) protein levels nor did it alter the transmigration of cells as compared to the control (untransfected) or mock (GFPsiRNA) transfected cells (Figure 21B). Additionally, similar results were obtained using two specific albeit different PKCδsiRNA sequences (Figures 21A and 23). These results indicated that the resultant reduced motility and invasiveness were due to disruption of PKCδ signaling.

EGFR autocrine signaling drives prostate tumor cell invasiveness both *in vitro* and *in vivo* [60, 162]. Of the two cell lines utilized, DU145WT cells present substantially higher autocrine signaling predominantly due to higher number of EGF receptor expression on the cell surface and exhibit significantly higher invasiveness compared to both PC3 and DU145 parental cells [60, 162]. Such autocrine signaling evidently induces sustained activation of numerous

intracellular signaling molecules enabling tumor proliferation and invasion. In above studies and in contrast to PC3 cells, Rottlerin did not abrogate the basal (unstimulated) migration of DU145WT cells. In a "closed" cell system enriched with active autocrine circuit loops, it is likely that additional signaling molecules other than PKCδ provide the necessary signals during such basal migratory conditions.

Of interest, the previously described activation of myosin contractility due to phosphorylation of MLC [34], was downregulated in PC3 cells but was muted in DU145WT cells. This argues for additional inputs into transcellular contractility in the DU145WT cells. Rho kinase, zipper-interacting protein kinase (ZIP Kinase), and integrin linked kinase (ILK) are obvious candidates under investigation [37, 38, 42, 170], but such ongoing studies lie beyond the realm of the present communication. As abrogation of PKCδ signaling is effective in limiting motility and invasiveness in DU145WT cells, it also raises the questions of additional targets for PKCδ in these cells. MARCKS, when phosphorylated, can associate with actin and actin modifying proteins and can affect cellular cytoskeletal reorganization as well as motility of a variety of cells[171]. Additionally, when activated using PMA or EGF, PKCδ translocates to the plasma membrane and the cytoskeleton [172], where it can modify by phosphorylation, the activity of numerous cytoskeleton associated proteins such as adducins that are crucial for motility [159]. In this vein, PKCδ is known to affect focal adhesion protein turnover by modulating focal adhesion kinase and paxillin activation as well as subcellular vinculin distribution [173]. Thus, a subtotal depletion of PKCδ as achieved by siRNA can affect cell motility not only by downregulating PKCδ activity but also by inhibiting PKCδ – cytoskeletal / focal adhesion protein interactions. Future investigations beyond the current scope of this manuscript are aimed at elucidating different focal adhesion proteins that are affected by PKCδ in prostate cancer cells.

PKCδ has been implicated in PMA induced apoptosis of androgen responsive prostate cancer (LnCaP and C4-2) cells [80], leading to the question of whether apoptosis may play a role in our decreased invasiveness. We do not note decreased number of PC3 and DU145 cells in our studies; this may be due to the fact that these are both androgen independent cell lines which present vigorous autocrine growth factor receptor stimulatory loops. While these cells were never challenged with PMA in our studies, activation of PKCδ, via EGF failed to induce spontaneous apoptosis but rather induced motility, at least in part by downstream

activation of MLC. Thus, it is clear that mere activation of PKCδ is not sufficient to dictate a signal for apoptosis, though it is possible that the pools of PKCδ activated by PMA and EGFR are distinct. Additionally, although PKCδ is required for apoptosis in LnCaP cells in response to PMA, it alone is insufficient in doing so and requires additional and redundant signaling from either PKCα or PKCε [156]. These underline the fact that the role PKCδ plays in survival / apoptosis of prostate cancer cells is yet to be clearly elucidated, but preliminary reports suggest, that apoptosis is actively signaled by this kinase in androgen-dependent cells, whereas more mesenchymally-transitioned cancer cells utilize it for promoting their invasive properties. In such cells, PKCδ may be preferentially activated by continual EGFR autocrine signaling.

Our experimental findings are supported by the analysis of PKCδ levels (total and activated) in prostate cancer tissue. Although some expected quantitative variation in total PKCδ levels within the samples is observed, both total and activated PKCδ levels are substantially higher in prostate cancer relative to normal donor prostate. Interestingly, activated PKCδ is readily found within the nucleus of tumor cells. Nuclear PKCδ has been shown to phosphorylate acidic fibroblast growth factor (aFGF or FGF-1) and control its export to the cytoplasm thereby modulating cell growth and migration [174]. Reinforcing autocrine signaling cascades have been defined in prostate cancer lines downstream of EGFR signaling [3]. This finding opens a new avenue for exploration in prostate cancer biology.

Our data present for the first time that PKCδ signaling is critical for prostate cancer cell invasiveness. This agrees with an earlier report employing breast cancer cell lines [159, 161], but herein we further demonstrate that this signaling results from EGFR activation that functions in an autocrine manner in practically all prostate carcinomas. Further, we provide evidence linking PKCδ activity to prostate cancer in *de novo*-occurring human prostate cancers. These findings support the potential of PKCδ as a target for rationale cancer interventions aimed at limiting tumor dissemination.

4.6 ACKNOWLEDGEMENTS

We thank Dr. Akihiro Iwabu, Dr. Asmaa Mamoune and all the members of Wells and Lauffenburger laboratories for insightful discussions that helped prepare this manuscript. We greatly appreciate the help of Aprell Delo and Marie Acquafondata in selecting and staining the prostate biopsies. This work was supported by grants from the DoD Prostate Cancer Program and the NCI (NIH).

5.0 PERSPECTIVES AND CONCLUSIONS

5.1 MODELING APPROACH TO CELL BIOLOGY

Our studies consist of two distinct approaches driven towards cell biology—first, identification of critical signaling pathways that govern cellular responses and second, targeting such signaling switches to alter cell behavior. Such an approach is necessary to answer some of the questions in cell biology. The answers are obviously difficult to extract, given the complexity of signal transduction events, and hence demand innovative techniques that can help 'sort' the laboratory data into stratifications which explain signal-response associations. We applied such an approach, Decision tree analysis, towards understanding cellular biophysical events during migration. This approach is innovative since for the first time, this methodology has been utilized to elucidate proteomic events with biophysical responses. Our observations shed light upon how crucial activated signaling switches collaborate during cell migration over a range of extracellular conditions.

The limited amount of data from experimental work, such as ours, often prevents derivation of insightful conclusions about intricate biological cell phenomena. In our case, we measured, using western blotting, the activated status of key signaling proteins using phosphorylated-specific antibodies. These proteins are actively recruited by cells during one or more biophysical events during motility. Quantitative measurements of such proteins were interpolated using polynomial modeling to create simulation data points that accurately represented (and expanded) actual measurements. We utilized a strict quality control method, using 95% confidence interval to such replicates and eliminated any outliers. Thus, the data was 'filtered' and expanded using polynomial modeling. Such mathematical solutions are critical since traditional biology experiments need tremendous effort and technical consistency to

produce few data points in a reasonable amount of time. Similar replicates were created for cell speed measurements. These data were then utilized to create decision trees that mapped the hierarchy of different signaling proteins in mediating cell migration speed. Our model discretized speed into three different categories; low, medium and high. Each of the molecules within the decision tree model was also discretized into the above three levels depending upon their activation status. Our model showed a 'step-by-step' regulation of cell speed by different signaling molecules. More importantly, it addressed the quantitative contributions of signaling proteins in dictating such a response. Central to this model was the EGFR activation. Each node was further divided into two branches corresponding to the activated status of downstream signaling proteins. The final speed of migrating cells was predicted by traversing down the branches of the decision trees and delineated specific molecular quantitative contributions. Our model predicted 76% observations (for cell speed) within the training data set accurately based solely on the activated state of these five proteins. Even more non-intuitively the model pointed out the role of transcellular contractility as a critical 'balancing switch' in cell motility. Specific 'quantitative' abrogation of myosin light chain activation is the key to desired cell response. This prediction applied to cancer cells as well. Complete limitation of cellular contractility retarded tumor cell motility and invasion whereas its partial abrogation paradoxically increased these two properties. These findings have immense implications in therapy of diseases, such as cancer where dysregulation of cell motility is central to its pathogenesis and progression. Our approach is unique in that it is one of the first few systems biology studies applied to cell migration.

5.2 UNDERSTANDING CELL MOTILITY USING SYSTEMS BIOLOGY APPROACH

Traditional biology experiments have grown on simplified models of cell migration. One such model is to divide motility into individual biophysical events (as elaborated in the introduction section). These individual biophysical events are very complicated and dictated by a myriad of signaling proteins. Thus, integrating these events to study migration is a daunting task. Almost all of these studies were carried out by abrogating or 'knocking out' one or more individual

molecules and showing that motility is retarded under such conditions. While such experiments provide a proof of principle about the molecule's requirement in motility, it sheds little light, if any, on how such a molecule interacts with other similar crucial molecules within the motility–signaling network. It is vital to understand such 'complementary signaling' within a proteomic network to target specific 'un-expendable' signaling nodes needed to manipulate cell behaviors. In cancer for example, this can minimize signaling from alternative pathways and lead to more effective therapy. Thus, a holistic approach is required to understand protein-protein interactions in a variety of biological phenomena, motility being one of them.

We took a systems biology path towards understanding cell migration based on the level of activation of intracellular signals. This approach is relevant for many purposes, but most significantly because intracellular signals transmit extracellular information within the network that leads to the ultimate cell decisions. Secondly, they can be abrogated or amplified as required using pharmacological or genetic methods to manipulate cell fate. Pharmaceutical agents have exploited this property and have been designed precisely on such a concept. But pharmacological agents have a limited impact in therapy because, at least in chronic diseases, multiple signaling switches are dysregulated. Hence, targeting nodes, where these signaling pathways intersect, holds the key to future therapeutic goals.

Systems biology answers this particular concern. This is because it involves a simultaneous measurement of multiple signaling events over a range of stimuli (inputs) within the 'cell system'. Such signals are then integrated and processed to generate important output responses. Using a variety of mathematical or statistical applications, the signaling events that are most relevant to the response are mapped out; e.g. Janes et al have elucidated, using principal component analysis, the set of activated signals that govern apoptosis or survival (output) upon simultaneous stimulation with apoptotic (TNF-α) or survival (Insulin, EGF) signals (inputs). Despite the computational modeling and (pseudo)expansion of data sets, the signaling events are able to accurately predict the biological phenomena since the original data is measured from such experimental conditions. The net concept is to generate a holistic model of *input-signal-response* associations and accurately predict responses from a signature of activation profile of intracellular switches. Thus, signaling events are studied as a whole and represents a final outcome of such proteomic interactions rather than individual linear phenomena.

5.3 IMPLICATIONS FOR CELL BIOLOGY

We generated an *extracellular cue-signal-motility* model using decision tree analysis. This model was based on measuring activation profile of five signaling proteins over a range of extracellular cues (EGF, FN). Both EGF and Fibronectin play a crucial role in motility of a variety of cells including fibroblasts and keratinocytes during wound repair. As such, the experimental conditions as well as the findings are applicable to the *in vivo* environment. Additionally, there is substantial crosstalk between these two physiological ligands and their respective receptors during wound healing; e.g. EGF increases fibronectin expression via PKCδ in human dermal fibroblasts during wound healing [175]. Similarly, fibronectin concentration of substratum influences EGF induced directional persistence in migrating cells by mediating cellular lamellipodal attachments to the substratum [176]. Thus the paradigm "adhesion is required for migration" holds true even further and warrants the need to study these two events together.

The physiological implications of our model can be summarized into three points:

[1] Our model delineates the quantitative signaling events that govern cell speed. Since most pharmacological interventions depend on quantitative parameters, such an approach is essential to further understanding of 'how much' an individual signaling switch contributes to motility. This contribution is relative to and dependent on that of other signaling proteins. The decision tree also explains how each molecule further depends on the others to mediate speed; e.g. interpreting from Figure 16, if EGFR is actively signaling (1 and/or 2) and MLC activation is low (0), then cells move with higher (2) speed. However if the MLC activation is intermediate (1) then the cells move with higher speed (2) if PKCδ activity is low (0). Thus, by reading the decision tree down each node, more proteins get incorporated into the equation. However, this does not mean that the predictions solely based on maximum number of signaling proteins are more accurate. As previously explained, models have to be utilized to extract the most significant information while reducing unnecessary complexity. This is because the experimental designs of such complex predictions are too complicated and as such, the use of many pharmacological agents to test such predictions can have unwanted 'off-target' effects confounding the experimental results. Furthermore, it is important to assess the number (or percentage) of observations that the model correctly predicts from the training data-set. If that percentage is very low, it means that more information is needed, probably by incorporating additional

molecules within the model to make such predictions. Our model will be 'refined' by incorporating other key players that have shown to be involved in EGFR-integrin receptor crosstalk during motility. Some of these are FAK, calpain and Rho GTPases. Thus, a future 'fine-tuned' model will seek to incorporate the above mentioned players, at the least, and will be expected to predict cell speed with better accuracy.

[2] Simplistically, our model places MLC mediated transcellular contractility as the most significant event in EGFR induced motility. Total disruption of any of the key biophysical events, including contractility, has shown to arrest cell motility. However, our model also predicted that the effects of subtotal MLC inhibition are completely opposite. We found that subtotal interventions increased motility of both fibroblasts and breast cancer cells rather than decreasing it linearly. The effect of MLC inhibition on cell migration is thus biphasic with highest motility found at relatively lower contractile state. This reflects on the importance of *adhesion versus contractility* balance that controls cell locomotion. Cells cannot move if surface adhesiveness is too less or too high for the cells to break adhesions and form newer ones with the substratum. Thus only when the contractility/adhesion balance is optimal, does the cell move fastest. This level is represented by a surface coated with 1 μg/ml of fibronectin in our experimental set up.

[3] Our model was based on data generated through laboratory experiments. The model can be utilized to extract and answer other non-intuitive questions; e.g. what are the effects of subtotal inhibition of ERK (or rear detachment) and PLCγ (or lamellipodal protrusion) on cell speed? Can lowering of membrane detachment increase contractility and retard cell speed? More importantly, does subtotal MLC inhibition modulate cell adhesiveness and how is it different at the two extreme fibronectin conditions? How does surface fibronectin concentration affect EGF induced contractility during total and subtotal MLC inhibition? Thus the model answers some questions but posits newer ones. Answering some of these questions can lead to a great insight into how migration is coordinately regulated. From our experiments that involve both cell contractility and adhesiveness, it is apparent that MLC inhibition shifts the balance between adhesion and contractility towards the optimal level. When cells find the optimum contractility/adhesion ratio, they attain a maximum speed of movement.

In summary, we present a systems biology approach to understanding cell migration. Such approaches warrant further employment in cell biology since a single response (or the

ultimate cellular outcome) is mediated by numerous interacting signaling molecules that fluctuate in their activation status within the 'cell system'. Thus, studying such molecules together can shed light on their contributions during the most complex biological events. Our model elucidates the interaction between the key players in a simplified motility-signaling network and explains how these proteins together control cell migration.

5.4 IMPLICATIONS FOR CANCER BIOLOGY AND THERAPY

Tumor cell motility is a crucial determinant of cancer progression to metastasis and results from a cell's acquisition of a variety of properties within its proteome. This is possible since a tumor cell can epigenetically modify signaling through its protein networks via vigorous upregulation of certain receptors and/or ligands that transmit 'beneficial' signals downstream. Most chemotherapeutic agents target the proliferative property of tumor cells. Hence, in cancers like that of prostate, where tumor growth is evidently slow, such agents are of limited efficacy and rather produce pronounced systemic toxicities. Thus, novel agents that target 'motility-specific' signaling switches are needed so they can be more effectively utilized in preventing cancer metastasis by increasing patient compliance.

One difficult question however, is to identify which signaling node, amongst hundreds, to disrupt in an attempt to successfully retard tumor cell migration. Our modeling approach identified MLC mediated contractility as a crucial biophysical event during migration. Based on such a prediction, we targeted one such switch, PKCδ, and found that such selective nodal disruption retards tumor invasion principally by decreasing cell contractility and motility. Abrogation of PKCδ did not affect proliferation or survival of tumor cells. Such 'motility-specific' molecular abrogation offers a promising approach in treatment of cancers such as those of prostate where much of the morbidity and mortality results from invasion of tumor cells through the stroma and forming metastatic foci. PKCδ additionally represents an important target since it is overexpressed in prostate cancer cells but not in normal prostate epithelial cells and is activated downstream of EGFR, signaling through which is vigorously upregulated in a variety of solid tumors including that of prostate. Reinforcing autocrine signaling through EGFR is

evident in prostate cancer progression with this property correlating with the emergence of a more aggressive 'androgen independent' cancer phenotype typically seen after few years of conventional androgen deprivation therapy [68, 177]. At such a stage, tumors depend on autocrine growth factors rather than androgens for their growth and survival. Thus, innovative therapy of aggressive prostate cancer should target such growth factor receptors that provide aberrant signaling inputs favoring tumor invasion. Such multiple and continual inputs cause tremendous alterations in the cellular interactive proteome. A systems biology approach is particularly relevant to studying tumor biology since continual autocrine signaling prevents rational interpretation of experiments involving growth factor stimulation. Thus a simultaneous measurement of protein clusters involved in tumor proliferation or migration can enable creation of a systems model for a cancer cell. Mathematical modeling techniques such as decision tree analysis constructed in fibroblasts can serve as a reference or 'normal' model with predictions being applied to a variety of cancer cells. The discrepancies, obviously apparent, can enable further understanding of how motility is dysregulated in cancer invasion with such findings confirmed using laboratory experiments. More importantly, it can identify quantitative proteomic alterations underlying such altered cell behavior. Such proteins can be targeted together or singularly to arrest tumor invasion.

Our study targeting PKCδ in tumor invasion opens a new avenue for cancer drug research. Since cancer is now known to be an aberrant overgrowth of cell populations that are heterogenous in their proteomic as well as genomic signatures [64], future treatments need to be personalized by further stratification of tumor phenotypes depending on their 'kinome'[178]. Reverse protein microarrays offer one such technology that enables detection of numerous intracellular activated kinases in selected cancer cells isolated using laser capture microdissection of tumor tissue samples [179, 180]. Depending upon the signature of signaling proteins, each tumor tissue sample can then be further sub-categorized. Thus, patients would then receive a particular chemotherapeutic regimen depending upon a 'set of active signaling protein profile' detected within the tumor biopsy [132]. This phenomenon was clearly responsible for rapid symptomatic improvement and tumor regression in a subset of lung cancer patients treated with EGFR tyrosine kinase inhibitor, Gefitinib (Iressa) [191, 192]. Majority of these patients harbored activating mutations in EGF receptor tyrosine kinase catalytic domain that enabled signaling events leading to tumor survival and proliferation. Such mutant receptor was readily

abrogated using Gefitinib [192]. Thus, by further dissecting the molecular aspects of cancer cells, treatment options can be individualized for patients. Such personalized therapy is expected to grow in future. With newer biological agents arriving in the market everyday, it is clear that careful selection of targets and their disruption within proteomic networks is key to the treatment of this deadly disease.

APPENDIX A

POLYNOMIAL EQUATIONS FOR PARAMETRIC MODEL DATA

For data simulation we assume the following model for *i*th value of *j*th observed variable such as the migration speed or a signaling protein:

$$g_2(y_{i(j)}) = f_{p(j)}\,(g_1(x_{i(j)})) + \varepsilon_j, \qquad (1)$$

where x_i denotes *i*th experimental condition (e.g. the level of fibronectin and the absence or presence of EGF) for *j*th variable, $f_{p(j)}$ is *p*th order polynomial for *j*th variable with parameters

$\beta^{(p)T} = [\beta_p\ \beta_{p-1},\ldots,\ \beta_0]$, $g_{1,2}$ are transformation functions and ε_j is an error term. For example, when p=2 and g_1, and g_2 are identity functions, $y_{i(j)} = f_{2(j)}\,(x_{i(j)}) + \varepsilon_j = \ldots$. Although transformation functions are usually identity functions, sometimes it is beneficial to perform the fitting in log-space (g1(x) = log(x)) or in log-log space (g1(x)= g2(x)= log(x)). In general, experimental conditions may vary between the variables and therefore quantities in Eq.1 depend on *j*. In the subsequent discussion, however, the subscript *j* is dropped for notational convenience.

The challenge with polynomial modeling is to choose order of the polynomial (*p*) that describes the data best without overfitting. In order to solve this challenge we use the minimum description length (MDL) principle [188]. The basic idea behind the MDL principle in model selection is to find the model that gives the minimum stochastic complexity relative to the model class [189]. Stochastic complexity can be understood as a measure of the goodness-of-fit of a model based on the model's ability to compress the data given a model class. As statistical inference is viewed as a data compression problem, there is no need to assume underlying, ``true'' data generating distributions. Thus, apart from choosing the model class, there is no need to make subjective assessments.

To be more precise, we use normalized maximum likelihood (NML) approach [190], which follows when the MDL principle is applied to the maximum likelihood estimation. Let γ belong to Ω and be a restricted set of indices for the current polynomial order k and $X_\gamma \in R^{nxk}$ $\mathbf{R}^{nxk}$ be a matrix of predictor values with indices gamma. For example, in our case study when k=2 (line fitting) the first column of X_γ is [0.1 0.3 1 3]^T and the second **1**.

We assume that $\varepsilon_i \sim N(0,\tau)$, so the response data ($y^n = y_1,\ldots, y_n$) are also Gaussian distributed with density function $f(y^n,\gamma,\beta,\tau) = \frac{1}{(2\pi\tau)}\exp(\frac{1}{2\tau}\sum_i (y_i - \beta^T x_i)^2)$. Thus, maximum likelihood solutions for a fixed γ are $\hat{\beta}(y^n) = Z^{-1} X_\gamma^T y^n$, (2)

$$\hat{\tau} = \frac{1}{n}\sum (y_i - \hat{\beta}(y^n)^T x_i)^2 , \quad (3)$$

where $Z = X_\gamma^T X_\gamma = n\Sigma_\gamma$. In subsequent discussion the subscript γ is dropped.

The NML density function is defined as

$$\hat{f}(y^n;\gamma) = \frac{f(y^n,\gamma,\hat{\beta},\hat{\tau})}{\int_{Y(\tau_0,R)} f(z^n,\gamma,\hat{\beta}(z^n),\hat{\tau}(z^n))}, \quad (4)$$

where y^n is restricted to the set $Y(\tau_0,R) = \{z^n : \hat{\tau}(z^n) \geq \tau_0, \hat{\beta}^T(y^n)\Sigma\hat{\beta}(y^n) \leq R\}$. Parameters τ_0 and R are determined so that the maximum likelihood estimates are within $Y(\tau_0,R)$.

The NML density function is unique solution to the minmax problem:

where q range over any distributions [190]. Therefore, solving Eq.4 results in the best model for the data relative to the chosen model class. Evaluation of Eq. 4 (for details, see~ [190] gives the final decomposition for finding the best polynomial order:

$$\min_{\gamma\in\Omega}\{(n-k)\ln(\hat{\tau}) + k\ln(n\hat{R}) + (n-k-1)\ln(\frac{n}{n-k}) - (k+1)\}, \quad (6)$$

where $\hat{R} = \hat{\beta}^T(y^n)\Sigma\hat{\beta}(y^n)$. In our case study, the NML criteion (Eq. 6) is used to find the best polynomial order (p).

After finding the best polynomial order, the parameters for that model are known from Eq. 2 and these are taken to estimate β. In addition to estimating the parameters for f_p in Eq. 1, it is

necessary to have an estimate for the standard deviation for ε_j. One way to get this is to first pool individual bootstrap error estimates:

$$s^{(pooled)}j = \sqrt{\frac{\sum_{i}^{m}(n_i - 1)s_i^2}{\sum_{i}^{m} n_i - m}}, \qquad (7)$$

where s_i is a bootstrap error estimate and then either use $s^{(pooled)j}{}_i$ directly or squared.

BIBLIOGRAPHY

1. Tran, K.T., L. Griffith, and A. Wells, Extracellular matrix signaling through growth factor receptors during wound healing. Wound Repair Regen, 2004. **12**(3): p. 262-8.

2. Satish, L., D. Yager, and A. Wells, Glu-Leu-Arg-negative CXC chemokine interferon gamma inducible protein-9 as a mediator of epidermal-dermal communication during wound repair. J Invest Dermatol, 2003. **120**(6): p. 1110-7.

3. Wells, A., Tumor invasion: role of growth factor-induced cell motility. Adv Cancer Res, 2000. **78**: p. 31-101.

4. Norman, M.U. and M.J. Hickey, Mechanisms of lymphocyte migration in autoimmune disease. Tissue Antigens, 2005. **66**(3): p. 163-72.

5. Sanchez-Madrid, F. and R. Gonzalez-Amaro, Drugs, inflammation and cell adhesion receptors. Expert Opin Pharmacother, 2001. **2**(1): p. 3-17.

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

7. Lauffenburger, D.A., Cell motility. Making connections count. Nature, 1996. **383**(6599): p. 390-1.

8. Petricoin, E.F., 3rd, et al., Mapping molecular networks using proteomics: a vision for patient-tailored combination therapy. J Clin Oncol, 2005. **23**(15): p. 3614-21.

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

10. DiMilla, P.A., K. Barbee, and D.A. Lauffenburger, Mathematical model for the effects of adhesion and mechanics on cell migration speed. Biophys J, 1991. **60**(1): p. 15-37.

11. Hautaniemi, S., et al., Modeling of signal-response cascades using decision tree analysis. Bioinformatics, 2005. **21**(9): p. 2027-35.

12. Wells, A., et al., Epidermal growth factor receptor-mediated motility in fibroblasts. Microsc Res Tech, 1998. **43**(5): p. 395-411.

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

14. Mamoune, A., et al., Calpain-2 as a target for limiting prostate cancer invasion. Cancer Res, 2003. **63**(15): p. 4632-40.

15. Kassis, J., et al., A role for phospholipase C-gamma-mediated signaling in tumor cell invasion. Clin Cancer Res, 1999. **5**(8): p. 2251-60.

16. Kassis, J., R. Radinsky, and A. Wells, Motility is rate-limiting for invasion of bladder carcinoma cell lines. Int J Biochem Cell Biol, 2002. **34**(7): p. 762-75.

17. Pollard, T.D., The cytoskeleton, cellular motility and the reductionist agenda. Nature, 2003. **422**(6933): p. 741-5.

18. Condeelis, J.S., et al., Lamellipodia in invasion. Semin Cancer Biol, 2001. **11**(2): p. 119-28.

19. Zebda, N., et al., Phosphorylation of ADF/cofilin abolishes EGF-induced actin nucleation at the leading edge and subsequent lamellipod extension. J Cell Biol, 2000. **151**(5): p. 1119-28.

20. Meyer, G. and E.L. Feldman, Signaling mechanisms that regulate actin-based motility processes in the nervous system. J Neurochem, 2002. **83**(3): p. 490-503.

21. Chou, J., et al., Distribution of gelsolin and phosphoinositol 4,5-bisphosphate in lamellipodia during EGF-induced motility. Int J Biochem Cell Biol, 2002. **34**(7): p. 776-90.

22. Nobes, C.D. and A. Hall, Rho, rac and cdc42 GTPases: regulators of actin structures, cell adhesion and motility. Biochem Soc Trans, 1995. **23**(3): p. 456-9.

23. Nobes, C.D. and A. Hall, Rho GTPases control polarity, protrusion, and adhesion during cell movement. J Cell Biol, 1999. **144**(6): p. 1235-44.

24. Hall, A. and C.D. Nobes, Rho GTPases: molecular switches that control the organization and dynamics of the actin cytoskeleton. Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1399): p. 965-70.

25. Chou, J., et al., Directional motility induced by epidermal growth factor requires Cdc42. Exp Cell Res, 2003. **287**(1): p. 47-56.

26. Turner, C.E., Paxillin interactions. J Cell Sci, 2000. **113 Pt 23**: p. 4139-40.

27. Parsons, J.T., et al., Focal adhesion kinase: a regulator of focal adhesion dynamics and cell movement. Oncogene, 2000. **19**(49): p. 5606-13.

28. Shen, T.L. and J.L. Guan, Differential regulation of cell migration and cell cycle progression by FAK complexes with Src, PI3K, Grb7 and Grb2 in focal contacts. FEBS Lett, 2001. **499**(1-2): p. 176-81.

29. 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.

30. Palecek, S.P., et al., Integrin-ligand binding properties govern cell migration speed through cell-substratum adhesiveness. Nature, 1997. **385**(6616): p. 537-40.

31. Maheshwari, G., et al., Biophysical integration of effects of epidermal growth factor and fibronectin on fibroblast migration. Biophys J, 1999. **76**(5): p. 2814-23.

32. Palecek, S.P., et al., Integrin dynamics on the tail region of migrating fibroblasts. J Cell Sci, 1996. **109 (Pt 5)**: p. 941-52.

33. Palecek, S.P., et al., Physical and biochemical regulation of integrin release during rear detachment of migrating cells. J Cell Sci, 1998. **111 (Pt 7)**: p. 929-40.

34. Iwabu, A., et al., Epidermal growth factor induces fibroblast contractility and motility via a protein kinase C delta-dependent pathway. J Biol Chem, 2004. **279**(15): p. 14551-60.

35. Darenfed, H. and C.A. Mandato, Wound-induced contractile ring: a model for cytokinesis. Biochem Cell Biol, 2005. **83**(6): p. 711-20.

36. Levinson, H., et al., Calmodulin-myosin light chain kinase inhibition changes fibroblast-populated collagen lattice contraction, cell migration, focal adhesion formation, and wound contraction. Wound Repair Regen, 2004. **12**(5): p. 505-11.

37. Riento, K. and A.J. Ridley, Rocks: multifunctional kinases in cell behaviour. Nat Rev Mol Cell Biol, 2003. **4**(6): p. 446-56.

38. Totsukawa, G., et al., Distinct roles of ROCK (Rho-kinase) and MLCK in spatial regulation of MLC phosphorylation for assembly of stress fibers and focal adhesions in 3T3 fibroblasts. J Cell Biol, 2000. **150**(4): p. 797-806.

39. Brzeska, H., et al., Rac-induced increase of phosphorylation of myosin regulatory light chain in HeLa cells. Cell Motil Cytoskeleton, 2004. **58**(3): p. 186-99.

40. Totsukawa, G., et al., Distinct roles of MLCK and ROCK in the regulation of membrane protrusions and focal adhesion dynamics during cell migration of fibroblasts. J Cell Biol, 2004. **164**(3): p. 427-39.

41. Somlyo, A.V., et al., Rho-kinase inhibitor retards migration and in vivo dissemination of human prostate cancer cells. Biochem Biophys Res Commun, 2000. **269**(3): p. 652-9.

42. Komatsu, S. and M. Ikebe, ZIP kinase is responsible for the phosphorylation of myosin II and necessary for cell motility in mammalian fibroblasts. J Cell Biol, 2004. **165**(2): p. 243-54.

43. Wilson, D.P., et al., Integrin-linked kinase is responsible for Ca2+-independent myosin diphosphorylation and contraction of vascular smooth muscle. Biochem J, 2005. **392**(Pt 3): p. 641-8.

44. Huttenlocher, A., et al., Regulation of cell migration by the calcium-dependent protease calpain. J Biol Chem, 1997. **272**(52): p. 32719-22.

45. Glading, A., et al., Epidermal growth factor activates m-calpain (calpain II), at least in part, by extracellular signal-regulated kinase-mediated phosphorylation. Mol Cell Biol, 2004. **24**(6): p. 2499-512.

46. Lee, J., et al., Regulation of cell movement is mediated by stretch-activated calcium channels. Nature, 1999. **400**(6742): p. 382-6.

47. Palecek, S.P., A.F. Horwitz, and D.A. Lauffenburger, Kinetic model for integrin-mediated adhesion release during cell migration. Ann Biomed Eng, 1999. **27**(2): p. 219-35.

48. Rios-Doria, J., et al., The role of calpain in the proteolytic cleavage of E-cadherin in prostate and mammary epithelial cells. J Biol Chem, 2003. **278**(2): p. 1372-9.

49. Lokeshwar, B.L., MMP inhibition in prostate cancer. Ann N Y Acad Sci, 1999. **878**: p. 271-89.

50. Zhao, Y.G., et al., Activation of pro-gelatinase B by endometase/matrilysin-2 promotes invasion of human prostate cancer cells. J Biol Chem, 2003. **278**(17): p. 15056-64.

51. Bok, R.A., et al., Patterns of protease production during prostate cancer progression: proteomic evidence for cascades in a transgenic model. Prostate Cancer Prostatic Dis, 2003. **6**(4): p. 272-80.

52. Paule, B. and N. Brion, [EGF receptors in urological cancer. Molecular basis and therapeutic involvements]. Ann Med Interne (Paris), 2003. **154**(7): p. 448-56.

53. Onn, A., et al., Synchronous overexpression of epidermal growth factor receptor and HER2-neu protein is a predictor of poor outcome in patients with stage I non-small cell lung cancer. Clin Cancer Res, 2004. **10**(1 Pt 1): p. 136-43.

54. Shinojima, N., et al., Prognostic value of epidermal growth factor receptor in patients with glioblastoma multiforme. Cancer Res, 2003. **63**(20): p. 6962-70.

55. Wells, A., J.M. Bishop, and D. Helmeste, Amplified gene for the epidermal growth factor receptor in a human glioblastoma cell line encodes an enzymatically inactive protein. Mol Cell Biol, 1988. **8**(10): p. 4561-5.

56. Prati, R., et al., Histopathologic characteristics predicting HER-2/neu amplification in breast cancer. Breast J, 2005. **11**(6): p. 433-9.

57. Myers, R.B., J.E. Kudlow, and W.E. Grizzle, Expression of transforming growth factor-alpha, epidermal growth factor and the epidermal growth factor receptor in adenocarcinoma of the prostate and benign prostatic hyperplasia. Mod Pathol, 1993. **6**(6): p. 733-7.

58. Wells, A., et al., Growth factor-induced cell motility in tumor invasion. Acta Oncol, 2002. **41**(2): p. 124-30.

59. Kim, H.G., et al., EGF receptor signaling in prostate morphogenesis and tumorigenesis. Histol Histopathol, 1999. **14**(4): p. 1175-82.

60. Turner, T., et al., EGF receptor signaling enhances in vivo invasiveness of DU-145 human prostate carcinoma cells. Clin Exp Metastasis, 1996. **14**(4): p. 409-18.

61. Morton, G.C., Early prostate cancer. Curr Probl Cancer, 2000. **24**(1): p. 5-51.

62. Schild, S.E., Radiation therapy (RT) after prostatectomy: The case for salvage therapy as opposed to adjuvant therapy. Int J Cancer, 2001. **96**(2): p. 94-8.

63. Porter, A.T., A. Williams, and J.D. Forman, Adjuvant radiation therapy for pathologic T3 prostate cancer. Can J Urol, 1997. **4**(2 Supp 1): p. 57-60.

64. Shah, R.B., et al., Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. Cancer Res, 2004. **64**(24): p. 9209-16.

65. Banyard, J. and B.R. Zetter, The role of cell motility in prostate cancer. Cancer Metastasis Rev, 1998. **17**(4): p. 449-58.

66. Nakashiro, K., et al., Phenotypic switch from paracrine to autocrine role of hepatocyte growth factor in an androgen-independent human prostatic carcinoma cell line, CWR22R. Am J Pathol, 2004. **165**(2): p. 533-40.

67. Hurle, R.A., et al., Hepatocyte growth factor/scatter factor and prostate cancer: a review. Histol Histopathol, 2005. **20**(4): p. 1339-49.

68. Oosterhoff, J.K., J.A. Grootegoed, and L.J. Blok, Expression profiling of androgen-dependent and -independent LNCaP cells: EGF versus androgen signalling. Endocr Relat Cancer, 2005. **12**(1): p. 135-48.

69. Mamoune, A., et al., DU145 human prostate carcinoma invasiveness is modulated by urokinase receptor (uPAR) downstream of epidermal growth factor receptor (EGFR) signaling. Exp Cell Res, 2004. **299**(1): p. 91-100.

70. Kondapaka, S.B., R. Fridman, and K.B. Reddy, Epidermal growth factor and amphiregulin up-regulate matrix metalloproteinase-9 (MMP-9) in human breast cancer cells. Int J Cancer, 1997. **70**(6): p. 722-6.

71. Fujiuchi, Y., et al., Effect of hepatocyte growth factor on invasion of prostate cancer cell lines. Oncol Rep, 2003. **10**(4): p. 1001-6.

72. Aaronson, S.A., Growth factors and cancer. Science, 1991. **254**(5035): p. 1146-53.

73. Tzahar, E., et al., Bivalence of EGF-like ligands drives the ErbB signaling network. Embo J, 1997. **16**(16): p. 4938-50.

74. Wells, A., EGF receptor. Int J Biochem Cell Biol, 1999. **31**(6): p. 637-43.

75. Schlegel, J., et al., Amplification of the epidermal-growth-factor-receptor gene correlates with different growth behaviour in human glioblastoma. Int J Cancer, 1994. **56**(1): p. 72-7.

76. Bargmann, C.I. and R.A. Weinberg, Increased tyrosine kinase activity associated with the protein encoded by the activated neu oncogene. Proc Natl Acad Sci U S A, 1988. **85**(15): p. 5394-8.

77. van der Poel, H.G., Smart drugs in prostate cancer. Eur Urol, 2004. **45**(1): p. 1-17.

78. Kim, S.J., et al., Blockade of epidermal growth factor receptor signaling in tumor cells and tumor-associated endothelial cells for therapy of androgen-independent human prostate cancer growing in the bone of nude mice. Clin Cancer Res, 2003. **9**(3): p. 1200-10.

79. Glading, A., et al., Membrane proximal ERK signaling is required for M-calpain activation downstream of epidermal growth factor receptor signaling. J Biol Chem, 2001. **276**(26): p. 23341-8.

80. Gonzalez-Guerrico, A.M. and M.G. Kazanietz, Phorbol ester-induced apoptosis in prostate cancer cells via autocrine activation of the extrinsic apoptotic cascade: a key role for protein kinase C delta. J Biol Chem, 2005. **280**(47): p. 38982-91.

81. Shelton, J.G., et al., Effects of endogenous epidermal growth factor receptor signaling on DNA synthesis and ERK activation in a cytokine-dependent hematopoietic cell line. Cell Cycle, 2005. **4**(6): p. 818-21.

82. El Sheikh, S.S., et al., Phosphorylation of both EGFR and ErbB2 is a reliable predictor of prostate cancer cell proliferation in response to EGF. Neoplasia, 2004. **6**(6): p. 846-53.

83. Chen, P., et al., Epidermal growth factor receptor-mediated cell motility: phospholipase C activity is required, but mitogen-activated protein kinase activity is not sufficient for induced cell movement. J Cell Biol, 1994. **127**(3): p. 847-57.

84. Chen, P., J.E. Murphy-Ullrich, and A. Wells, A role for gelsolin in actuating epidermal growth factor receptor-mediated cell motility. J Cell Biol, 1996. **134**(3): p. 689-98.

85. Wells, A., et al., Shaping up for shipping out: PLCgamma signaling of morphology changes in EGF-stimulated fibroblast migration. Cell Motil Cytoskeleton, 1999. **44**(4): p. 227-33.

86. Vivanco, I. and C.L. Sawyers, The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer, 2002. **2**(7): p. 489-501.

87. Goberdhan, D.C. and C. Wilson, PTEN: tumour suppressor, multifunctional growth regulator and more. Hum Mol Genet, 2003. **12 Spec No 2**: p. R239-48.

88. Kauffmann-Zeh, A., et al., Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. Nature, 1997. **385**(6616): p. 544-8.

89. Kubiatowski, T., et al., Association of increased phosphatidylinositol 3-kinase signaling with increased invasiveness and gelatinase activity in malignant gliomas. J Neurosurg, 2001. **95**(3): p. 480-8.

90. Ellerbroek, S.M., et al., Phosphatidylinositol 3-kinase activity in epidermal growth factor-stimulated matrix metalloproteinase-9 production and cell surface association. Cancer Res, 2001. **61**(5): p. 1855-61.

91. Park, B.K., X. Zeng, and R.I. Glazer, Akt1 induces extracellular matrix invasion and matrix metalloproteinase-2 activity in mouse mammary epithelial cells. Cancer Res, 2001. **61**(20): p. 7647-53.

92. Kim, D., et al., Akt/PKB promotes cancer cell invasion via increased motility and metalloproteinase production. Faseb J, 2001. **15**(11): p. 1953-62.

93. Arboleda, M.J., et al., Overexpression of AKT2/protein kinase Bbeta leads to up-regulation of beta1 integrins, increased invasion, and metastasis of human breast and ovarian cancer cells. Cancer Res, 2003. **63**(1): p. 196-206.

94. Cooper, C.R., C.H. Chay, and K.J. Pienta, The role of alpha(v)beta(3) in prostate cancer progression. Neoplasia, 2002. **4**(3): p. 191-4.

95. Wells, A., et al., Ligand-induced transformation by a noninternalizing epidermal growth factor receptor. Science, 1990. **247**(4945): p. 962-4.

96. Haugh, J.M., et al., Effect of epidermal growth factor receptor internalization on regulation of the phospholipase C-gamma1 signaling pathway. J Biol Chem, 1999. **274**(13): p. 8958-65.

97. Wang, Y., et al., Endosomal signaling of epidermal growth factor receptor stimulates signal transduction pathways leading to cell survival. Mol Cell Biol, 2002. **22**(20): p. 7279-90.

98. Reddy, C.C., et al., Engineering epidermal growth factor for enhanced mitogenic potency. Nat Biotechnol, 1996. **14**(13): p. 1696-9.

99. Swindle, C.S., et al., Epidermal growth factor (EGF)-like repeats of human tenascin-C as ligands for EGF receptor. J Cell Biol, 2001. **154**(2): p. 459-68.

100. Schenk, S., et al., Binding to EGF receptor of a laminin-5 EGF-like fragment liberated during MMP-dependent mammary gland involution. J Cell Biol, 2003. **161**(1): p. 197-209.

101. Iozzo, R.V., et al., Decorin is a biological ligand for the epidermal growth factor receptor. J Biol Chem, 1999. **274**(8): p. 4489-92.

102. Stetler-Stevenson, W.G., S. Aznavoorian, and L.A. Liotta, Tumor cell interactions with the extracellular matrix during invasion and metastasis. Annu Rev Cell Biol, 1993. **9**: p. 541-73.

103. Danen, E.H., Integrins: regulators of tissue function and cancer progression. Curr Pharm Des, 2005. **11**(7): p. 881-91.

104. Fornaro, M., T. Manes, and L.R. Languino, Integrins and prostate cancer metastases. Cancer Metastasis Rev, 2001. **20**(3-4): p. 321-31.

105. Cox, E.A. and A. Huttenlocher, Regulation of integrin-mediated adhesion during cell migration. Microsc Res Tech, 1998. **43**(5): p. 412-9.

106. Geiger, B., Cell biology. Encounters in space. Science, 2001. **294**(5547): p. 1661-3.

107. Geiger, B., et al., Transmembrane crosstalk between the extracellular matrix--cytoskeleton crosstalk. Nat Rev Mol Cell Biol, 2001. **2**(11): p. 793-805.

108. Cukierman, E., et al., Taking cell-matrix adhesions to the third dimension. Science, 2001. **294**(5547): p. 1708-12.

109. Smilenov, L.B., et al., Focal adhesion motility revealed in stationary fibroblasts. Science, 1999. **286**(5442): p. 1172-4.

110. Bussey, H., Rho returns: its targets in focal adhesions. Science, 1996. **273**(5272): p. 203.

111. Parsons, J.T., Focal adhesion kinase: the first ten years. J Cell Sci, 2003. **116**(Pt 8): p. 1409-16.

112. Martin, K.H., et al., Integrin connections map: to infinity and beyond. Science, 2002. **296**(5573): p. 1652-3.

113. Wild-Bode, C., M. Weller, and W. Wick, Molecular determinants of glioma cell migration and invasion. J Neurosurg, 2001. **94**(6): p. 978-84.

114. Demetriou, M.C. and A.E. Cress, Integrin clipping: a novel adhesion switch? J Cell Biochem, 2004. **91**(1): p. 26-35.

115. Cress, A.E., et al., The alpha 6 beta 1 and alpha 6 beta 4 integrins in human prostate cancer progression. Cancer Metastasis Rev, 1995. **14**(3): p. 219-28.

116. Schwartz, M.A., Integrins, oncogenes, and anchorage independence. J Cell Biol, 1997. **139**(3): p. 575-8.

117. Zahir, N., et al., Autocrine laminin-5 ligates alpha6beta4 integrin and activates RAC and NFkappaB to mediate anchorage-independent survival of mammary tumors. J Cell Biol, 2003. **163**(6): p. 1397-407.

118. Schwartz, M.A., M.D. Schaller, and M.H. Ginsberg, Integrins: emerging paradigms of signal transduction. Annu Rev Cell Dev Biol, 1995. **11**: p. 549-99.

119. Slack-Davis, J.K. and J.T. Parsons, Emerging views of integrin signaling: implications for prostate cancer. J Cell Biochem, 2004. **91**(1): p. 41-6.

120. Roberts, P.C., et al., Sequential molecular and cellular events during neoplastic progression: a mouse syngeneic ovarian cancer model. Neoplasia, 2005. **7**(10): p. 944-56.

121. Larue, L. and A. Bellacosa, Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. Oncogene, 2005. **24**(50): p. 7443-54.

122. Matsumura, T., R. Makino, and K. Mitamura, Frequent down-regulation of E-cadherin by genetic and epigenetic changes in the malignant progression of hepatocellular carcinomas. Clin Cancer Res, 2001. **7**(3): p. 594-9.

123. Berx, G. and F. Van Roy, The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. Breast Cancer Res, 2001. **3**(5): p. 289-93.

124. Matabuena de Yzaguirre, M., et al., Epigenetic silencing of E- and N-cadherins in the stroma of mouse thymic lymphomas. Carcinogenesis, 2005.

125. Kallakury, B.V., C.E. Sheehan, and J.S. Ross, Co-downregulation of cell adhesion proteins alpha- and beta-catenins, p120CTN, E-cadherin, and CD44 in prostatic adenocarcinomas. Hum Pathol, 2001. **32**(8): p. 849-55.

126. Cohen, M.B., et al., Cellular adhesion molecules in urologic malignancies. Am J Clin Pathol, 1997. **107**(1): p. 56-63.

127. Yates, C., A. Wells, and T. Turner, Luteinising hormone-releasing hormone analogue reverses the cell adhesion profile of EGFR overexpressing DU-145 human prostate carcinoma subline. Br J Cancer, 2005. **92**(2): p. 366-75.

128. Knox, J.D., et al., Differential expression of extracellular matrix molecules and the alpha 6-integrins in the normal and neoplastic prostate. Am J Pathol, 1994. **145**(1): p. 167-74.

129. von Bredow, D.C., et al., Cleavage of beta 4 integrin by matrilysin. Exp Cell Res, 1997. **236**(1): p. 341-5.

130. Kolanus, W. and B. Seed, Integrins and inside-out signal transduction: converging signals from PKC and PIP3. Curr Opin Cell Biol, 1997. **9**(5): p. 725-31.

131. Araujo, R.P. and L.A. Liotta, A control theoretic paradigm for cell signaling networks: a simple complexity for a sensitive robustness. Curr Opin Chem Biol, 2006. **10**(1): p. 81-7.

132. Calvo, K.R., L.A. Liotta, and E.F. Petricoin, Clinical proteomics: from biomarker discovery and cell signaling profiles to individualized personal therapy. Biosci Rep, 2005. **25**(1-2): p. 107-25.

133. Glading, A., et al., Epidermal growth factor receptor activation of calpain is required for fibroblast motility and occurs via an ERK/MAP kinase signaling pathway. J Biol Chem, 2000. **275**(4): p. 2390-8.

134. Ideker, T. and D. Lauffenburger, Building with a scaffold: emerging strategies for high- to low-level cellular modeling. Trends Biotechnol, 2003. **21**(6): p. 255-62.

135. Janes, K.A. and D.A. Lauffenburger, A biological approach to computational models of proteomic networks. Curr Opin Chem Biol, 2006. **10**(1): p. 73-80.

136. Janes, K.A., et al., A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science, 2005. **310**(5754): p. 1646-53.

137. Woolf, P.J., et al., Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. Bioinformatics, 2005. **21**(6): p. 741-53.

138. Ridley, A.J., et al., Cell migration: integrating signals from front to back. Science, 2003. **302**(5651): p. 1704-9.

139. Chen, P., K. Gupta, and A. Wells, Cell movement elicited by epidermal growth factor receptor requires kinase and autophosphorylation but is separable from mitogenesis. J Cell Biol, 1994. **124**(4): p. 547-55.

140. Webb, D.J., et al., FAK-Src signalling through paxillin, ERK and MLCK regulates adhesion disassembly. Nat Cell Biol, 2004. **6**(2): p. 154-61.

141. Hautaniemi, S., et al., A novel strategy for microarray quality control using Bayesian networks. Bioinformatics, 2003. **19**(16): p. 2031-8.

142. Huang, C., K. Jacobson, and M.D. Schaller, MAP kinases and cell migration. J Cell Sci, 2004. **117**(Pt 20): p. 4619-28.

143. Grubb, R.L., et al., Signal pathway profiling of prostate cancer using reverse phase protein arrays. Proteomics, 2003. **3**(11): p. 2142-6.

144. Barrett, J.C., et al., Linking laboratory and clinical research: the development of molecularly targeted therapeutics inside the national cancer institute center for cancer research. Clin Adv Hematol Oncol, 2003. **1**(5): p. 302-6.

145. Wulfkuhle, J., et al., Genomic and proteomic technologies for individualisation and improvement of cancer treatment. Eur J Cancer, 2004. **40**(17): p. 2623-32.

146. Jackson, D.N. and D.A. Foster, The enigmatic protein kinase Cdelta: complex roles in cell proliferation and survival. Faseb J, 2004. **18**(6): p. 627-36.

147. Long, B.J. and D.P. Rose, Invasive capacity and regulation of urokinase-type plasminogen activator in estrogen receptor (ER)-negative MDA-MB-231 human breast cancer cells, and a transfectant (S30) stably expressing ER. Cancer Lett, 1996. **99**(2): p. 209-15.

148. Bichsel, V.E., L.A. Liotta, and E.F. Petricoin, 3rd, Cancer proteomics: from biomarker discovery to signal pathway profiling. Cancer J, 2001. **7**(1): p. 69-78.

149. Petricoin, E.F. and L.A. Liotta, Clinical applications of proteomics. J Nutr, 2003. **133**(7 Suppl): p. 2476S-2484S.

150. Clark, R.A., et al., Fibroblast migration on fibronectin requires three distinct functional domains. J Invest Dermatol, 2003. **121**(4): p. 695-705.

151. Allen, F.D., et al., Epidermal growth factor induces acute matrix contraction and subsequent calpain-modulated relaxation. Wound Repair Regen, 2002. **10**(1): p. 67-76.

152. Huang, S., et al., Dual-agent molecular targeting of the epidermal growth factor receptor (EGFR): combining anti-EGFR antibody with tyrosine kinase inhibitor. Cancer Res, 2004. **64**(15): p. 5355-62.

153. Blackledge, G., et al., Anti-EGF receptor therapy. Prostate Cancer Prostatic Dis, 2000. **3**(4): p. 296-302.

154. Geho, D.H., et al., Physiological mechanisms of tumor-cell invasion and migration. Physiology (Bethesda), 2005. **20**: p. 194-200.

155. Kaneko, K., et al., Myosin light chain kinase inhibitors can block invasion and adhesion of human pancreatic cancer cell lines. Pancreas, 2002. **24**(1): p. 34-41.

156. Yin, L., N. Bennani-Baiti, and C.T. Powell, Phorbol ester-induced apoptosis of C4-2 cells requires both a unique and a redundant protein kinase C signaling pathway. J Biol Chem, 2005. **280**(7): p. 5533-41.

157. Nabha, S.M., et al., Upregulation of PKC-delta contributes to antiestrogen resistance in mammary tumor cells. Oncogene, 2005. **24**(19): p. 3166-76.

158. Kruger, J.S. and K.B. Reddy, Distinct mechanisms mediate the initial and sustained phases of cell migration in epidermal growth factor receptor-overexpressing cells. Mol Cancer Res, 2003. **1**(11): p. 801-9.

159. Kiley, S.C., et al., Increased protein kinase C delta in mammary tumor cells: relationship to transformtion and metastatic progression. Oncogene, 1999. **18**(48): p. 6748-57.

160. Koren, R., et al., Expression of protein kinase C isoenzymes in benign hyperplasia and carcinoma of prostate. Oncol Rep, 2004. **11**(2): p. 321-6.

161. Kiley, S.C., et al., Protein kinase C delta involvement in mammary tumor cell metastasis. Cancer Res, 1999. **59**(13): p. 3230-8.

162. Xie, H., et al., In vitro invasiveness of DU-145 human prostate carcinoma cells is modulated by EGF receptor-mediated signals. Clin Exp Metastasis, 1995. **13**(6): p. 407-19.

163. Kaighn, M.E., et al., Establishment and characterization of a human prostatic carcinoma cell line (PC-3). Invest Urol, 1979. **17**(1): p. 16-23.

164. Balendran, A., et al., Further evidence that 3-phosphoinositide-dependent protein kinase-1 (PDK1) is required for the stability and phosphorylation of protein kinase C (PKC) isoforms. FEBS Lett, 2000. **484**(3): p. 217-23.

165. Sumitomo, M., et al., Protein kinase Cdelta amplifies ceramide formation via mitochondrial signaling in prostate cancer cells. J Clin Invest, 2002. **109**(6): p. 827-36.

166. Straussman, R., L. Even, and S. Ravid, Myosin II heavy chain isoforms are phosphorylated in an EGF-dependent manner: involvement of protein kinase C. J Cell Sci, 2001. **114**(Pt 16): p. 3047-57.

167. Gillespie, G.Y., et al., Glioma migration can be blocked by nontoxic inhibitors of myosin II. Cancer Res, 1999. **59**(9): p. 2076-82.

168. Tohtong, R., et al., Dependence of metastatic cancer cell invasion on MLCK-catalyzed phosphorylation of myosin regulatory light chain. Prostate Cancer Prostatic Dis, 2003. **6**(3): p. 212-6.

169. Kassis, J., et al., Tumor invasion as dysregulated cell motility. Semin Cancer Biol, 2001. **11**(2): p. 105-17.

170. Deng, J.T., et al., Ca2+-independent smooth muscle contraction. a novel function for integrin-linked kinase. J Biol Chem, 2001. **276**(19): p. 16365-73.

171. Myat, M.M., et al., MARCKS regulates membrane ruffling and cell spreading. Curr Biol, 1997. **7**(8): p. 611-4.

172. Barry, S.T. and D.R. Critchley, The RhoA-dependent assembly of focal adhesions in Swiss 3T3 cells is associated with increased tyrosine phosphorylation and the recruitment of both pp125FAK and protein kinase C-delta to focal adhesions. J Cell Sci, 1994. **107 (Pt 7)**: p. 2033-45.

173. Li, C., et al., Mechanical stress-activated PKCdelta regulates smooth muscle cell migration. Faseb J, 2003. **17**(14): p. 2106-8.

174. Wiedlocha, A., et al., Phosphorylation-regulated nucleocytoplasmic trafficking of internalized fibroblast growth factor-1. Mol Biol Cell, 2005. **16**(2): p. 794-810.

175. Mimura, Y., et al., Epidermal growth factor induces fibronectin expression in human dermal fibroblasts via protein kinase C delta signaling pathway. J Invest Dermatol, 2004. **122**(6): p. 1390-8.

176. Harms, B.D., et al., Directional persistence of EGF-induced cell migration is associated with stabilization of lamellipodial protrusions. Biophys J, 2005. **88**(2): p. 1479-88.

177. Kaplan, P.J., et al., Involvement of transforming growth factor alpha (TGFalpha) and epidermal growth factor receptor (EGFR) in sex hormone-induced prostatic dysplasia and the growth of an androgen-independent transplantable carcinoma of the prostate. Carcinogenesis, 1996. **17**(12): p. 2571-9.

178. Espina, V., et al., Pathology of the future: molecular profiling for targeted therapy. Cancer Invest, 2005. **23**(1): p. 36-46.

179. Cowherd, S.M., et al., Proteomic analysis of human breast cancer tissue with laser-capture microdissection and reverse-phase protein microarrays. Clin Breast Cancer, 2004. **5**(5): p. 385-92.

180. Espina, V., et al., Protein microarray detection strategies: focus on direct detection technologies. J Immunol Methods, 2004. **290**(1-2): p. 121-33.

181. Lodish, H., et al., Molecular Cell Biology. W.H.Freeman & Co, New York.

182. Breiman, L., et al., Classification and Regression Trees. 1984. Wadsworth.

183. Efron, B., et al. An Introduction to Bootstrap. 1994. Chapman & Hall, London.

184. Haykin, S. Neural Networks: A comprehensive Foundation, 2nd Edition. 1999. Prentice Hall, Inc., Upper Saddle River, NJ.

185. Hochberg, Y et al. Multiple Comparision Procedures. 1987. John & Wiley Sons, New York.

186. Lloyd, S. Least Square quantization in PCM. IEEE Transactions on Information Theory, IT-28, 129-137.

187. Pearl, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Interference. 1988. Morgan Kaufmann Publishers Sam Mateo, CA.

188. Rissanen, J. Modeling by shortest data description. 1978. Automatica, 14, 465-471.

189. Rissanen, J. Stochastic complexity in Statistical Inquiry. 1998. World Scientific, Singapore.

190. Rissanen, J. MDL denoising. IEEE Transactions on Information Theory. 2000; 46, 2537-2543.

191. Lynch, T. J. *et al*. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 2004. **350**, 2129−2139

192. Paez, J. G. *et al*. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 2004. **304**, 1497−1500 (2004).